Dissertation

Least-Squares Monte Carlo Methods in the Life Insurance Sector

Anne-Sophie Krah

Vom Fachbereich Mathematik der Technischen Universität Kaiserslautern
zur Verleihung des akademischen Grades Doktor der Naturwissenschaften
(Doctor rerum naturalium, Dr. rer. nat.) genehmigte Dissertation

1. Gutachter: Prof. Dr. Ralf Korn
2. Gutachter: Prof. Dr. Ralf Werner

Datum der Disputation: 30. April 2021

D 386

Acknowledgements

Foremost, I would like to express my deep respect and genuine gratitude to my supervisor Professor Dr. Ralf Korn from TU Kaiserslautern for his enthusiasm, active interest, continuous support and very constructive advice throughout my doctoral journey. At all times, working with you has been insightful and felt natural. Not only have you always been there for any kind of help but you have also given me the trust and freedom to make progress autonomously. I would like to emphasize my thanks for your openness to this external research project, reading my drafts very closely and commenting on them in detail, inviting me to present parts of this thesis at DFG-Graduiertenkolleg 1932 on May 2, 2017, in Kaiserslautern and the collaboration on our three joint LSMC articles published in Risks in June 2018, February 2020 and November 2020. It has been a great privilege to write this external thesis under your guidance.

My sincere thanks also goes to Professor Dr. Ralf Werner from Augsburg University who acts as a referee for this dissertation and invited me to present the results in his seminars. Thank you for your perspective: I am glad to have enriched the manuscript for publication with additional resources based on your input. Moreover, I am grateful to Professor Dr. Caroline Lassueur and Professor Dr. Jörn Saß from the thesis committee of TU Kaiserslautern for participating at my journey.

Next, I wish to acknowledge my former head of department Dr. Zoran Nikolić for having given me the opportunity to pursue simultaneously my professional and academic goals while realizing precious synergies. I appreciate your keen interest in conducting research to drive the methodological development of the insurance industry forward and am glad to have taken an active part in it. In particular, I am thankful for the extensive data sets, your collaboration on our three joint articles with my supervisor and the invitation to present parts of this thesis at max.99 event of DAV on September 16, 2019, in Cologne.

Furthermore, I would like to convey my gratitude to Professor Dr. Christian Weiß from Hochschule Ruhr West for recognizing my desire to write a doctoral thesis, establishing the contact to my supervisor, and proofreading. I very much value your attentiveness, honest opinion on my drafts and the inspiring suggestions. In addition, I would like to thank Dr. Christian Jonen and Boris Gendler for beneficial discussions and proofreading.

My gratitude also goes to my former and current direct superiors Dr. Magdalena Roth and Dr. Tamino Meyhöfer from Generali Deutschland AG who have been supportive by assigning topics related to this thesis to me and who have taken into account my special working time model in their planning. Moreover, I am thankful to all my colleagues for allocating their excess computational capacities to me for this research project.

Lastly, I wish to thank my dear family, friends and teachers for their encouragement and strong belief in me. I am grateful to my parents who have supported me in all my pursuits, to my friends who have appreciated my affection for science and to my teachers who have contributed to giving me the confidence to take this path.

Abstract

Life insurance companies are asked by the Solvency II regime to retain capital requirements against economically adverse developments. This ensures that they are continuously able to meet their payment obligations towards the policyholders. When relying on an internal model approach, an insurer's solvency capital requirement is defined as the 99.5% value-at-risk of its full loss probability distribution over the coming year. In the introductory part of this thesis, we provide the actuarial modeling tools and risk aggregation methods by which the companies can accomplish the derivations of these forecasts.

Since the industry still lacks the computational capacities to fully simulate these distributions, the insurers have to refer to suitable approximation techniques such as the least-squares Monte Carlo (LSMC) method. The key idea of LSMC is to run only a few wisely selected simulations and to process their output further to obtain a risk-dependent proxy function of the loss. We dedicate the first part of this thesis to establishing a theoretical framework of the LSMC method. We start with how LSMC for calculating capital requirements is related to its original use in American option pricing. Then we decompose LSMC into four steps. In the first one, the Monte Carlo simulation setting is defined. The second and third steps serve the calibration and validation of the proxy function, and the fourth step yields the loss distribution forecast by evaluating the proxy model. When guiding through the steps, we address practical challenges and propose an adaptive calibration algorithm. We complete with a slightly disguised real-world application.

The second part builds upon the first one by taking up the LSMC framework and diving deeper into its calibration step. After a literature review and a basic recapitulation, various adaptive machine learning approaches relying on least-squares regression and model selection criteria are presented as solutions to the proxy modeling task. The studied approaches range from ordinary and generalized least-squares regression variants over GLM and GAM methods to MARS and kernel regression routines. We justify the combinability of the regression ingredients mathematically and compare their approximation quality in slightly altered real-world experiments. Thereby, we perform sensitivity analyses, discuss numerical stability and run comprehensive out-of-sample tests. The scope of the analyzed regression variants extends to other high-dimensional variable selection applications.

Life insurance contracts with early exercise features can be priced by LSMC as well due to their analogies to American options. In the third part of this thesis, equity-linked contracts with American-style surrender options and minimum interest rate guarantees payable upon contract termination are valued. We allow randomness and jumps in the movements of the interest rate, stochastic volatility, stock market and mortality. For the simultaneous valuation of numerous insurance contracts, a hybrid probability measure and an additional regression function are introduced. Furthermore, an efficient seed-related simulation procedure accounting for the forward discretization bias and a validation concept are proposed. An extensive numerical example rounds off the last part.

Zusammenfassung

Zur Gewährleistung ihrer Solvabilität auch in nachteiligen ökonomischen Szenarien werden Lebensversicherer unter Solvency II aufgefordert, ein Solvenzkapital vorzuhalten. Für Versicherer mit einem internen Modell ist dieses als der 99.5%-Value-at-Risk ihrer einjährigen Verlustverteilung definiert. Im Einführungsteil dieser Dissertation werden die Grundlagen der aktuariellen Modellierung zur Bestimmung dieser Wahrscheinlichkeitsverteilung einschließlich bekannter Methoden zur Risikoaggregation vorgestellt.

Derzeit verfügen die Versicherer bei Weitem nicht über die Computerkapazitäten, die zur vollständigen Simulation ihrer Verlustverteilung notwendig wären. Deswegen müssen sie auf Approximationstechniken wie die Least-Squares Monte Carlo (LSMC)-Methodik zurückgreifen. Die Idee von LSMC besteht darin, nur wenige geschickt gewählte Simulationen durchzuführen und daraus approximativ den Zusammenhang zwischen Verlust und Risiken herzuleiten. Im ersten Teil dieser Dissertation wird ein theoretischer Rahmen für den LSMC-Ansatz zur Solvenzkapitalberechnung geschaffen. Nachdem die Verbindung zur Bewertung amerikanischer Optionen steht, wird der Ansatz in vier Schritte zerlegt. Im ersten Schritt wird das Monte Carlo Setting definiert. Dann folgen die Schritte zwei und drei zur Kalibrierung und Validierung der Proxyfunktion. Im vierten Schritt wird diese zur Vorhersage der Verlustverteilung angewendet. Parallel werden die Herausforderungen in der Praxis thematisiert sowie ein adaptiver Regressionsalgorithmus präsentiert. Abschließend wird die LSMC-Methodik mit Hilfe eines praktischen Beispiels illustriert.

Der zweite Teil baut auf dem ersten auf und beleuchet den Kalibrierungsschritt tiefergehend. Nach einem Literaturüberblick und methodischen Basics werden diverse adaptive Machine Learning Ansätze basierend auf der Methode der kleinsten Quadrate und Modellwahlkriterien zur Herleitung der Proxyfunktion eingeführt. Die untersuchten Regressionstechniken reichen von der gewöhnlichen und verallgemeinerten Kleinste-Quadrate-Schätzung über GLM und GAM Methoden bis hin zur MARS Routine und Kernel-Regression. Alle Verfahren werden mathematisch legitimiert sowie anhand von praktischen Experimenten hinsichtlich ihres Verhaltens, ihrer numerischen Stabilität und Out-of-Sample Performance analysiert. Neben dem LSMC-Kontext kommen für diese Verfahren auch andere hochdimensionale Anwendungen mit Variablenselektion infrage.

Lebensversicherungsverträge mit der Option zur vorzeitigen Beendigung können ebenfalls durch einen LSMC-Algorithmus bewertet werden. Im dritten Teil dieser Dissertation werden fondgebundene Verträge mit Mindestgarantiezinsen, die Rückkaufsrechte nach dem Vorbild amerikanischer Optionen enthalten, bewertet. Die Modellierung lässt Zufall und Sprünge in den Zins-, Volatilitäts-, Aktienmarkt- und Sterblichkeitsentwicklungen zu. Zur gleichzeitigen Bewertung zahlreicher Versicherungsverträge wird ein hybrides Wahrscheinlichkeitsmaß verwendet. Weiter werden eine seedbasierte Simulationsprozedur zur Kontrolle des Bias bei der Forwärtsdiskretisierung und ein Validierungskonzept vorgeschlagen. Ein numerisches Beispiel rundet den letzten Teil ab.

Contents

List of Tables

List of Figures

Introductory Part

Actuarial Modeling in Risk Management

Résumé

First of all, we give a thorough insight into the world of actuarial risk modeling, which serves as the mathematical foundation for risk management in the life insurance business. Thereby, five sub-areas are illuminated. We start with the cash-flow-projection models as the platforms for the actuarial modeling under Solvency II. Then, we move on to the stochastic modeling of selected capital market variables such as interest rates, stock indices and credit default by economic scenario generators. Thereafter, we dive deeper into the central themes for calculating capital requirements. As the first theme, the most common scenario types are described. This is followed by the modeling of shocks to selected risk factors such as interest rates, equities, bonds, mortality and lapse. Last but not least, the presented concepts are united in the outlines of some well-known risk aggregation techniques, including the least-squares Monte Carlo method which is the fundamental object of research of this thesis.

1 Cash-flow-Projection Models

1.1 Solvency II & CFP Models

Insurance companies have to derive the market values of their assets, best estimate liabilities, available capitals and other related balance sheet items under *Solvency II* by European Parliament & European Council (2009) to calculate their solvency capital requirements (SCRs). The legislative background of the Solvency II directive is EU insurance regulation aiming at policyholder protection by managing the companies' risks of insolvency. For the computations, the insurers make assumptions about the future developments on the capital market and the actuarial factors driving the movements in the product portfolios such as mortality, the insureds' lapse behavior or administration costs. Many of these assumptions are prescribed by authorities such as EIOPA.

In the life insurance business, the cash flows on the assets and liabilities sides are interdependent. One reason are profit sharing mechanisms according to which the insureds highly participate at generated surpluses while not equally experiencing losses due to their minimum interest rate guarantees (asymmetry). To account for the interdependencies, the insurers need to model the entirety of their cash flow streams in a common system. On the liabilities side, premiums are the main cash inflow source whereas benefits to policyholders, dividends to shareholders and administration costs are the main cash outflow sources. The cash flows on the assets side are mainly determined by investment decisions.

Furthermore, the insurers need to project their cash flows typically between 40 and 60 years into the future to reflect the long terms of life insurance contracts appropriately. Besides regulatory requirements, management actions should be considered for a possibly realistic modeling as well. We refer to the systems in which the complex cash flow interactions are modeled as the *cash-flow-projection (CFP) models*.

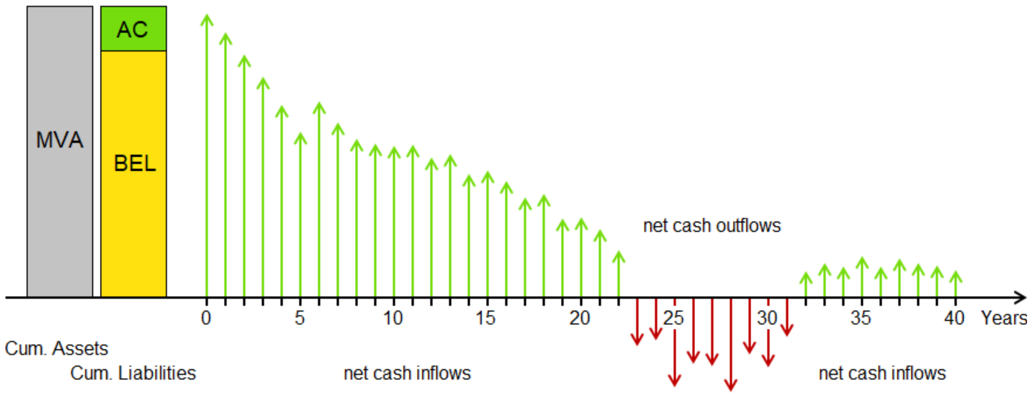

Figure 1: Discounted cumulative balance and net cash flow stream generated by a CFP model.

To compute the relevant balance sheet items, the CFP model has to be simulated. We call these calculations *Monte Carlo simulations*. Figure 1 displays the discounted net cash flow stream generated by the CFP model of an exemplary life insurer's existing portfolio over a projection horizon of 40 years. The cash flows depend on the CFP model characteristics, the *outer risk scenarios* and *inner capital market scenarios*. For more details on how the random ingredients of inner capital market scenarios can look like, see

Section 2, and for more on the different scenario types, see Section 3.

1.2 Balance Sheet Items & SCR

The Solvency II directive asks insurance companies to calculate their SCRs with the standard formula approach, a partial or a full internal model approach, see Section 5. Insurers with full internal models have to assess their available capitals AC_1^i after a one-year risk horizon and the related balance sheet items under several hundred thousand real-world scenarios i. Essentially, this means they must derive the market value of their assets MVA_1^i and best estimate liability BEL_1^i, that is, the present value of expected future cash flows to the policyholders, under each scenario. The available capital is then obtained scenario wise as the difference between the two positions, i.e.,

$$\mathrm{AC}_1^i = \mathrm{MVA}_1^i - \mathrm{BEL}_1^i. \tag{1}$$

While the BEL_1^i stem from the CFP model, the MVA_1^i are typically derived by more granular models maintained for asset management or accounting purposes. In these models, the assets can be valued by closed-formula solutions or other numerically efficient pricing algorithms. To keep the explanations in this thesis simple, we will sometimes implicitly assume that the BEL_1^i and MVA_1^i are produced by the same model.

Furthermore, let the initial available capital or *base available capital* be denoted by AC^0. Kochanski (2010) formalizes AC^0 as the value of the available capital at time $t = 0$, that is, the available capital under the *base scenario* which represents the current capital market situation and most likely actuarial model assumptions. The profit Δ^i made under scenario i is given as the difference between the discounted available capital $B_1\mathrm{AC}_1^i$ and initial available capital AC^0, i.e.,

$$\Delta^i = B_1\mathrm{AC}_1^i - \mathrm{AC}^0. \tag{2}$$

Articles 122(2) and 101(3) of the Solvency II directive define the SCR in the internal model approach as the 99.5% value-at-risk of the full loss probability distribution forecast, i.e.,

$$\mathrm{SCR} = \mathrm{VaR}_{99.5\%}\left(-\Delta\right) = \inf\left\{y \in \mathbb{R} \mid \mathbb{P}\left(-\Delta \leq y\right) \geq 99.5\%\right\}. \tag{3}$$

Here, $\mathbb{P}\left(-\Delta \leq y\right)$ denotes the cumulative distribution function of the loss $-\Delta$ under the real-world probability measure $\mathbb{P}$. By construction, having an initial available capital equal to the SCR, i.e., $\mathrm{AC}^0 = \mathrm{SCR}$, ensures that the insurance company will statistically default only with probability 0.5%. For an illustration of this case see Figure 2.

1.3 Risk Factors & Stresses

Risks are sensitivities of an insurer's business developments to adverse deviations from the calculation principles. Solvency II takes a purely economic viewpoint and focuses on the risks directly affecting the balance sheet items. This means, the risks associated with the capital market and actuarial assumptions that deteriorate the investment, actuarial and cost results need to be taken into account. An intact risk management continuously identifies and quantifies the factors determining success or failure of the business activities.

For the quantification, *risk factors*, often referred to as *risk drivers*, are introduced in the company's CFP model. Risk factors are objectifiable and measurable parameters with expectedly a significant impact on the business results. Ideal candidates for risk factors are

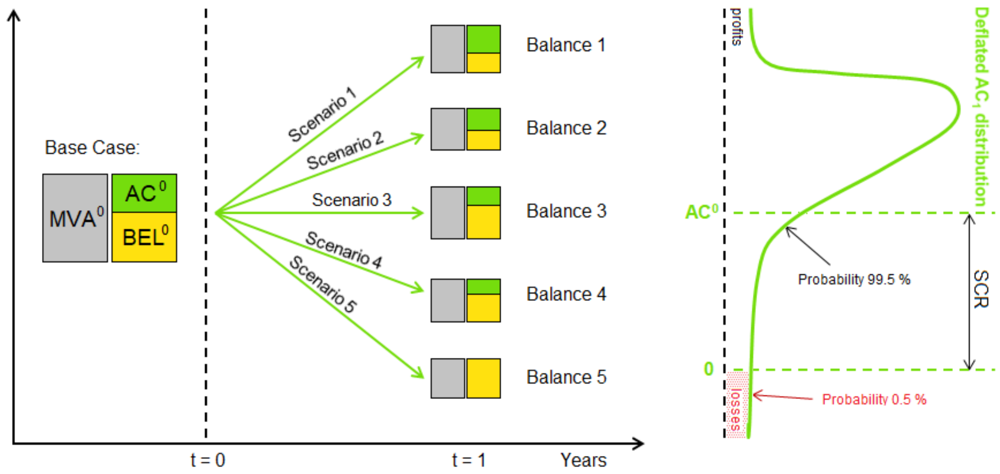

Figure 2: Real-world probability distribution forecast of $B_1\mathrm{AC}_1$ where $\mathrm{AC}^0 = \mathrm{SCR}$.

shocks to publicly accessible market quotations such as equity prices or shocks to statistically deducible parameters such as biometric variables, compare Investment Committee of DAV (2015). A distinction between *capital market* or *financial risk factors* and *actuarial risk factors* is often made. Prominent examples for capital market risk factors are shocks to the risk-free interest rates movement, equity prices or bond yields. For possible modeling approaches, see Sections 4.1–4.3. Typical actuarial risk factors are longevity, mortality or lapse shocks, for their modeling, see Sections 4.4 and 4.5. Mandatory risk factors in the derivation of an insurer's SCR are specified by Article 105 of the Solvency II directive. Barely quantifiable risks such as operational, reputation and strategic risks have to be treated separately and will not be considered here.

Based on historical data, probability distributions such as Student's t-distribution or the normal distribution are estimated for the modeling of the risk factor ranges. These distributions reflect the realistic stress levels which the risk factors can attain. To refer to a specific risk factor stress level, we simply use the term *stress* or *shock*. Stresses and shocks have to interpreted as values and can be detrimental or beneficial to a balance in most risk aggregation methods. But in the standard formula approach they are always detrimental. The term "stress" will in some parts of this thesis be employed as just another synonym for the term "risk factor". In Figure 3, the outer scenarios take on different stress values.

In summary, risks constitute sensitivities of an insurer's balance to adverse changes in the assumptions. They are modeled by risk factors, which can attain a wide range of stresses. Stresses can have a positive or negative impact on the business results.

1.4 Inner & Outer Scenarios

An *inner scenario* refers to a path of the capital market variables over the projection horizon. It can be defined as a two-dimensional matrix with the capital market variables in the rows and the projection times in the columns. Examples for capital market variables are prices of zero-coupon bonds with different terms to maturity, exchange rates, stock returns or property returns. Usually, many of these variables are also risk factors. The evolutions of the capital market variables highly depend on the model assumptions, that is, the outer scenarios, see the next paragraph. While an inner scenario belongs to the input of the CFP model, a corresponding *inner Monte Carlo simulation* describes an actual path of the CFP model in which all modeled economic interdependencies are processed. It is common to further specify inner scenarios conditional on the context. In Section 3.1, we will distinguish *stochastic scenarios* from *deterministic scenarios*, and in Section 3.2, we will dive deeper into *risk-neutral scenarios*. Later on in this thesis, we will use capital market, stochastic and risk-neutral scenarios interchangeably.

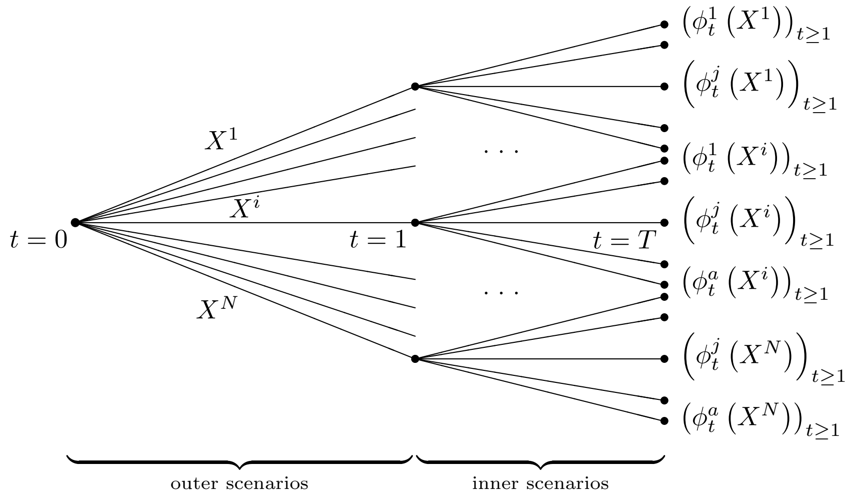

Figure 3: Outer and inner scenarios over the projection horizon.

By an *outer scenario*, we model the risk an insurer is exposed to in the first projection year. Mathematically, we define an outer scenario as a vector of which each component stands for a certain capital market or actuarial risk factor. Some of the capital market risk factors affect the inner scenario paths described above. The other ones and the actuarial risk factors take effect on other assumptions made in the CFP model. Typically, setting all components equal to zero yields the stress-neutral *base outer scenario* and thus an insurer's actual expectations about the model. Stresses unequal to zero are either positive or negative deviations from these expectations. In Sections 3.3–3.5, we will divide outer scenarios into *real-world scenarios*, *fitting scenarios* and *validation scenarios*.

In each Monte Carlo simulation of the CFP model, an outer scenario is complemented by an inner scenario. Figure 3 shows the relationship between N outer scenarios X^i with respectively a inner scenarios $\left(\phi_t^j\left(X^i\right)\right)_{t\geq1}$ from projection start $t=0$ until projection end $t=T$. While the purpose of inner scenarios is to value balance sheet items conditional on an outer scenario assumption, the purpose of an outer scenario is to obtain a prediction of the balance for a specific risk constellation.

2 Stochastic Modeling

2.1 Capital Market Scenarios & ESG

When using the term *capital market scenarios*, the focus lies on the economic meaning of inner scenarios. Capital market scenarios are either deterministic or stochastic. The scenarios are supposed to model the capital market evolutions over the projection horizon according to some industry-wide standards as set out by Investment Committee of DAV (2017) for Germany. Since this is a non-trivial requirement for insurance companies with complex business profiles, software solutions taking over the scenario generation task have been developed. These software solutions are called *economic scenario generators (ESGs)*.

ESGs need to be tailored to an insurer's business characteristics. This means, an ESG has to model the insurer's asset classes and risk factors neither in a too simple nor in a too complex way. For instance, only asset classes the insurer is invested in must be considered and the more important a class is in the portfolio, the preciser it should be modeled. Similarly, areas with higher risk exposures should be covered at higher levels of detail whereas low-risk areas should be captured by models that are easier to understand. For the modeling of the ingredients of stochastic scenarios in ESGs, see representatively Sections 2.2–2.4 for term structures of interest rates, stock indices and credit default. For more information on how property, commodities, foreign currencies, inflation and volatilities can be modeled, see the aforementioned source.

The firms can choose between standard and company-specific ESGs. When it comes to the generation of market consistent risk-neutral scenarios, ESGs include the determination of targets, a calibration procedure, the actual scenario generation process and a validation procedure. The targets comprise current market conditions such as, for example, the risk-free yield curve, swaption, bond, stock and option prices or implicit volatilities, and dependencies between these quantities. While the former targets are directly observable in the market or can be derived from it, the dependencies need to be estimated based on historical data. Figure 4 by the European Central Bank (2020) displays two government bond yield curves which can be decomposed into the risk-free yield curve plus a credit spread. In the calibration procedure, the model parameters of the ESG are fit to the targets to achieve market consistency. The scenarios are then generated conditional on these model parameters. If the validation indicates that the obtained scenarios are market consistent and arbitrage-free, they can be fed into the CFP model.

2.2 Interest Rate Modeling

In this section, we describe widely spread modeling approaches for the *term structures of interest rates* in ESGs according to Chapter 3.2.5.3 of Investment Committee of DAV (2017). Typically, these term structures are modeled by *short-rate models* or the *LIBOR market model*. Despite we present only one-factor models, multi-factor models are conceivable as well. For a proper modeling of negative interest rate environments such as the one from January 2020 illustrated in Figure 4, the possibility of negative interest rates should be given.

The *short rate* r_t denotes the instantaneous interest rate at time t at which money can be borrowed for an infinitesimally short period of time. If r_t is modeled as a stochastic process under a risk-neutral probability measure $\mathbb{Q}$, the price at the start year $t = 0$ of a zero-coupon bond with term T to maturity and a payoff of 1 is given by

$$P(0,T) = E^{\mathbb{Q}}\left[\exp\left(-\int_0^T r_s \mathrm{d}s\right)\right], \tag{4}$$

see e.g. Korn et al. (2010)[p. 276]. The interest rates or yields which correspond to these T-dependent zero-coupon bond prices are called *spot rates* $i(0,T)$. They follow from

$$P(0,T) = \frac{1}{(1 + i(0,T))^T}. \tag{5}$$

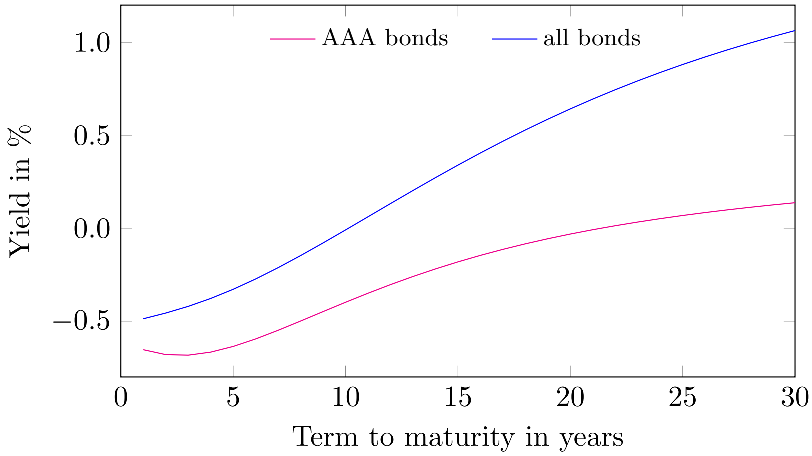

Figure 4: AAA-rated and overall Euro area government bond yield curves as of 31st of January 2020.

They form the risk-free yield curve with respect to T at the start year. Future zero-coupon bond prices are implicitly defined by this yield curve as well so that a short-rate model not only specifies the spot rates but also the *forward rates*. In this way, the zero-coupon yield curves for all projection years $t > 0$ are specified by exactly the same short-rate model. Short rates are not observable in the market.

The following popular short-rate models by Vasicek (1977), Cox et al. (1985) and Black & Karasinski (1991) are mean-reverting towards the long-run average value δ_r with speed of adjustment ζ_r, volatility σ_r, and standard Brownian motion Z^r. Thereby, the variables can be set to be constant, to vary deterministically over time, or they can be modeled as stochastic processes. Here are the definitions of the short-rate models:

- Vasicek model: $\mathrm{d}r_t = \zeta_r\,(\delta_r - r_t)\,\mathrm{d}t + \sigma_r \mathrm{d}Z_t^r$,
- Cox-Ingersoll-Ross (CIR) model: $\mathrm{d}r_t = \zeta_r\,(\delta_r - r_t)\,\mathrm{d}t + \sigma_r\sqrt{r_t}\mathrm{d}Z_t^r$,
- Black-Karasinski (BK) model: $\mathrm{d}\log r_t = \zeta_t^r\,(\delta_t^r - \log r_t)\,\mathrm{d}t + \sigma_t^r \mathrm{d}Z_t^r$.

The short rates of the Vasicek model are normally distributed and permit thus negative interest rates, the ones of the CIR model are chi-squared and ensure nonnegative interest rates, and the ones of the BK model are log-normally distributed and always positive. For an application of the CIR model, see also Section 17.3.

The LIBOR market model by Brace et al. (1997), also known as the Brace-Gątarek-Musiela (BGM) model, describes the evolutions of the LIBOR forward rates L_t^l on time intervals $[l, l+1]\,,\ l = 1, \ldots, L$:

- LIBOR market model (LMM): $\frac{\mathrm{d}L_t^l}{L_t^l} = \mu_t^l \mathrm{d}t + \sigma_t^l \mathrm{d}Z_t^l$ with drift term μ_t^l.

Here, the modeled quantities are directly observable in the market (forward LIBORs) and the model can be calibrated such that it matches the risk-free yield curve at the start year. The LIBOR forward rates are log-normally distributed and always positive.

2.3 Modeling of Stock Indices

Now we address the modeling of *stock indices* in ESGs according to Chapter 3.2.5.2 of Investment Committee of DAV (2017). Since the prices of single stocks can be designed in the same way, the following holds for them as well. Stock indices are modeled by the changes of their values over time, that is, by stochastic processes of their returns. Both discrete and logarithmic/continuously compounded returns are conceivable. The advantage of logarithmic returns is their time-additivity, meaning that the overall logarithmic return over two successive time periods is equal to the sum of the two individual logarithmic returns, see e.g. Hudson & Gregoriou (2015). Furthermore, positive and negative logarithmic returns of the same magnitude cancel each other out.

Insurance companies barely exposed to the risk of changes in dividend cash flows only need to model the total returns of the performance indices. If, however, dividend cash flows are a risk factor, a more granular approach has to be taken. Either the dividend yields or the returns of the pure price indices must be modeled in addition. For stochastic dividend yields, mean-reverting processes are recommended. Regardless of whether a performance or price index is set up, the same type of stochastic process can be chosen, the parameters resulting from calibration are then just different. An adequate stochastic process has to account for stylized facts observed on the capital market that the insurer's balance sheet is sensitive to. Important stylized facts of returns might be left skewed distributions with fat tails or stochastic volatilities.

The Committee on Finance Research of Society of Actuaries (2016) proposes in Chapter 13 the models by Black & Scholes (1973), Heston (1993) and Bates (1996). Let S_t denote the index price level so that $\frac{\mathrm{d}S_t}{S_t}$ corresponds to the discrete return. Moreover, let r_t be the short rate, μ_t the stochastic drift term, K_t the stochastic volatility, and let the constant parameters $\lambda_S, \sigma_S, \zeta_K, \delta_K$ be given. Let Z^S and Z^K be standard Brownian motions and let J^S be a jump process such as of the compound Poisson type. Suitable index models are then of the following forms:

- Black-Scholes model: $\frac{\mathrm{d}S_t}{S_t} = (r_t + \lambda_S \sigma_S)\,\mathrm{d}t + \sigma_S \mathrm{d}Z_t^S$, often written with $\mu_t = r_t + \lambda_S \sigma_S$,
- Heston model: $\frac{\mathrm{d}S_t}{S_t} = \mu_t \mathrm{d}t + \sqrt{K_t}\mathrm{d}Z_t^S$ and $\mathrm{d}K_t = \zeta_K(\delta_K - K_t)\,\mathrm{d}t + \sigma_K\sqrt{K_t}\mathrm{d}Z_t^K$,
- Bates model: $\frac{\mathrm{d}S_t}{S_t} = \mu_t \mathrm{d}t + \sqrt{K_t}\mathrm{d}Z_t^S + \mathrm{d}J_t^S$ and $\mathrm{d}K_t = \zeta_K(\delta_K - K_t)\,\mathrm{d}t + \sigma_K\sqrt{K_t}\mathrm{d}Z_t^K$.

In the Black-Scholes model, the instantaneous returns are normally distributed and have a constant volatility. Differently, the Heston and Bates models embed fat tails and stochastic volatilities as CIR models. Since the SCR is determined as the 99.5% value-at-risk of the loss distribution under Solvency II, compare (3), it is crucial for a risk analysis to model the tails of any potentially relevant distribution properly. Hence, if fat tails and stochastic volatilities are distinct features of the index returns, the Heston and Bates models should be favored. An example for an extended Bates model proposed by Bakshi et al. (1997) and implemented by Bacinello et al. (2009) can be found in Section 17.3. Further alternatives are (G)ARCH models, mixed Markov models, Lévy processes and other extensions of the presented models.

When transitioning from discrete to logarithmic returns, small changes in the dynamics of the models not affecting their general practicability have to be made, see e.g. Drăgulescu & Yakovenko (2002).

2.4 Modeling of Credit Default

Credit default refers in CFP models to the event that a credit borrower does not repay the loan which the insurance company has granted him in parts or in full. Since government bonds are most often modeled as risk-free investments, corporate bonds can typically be considered an insurer's main credit default risk. By following Chapter 3.2.5.4 of Investment Committee of DAV (2017), we limit ourselves here to credit defaults in ESGs that concern corporate bonds. The easiest way to factor in corporate bonds is to proceed with them like with government bonds, that is, to omit their risk of a default. Then, only their coupon payments need to be adjusted such that the risk-neutral market values at projection start hit the real-world ones. But only as long as the insurer's asset portfolio comprises nearly solely corporate bonds with a very high creditworthiness and the projection horizon is comparatively short, such an approach is justifiable. Compare the developments associated with the financial crisis of 2007-08.

A slightly more advanced modeling approach decomposes a corporate bond into risk-free coupon payments, a government bond, a stock and a short call. The four components are then calibrated such that their total market value at projection start equals the real-world one of the corporate bond and their total final payment attains at maximum the nominal value of the corporate bond. In addition, market consistency is required for the government bond and the stock. Conditional on the capital market scenario, the corporate bond can now default. Such an event occurs in scenarios in which the stock value increases significantly, which makes the buyer of the call option execute her right to receive the stock from the insurance company at the comparatively low pre-agreed strike price. However, this approach lacks the modeling of credit ratings, migrations thereof and reinvestments in new corporate bonds.

The events in connection with the financial crisis of 2007-08 have shown the high relevance of an explicit modeling of these features. A popular approach incorporating them is the Jarrow-Lando-Turnbull (JLT) model introduced by Jarrow et al. (1997). In this model, the credit rating dynamics are characterized by a Markov chain, meaning that future credit ratings only depend on current states but not past ones. The probabilities of the migrations between the different states are summarized by a real-world transition matrix. For an exemplary transition matrix, where element q_{ij} denotes the transitional probability from state i to j, see Table 1. To achieve risk-neutrality, all transitional prob-

	AAA	**AA**	**A**	**BBB**	**BB**	**B**	**CCC**	**D**
AAA	0.8899	0.0984	0.0071	0.0006	0.0019	0.0009	0.0011	0.0001
AA	0.0006	0.9055	0.0842	0.0075	0.0007	0.0003	0.0010	0.0002
A	0.0004	0.0114	0.9091	0.0700	0.0034	0.0028	0.0010	0.0019
BBB	0.0001	0.0019	0.0496	0.8747	0.0635	0.0049	0.0021	0.0032
BB	0.0001	0.0010	0.0017	0.0634	0.8328	0.0835	0.0070	0.0105
B	0.0002	0.0011	0.0013	0.0031	0.0578	0.8367	0.0545	0.0453
CCC	0.0001	0.0001	0.0020	0.0030	0.0063	0.1403	0.5685	0.2797
D	0	0	0	0	0	0	0	1

Table 1: Transition matrix with credit rating migrations between AAA and D (default).

abilities are scaled individually and time wise by risk premium adjustments conditional on each capital market scenario. Once, the Markov chain hits the default state, it stays there until projection end but still repays the insurance company a given fraction (i.e., the recovery rate) of its loan. Through a correlation approach, the occurrence of default can be coupled to stock market performance. Reinvestments in new corporate bonds are also implementable in the JLT model.

3 Scenario Types

3.1 Deterministic & Stochastic Scenarios

Inner scenarios can be divided into *deterministic scenarios* and *stochastic scenarios*. In dependence of the risk projection and aggregation technique, the one or the other type is applied, see Section 5. Investment Committee of DAV (2015) states that deterministic scenarios are capital market scenario paths relying on historical data or expert judgment that are used in what-if analyses such as impact assessments of low interest rate environments or certainty equivalent (CE) calculations in the market consistent embedded value (MCEV) context. Per outer scenario, a deterministic scenario is uniquely determined, in Figure 3 this means $a = 1$. For instance, the inner scenarios associated with the stresses from the Solvency II standard formula or the Swiss Solvency test are deterministic scenarios. In CE scenarios, all asset classes earn the risk-free interest rate proxy and reflect the market conditions at the valuation date.

In Parts I, II and III of this thesis, we will only consider stochastic scenarios. These kinds of scenarios are generated based on the stochastic processes modeled in the ESG, compare Section 2. Due to the randomness, there exist infinitely many stochastic capital market scenarios per outer scenario, in Figure 3 this means $a > 1$. An example of a set of stochastically simulated yield curves at projection year $t = 1$ is depicted in Figure 5 and a set of stochastically simulated total return values is depicted in Figure 6.

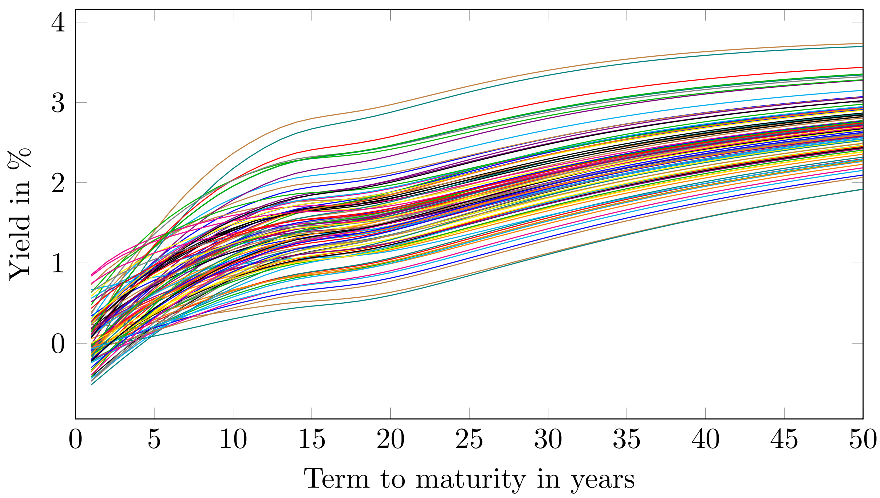

Figure 5: Set of stochastically simulated zero-coupon bond yield curves at projection year $t = 1$.

The crucial advantage of stochastic simulations coming as a full set is their ability to deal with uncertainty by defining various Monte Carlo paths in the CFP model. The downside of full stochastic simulations is their high computational costs. Typical application fields are the SCR calculations based on internal model approaches under Solvency II, asset liability management (ALM) analyses, strategic asset allocation (SAA) examinations and also MCEV calculations.

3.2 Risk-neutral Scenarios

Risk-neutral scenarios are stochastic scenarios with the property that each security earns on average the risk-free interest rate proxy so that there are no risk premiums regardless of the investment strategy. Under a *risk-neutral probability measure*, any security can be priced as the expectation of its discounted cash flows. A risk-neutral measure is therefore a *pricing measure*, and the purpose of risk-neutral scenarios drawn from it is the valuation of liabilities, assets or other balance sheet items by simulation. Risk-neutral scenarios must always be understood as full sets of inner scenarios. The Monte Carlo paths illustrated in Figures 5 and 6 belong to sets of risk-neutral scenarios. The capital market variables and their correlations are required to be market consistent for fair valuations. To achieve market consistency, the model parameters of the underlying stochastic processes are calibrated to appropriate targets in the ESG, compare Section 2.1.

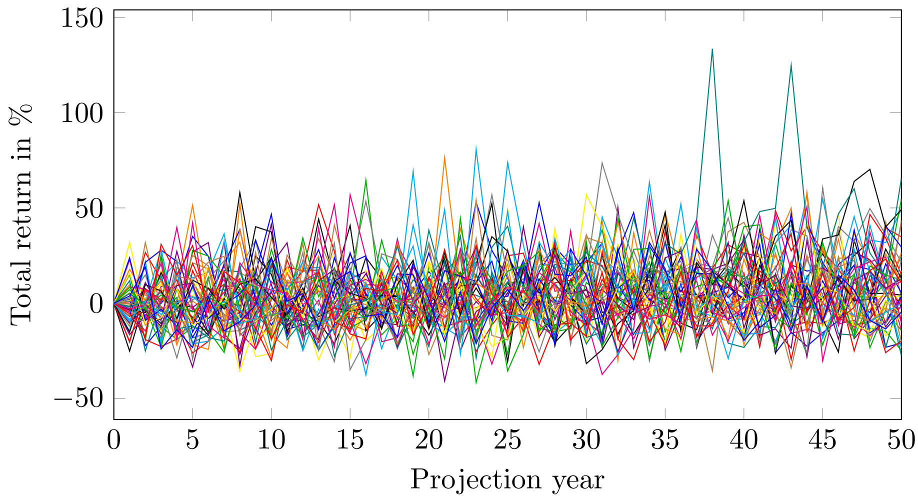

Figure 6: Set of stochastically simulated logarithmic total returns of a performance index.

An application involving both CE and risk-neutral scenarios arises in the MCEV context when calculating the time value of options and guarantees (TVOG). Since the policyholders participate at the profits whereas they do not incur any losses in Germany, the insurers' cash flow profiles are asymmetric so that the risk-neutral simulations yield on average a lower present value of future profits (PVFP) than the CE simulation. The TVOG captures this gap, i.e.,

$$\mathrm{TVOG} = \mathrm{PVFP}_{\mathrm{CE}} - \mathrm{PVFP}_{\mathrm{stoch.}}, \tag{6}$$

see e.g. Gürtler (2011). An exemplary life insurance contract featuring minimum interest rate guarantees and a surrender option can be found in Part III of this thesis.

3.3 Real-world Scenarios

In the following, *real-world scenarios* are outer scenarios, under which an insurer's liabilities, assets or other balance sheet items shall be projected into the future. The projection has the purpose of assessing the insurer's financial situation and deriving concrete business actions from it. A prominent example for such an assessment is the real-world loss probability distribution forecast and the SCR calculation by the internal model approach under Solvency II, compare Section 1.2. Like stochastic scenarios, real-world scenarios are able to cope with uncertainty when coming in full sets. But differently, a stand-alone real-world scenario can be interpreted reasonably (given a sufficient inner scenario valuation). Real-world scenarios are drawn from a *physical probability measure*, which is derived with the aid of historical data and expert judgment. Copulas are a way to model physical probability measures, see e.g. Mai & Scherer (2012). Capturing historical patterns of the risk factors including their correlations lies in the focus here. The challenge consists of estimating the parameters of the physical probability measure representing the joint real-world distribution of the insurer's risk factors realistically. A random set of real-world scenarios (given sufficient inner scenario valuations) lets risky assets reach on average an excess return over the risk-free interest rate and allows thus risk premiums.

In the CFP model, the physical and risk-neutral probability measures complement each other over two disjoint time intervals as can be seen in Figure 3 by the sequence of the outer and inner scenario realizations. The unification of the two measures yields the *hybrid probability measure* introduced by Bauer & Ha (2015) and Natolski & Werner (2016). This measure justifies the flexibility to switch between projection and pricing within one Monte Carlo path.

3.4 Fitting Scenarios

To transition from computationally expensive CFP models to cheap proxy models, *fitting scenarios* are required. These kinds of scenarios are selected outer scenarios under which a CFP model is simulated to obtain a representative data image. The generated *fitting points* serve as regression data for the derivation of the proxy model. For this derivation, suitable machine learning algorithms such as the least-squares Monte Carlo (LSMC) method can be applied. Appropriately, fitting scenarios are also referred to as *training* or *calibration scenarios*. The set of fitting scenarios has to cover a wide range of risk factor stress level combinations densely while being allowed to be complemented by only few inner scenarios in smart LSMC designs. In Figure 3, such a set-up corresponds to an allocation of, for instance, $N = 25,000$ outer with $a = 2$ inner scenarios.

Since the proxy model will be evaluated at the real-world scenarios, the fitting scenarios have to cover their space sufficiently well. We call the space on which the fitting scenarios are defined *fitting space*. Let a specific risk factor be given and let $q_{0.1\%}$ and $q_{99.9\%}$ denote the 0.1%- and 99.9%-quantiles of its univariate real-world distribution. If we aim to calculate the SCR as the 99.5% value-at-risk of the insurer's full loss distribution forecast, compare Equation (3), reasonable values for the fitting scenario component representing the given risk factor should range, for example, from $q_{0.1\%}$ to $q_{99.9\%}$. For a concrete example, see Section 4.2.

The fitting scenarios are not required to be drawn from the physical probability measure but can instead be drawn from a purely technical one. Common choices for such technical measures are low-discrepancy sequences such as quasi-random Sobol sequences. Not only do these sequences ensure optimal usage of the scenario budget in Monte Carlo simulations but also they are easy to implement, for details see Niederreiter (1992).

3.5 Validation Scenarios

For a reliable transitioning from CFP models to proxy models, *validation scenarios* are required in addition to the fitting scenarios. These kinds of scenarios are selected outer scenarios under which a CFP model is simulated to obtain a relevant data image. The generated *validation points* are used to assess the absolute goodness of fit of the previously calibrated proxy model. In smart LSMC regression set-ups, a set of validation scenarios aims at quantifying the expectations of the economic variable with respect to a few informative risk factor stress level combinations accurately. Such a set-up entails an allocation of, for instance, $N = 50$ outer with $a = 1,000$ inner scenarios.

Principally, *in-sample* and *out-of-sample validation scenarios* can be chosen. Since in-sample validation scenarios are technically not distinguishable from fitting scenarios, they are not suitable for measuring the absolute goodness of fit of a proxy model in a risk-neutral world. Therefore, we will only consider out-of-sample validation scenarios and refer to them as *the* validation scenarios. Due to the many required inner scenarios for the high accuracy, only few validation scenario specifications are within the scenario budget. For the SCR calculation under Solvency II, for example, the following paradigms exist:

- Points known to be in the *capital region*, that is, scenarios producing a risk capital close to the SCR estimate from previous risk capital calculations;
- Quasi-random points from the entire fitting space;
- One-dimensional risks leading to a 1-in-200 loss in the one-dimensional distribution;
- Two- or three-dimensional stresses for risk factors with high interdependency.

4 Shock Modeling

4.1 Shocks to Interest Rates

Article 105(5)a of the Solvency II directive defines the *interest rate risk* as

> "the sensitivity of the values of assets, liabilities and financial instruments to changes in the term structure of interest rates, or in the volatility of interest rates."

Shocks to interest rates concern the discounting of the cash flows with respect to the risk-free yield curve in the CFP model. The ranges of the shocks are determined based on historical data, which, by the way, applies to the ranges of any risk factors. The shocks to the term structure of interest rates are required as input information for the simulations of the short-rate model or LIBOR market model in the ESG, compare Section 2.2. This means that a specific stressed risk-free yield curve cannot be obtained by simply applying the shock ex post to an already existing risk-free yield curve such as the base yield curve.

While in one-factor models the shock to the pure interest rates movement is one-dimensional, in multi-factor models it is multidimensional. For instance, a shock to a three-factor model has been observed in practice to allow under certain conditions a component wise interpretation as follows: The first shock component represents a parallel shift to the base yield curve, the second component a change to the slope of that curve, and the third component a change to its curvature. Shocks to interest rate volatility can be modeled separately. For most currencies, the stressed risk-free yield curves with and without volatility adjustment (VA) have to be provided. The volatility adjustment itself can also be stressed by treating it as an additional risk factor.

4.2 Shocks to Equities

The *equity risk* is defined by Article 105(5)b of the Solvency II directive as

> "the sensitivity of the values of assets, liabilities and financial instruments to changes in the level or in the volatility of market prices of equities."

Shocks to the level of equities are modeled as percentage changes in the market values of the equities. The simplest approach is to tackle all equities of an insurance company jointly by applying the shocks to their total market value. Independently, shocks to equity volatility can be modeled. As opposed to shocks to term structures of interest rates and volatility shocks, equity level shocks are not required as input information for the simulations in the ESG as returns reflect only changes to indices.

Let us give a numerical example for the equity level stress corresponding to the one-dimensional 1-in-200 loss. At first, the real-world shocks need to be estimated based on historical data. Let these be found to follow Student's t-distribution with four degrees of freedom and scale parameter 0.2126. If S_t indicates the equity index level at time t and if the equity returns are logarithmic, compare Figure 6, the p%-quantile is given as $q_{p\%} = \mathrm{d}\log S = \log S_1 - \log S_0 = \log \frac{S_1}{S_0}$. Rewriting this equation yields $X_{p\%} = \frac{S_1}{S_0} - 1 = \exp q_{p\%} - 1$ for the percentage change in equity market price. The shock leading to the 1-in-200 loss in the univariate equity distribution corresponds to the 0.5%-quantile of the estimated t-distribution. This quantile is equal to $q_{0.5\%} = -0.9789$ here. The percentage change in equity market price associated with the 1-in-200 loss is thus

$$X_{0.5\%} = \exp q_{0.5\%} - 1 = \exp(-0.9789) - 1 = -62.43\%.$$

In Section 3.4, a reasonable fitting scenario range under Solvency II is said to correspond to the interval $[q_{0.1\%}, q_{99.9\%}]$. By using $q_{0.1\%} = -1.5252$ and $q_{99.9\%} = 1.5252$, the minimum and maximum percentage changes occurring in the fitting scenarios turn out to be

$$\begin{aligned} X_{0.1\%} &= \exp q_{0.1\%} - 1 = \exp(-1.5252) - 1 = -78.24\%, \\ X_{99.9\%} &= \exp q_{99.9\%} - 1 = \exp(1.5252) - 1 = 359.60\%. \end{aligned}$$

According to Chapter 5.1.2 of EIOPA (2019), the shocks can also be provided per country or geographical area (e.g., EU, US, ASIA) where the equity is listed. If an equity is listed in transnational stock exchanges, either the weighted average of the shocks across the involved countries or the shock of the country where the largest portion of the equity is listed shall be applied. For instance, DAX shall be shocked with respect to the German stress and EURO STOXX 50 with respect to the weighted average across the involved European countries.

4.3 Shocks to Bonds

Article 105(5)d of the Solvency II directive defines the *spread risk* as

> "the sensitivity of the values of assets, liabilities and financial instruments to changes in the level or in the volatility of credit spreads over the risk-free interest rate term structure."

Bonds are assets which produce fixed cash flows until they mature. Their average returns are described by yield curves such as the ones in Figure 4. A yield curve of a risky bond can be decomposed into the risk-free yield curve plus a credit spread on top. Typically, the credit spread is positive to compensate the investors for taking the risk that the bond defaults. The higher the risk of default, the higher is the credit spread, which is why the overall bond yield curve runs above the AAA-rated bond yield curve in the figure. Shocks to bonds per term to maturity can be modeled in the three equivalent ways: by a change in credit spreads, a change in yield, or a change in market prices.

Let the possibly stressed risk-free yield curve be fixed. If the levels of the credit spreads are reduced by a shock, the yield curve is shifted downwards, which represents a decrease of yields. A decrease of yields relative to the risk-free yields lowers the attractiveness of bonds as investments and could therefore also be modeled as a drop in bond market prices. Simultaneously, this means that if the risk-free yield curve is stressed, the yields of the bonds will change relative to the risk-free yields, so that a shock to the term structure of interest rates will always be accompanied by a shock to the bonds. For more clarity, let the risk-free yield curve be shifted upwards. Then the bonds earn less relative to the risk-free investment and as a result their market prices drop. As already explained, equivalently, this drop in market prices could be modeled as a downward shift of the bond yield curve or a decline in credit spreads.

Shocks to government bonds and corporate bonds can be implemented separately. According to Chapter 5.1.1 of EIOPA (2019), shocks to government bonds can be grouped in addition by credit rating, maturity, country and geographical area depending on the granularity. While shocks to government bonds with selected maturities are provided, the ones with missing maturities have to be interpolated by, for example, cubic splines or extrapolated constantly by the lastly provided shock. Shocks to corporate bonds can be clustered in addition by sector, credit rating and geographical area, and distinguished as financial/non-financial (ESA 2010 definition). A grouping by sector can thereby account for varying credit spread volatilities across sectors.

4.4 Shocks to Longevity & Mortality

The *mortality/longevity risks* are defined by Articles 105(3)a/b of the Solvency II directive as

> "the risk of loss, or of adverse change in the value of insurance liabilities, resulting from changes in the level, trend, or volatility of mortality rates, where an increase/decrease in the mortality rate leads to an increase in the value of insurance liabilities."

While the mortality risk is related to adverse business developments caused by a decrease in policyholders' life expectancy, the longevity risk is related to adverse developments resulting from an increase in life expectancy. Whether a specific shock to mortality rates has a detrimental or beneficial impact on the balance, depends on the characteristics of the insurer's product portfolio. For instance, life insurers with annuity business as their largest portion profit from a rise in mortality whereas they have to make additional payments to the insureds if mortality declines. Differently, insurers focusing on whole life insurance policies benefit from lower mortality and face more payments in case mortality goes up. Since a rise in payments stands for an increase in insurance liabilities not affecting the assets side, it entails a decrease in profits as can be seen in Equation (1).

The mortality/longevity risks must be calibrated to the company's individual demands based on historical data. A suitable risk factor design in terms of level, trend, and volatility stresses on the parameters of the insurer's mortality model has to be found. In the industry, the Lee-Carter model by Lee & Carter (1992) has become popular. An alternative accounting for cohort effects is the Cairns-Blake-Dowd model by Cairns et al. (2006). According to Chapter 5.2.1.1 of EIOPA (2019), the most frequently used mortality models have a random component for the modeling of the number of deaths, a deterministic component depending on age, calendar year and year of birth, and a link function relating the aforementioned two components to each other.

Shocks to mortality rates are typically applied directly to the best estimate mortality assumptions under which the best estimate liability (BEL) is derived. Currently, it is common practice to implement single mortality/longevity stresses that apply to all mortality rates. But EIOPA (2019) argue that a more granular approach would be needed to reflect the composition of BEL more appropriately. Shocks per age, gender, product type, socioeconomic factors (e.g., job, wealth) or geographical area are conceivable.

4.5 Shocks to Lapse

Article 105(3)f of the Solvency II directive defines the *lapse risk* as

> "the risk of loss, or of adverse change in the value of insurance liabilities, resulting from changes in the level or volatility of the rates of policy lapses, terminations, renewals and surrenders."

Wherever policyholders have the option to adapt the current conditions of their insurance contracts, the lapse risk comes into play. Articles 1(15) and 1(16) of the Delegated Regulation by European Commission (2014) divide these options into discontinuity and continuity options. The discontinuity options include all rights to fully or partly terminate, surrender, decrease, restrict, or suspend the insurance cover, or to permit the contract to lapse, whereas the continuity options include all rights to establish, renew, increase, extend or resume the insurance cover.

There are two different design possibilities for the modeling of lapse shocks according to Chapter 5.2.1.2 of EIOPA (2019). The idea of the first one is to model the shocks as instantaneous lapse events at projection start in the form of abrupt rises of lapses. In this case, the affected policies and levels of the lapse rates must be specified. The second possibility consists of a permanent change in lapse rates. This kind of shock is usually applied directly to the best estimate lapse assumptions by level and volatility adjustments but can also be related to the surrender benefits. Generally, the stress parameters can be modeled in dependence of policy features such as product type, financial guarantees, or lapse penalties. The ranges of the shocks are currently determined based on expert judgment because of a lack of market data.

The primary impact of a certain lapse stress strongly depends on the risk factor design and the involved contracts. For instance, surrender benefits falling due for payment instantaneously because of a massive lapse event require the immediate liquidation of, for instance, cash and government bond positions on the assets side and come along with a decrease in insurance liabilities. Differently, an instantaneous lapse event without suddenly due surrender benefits as well as a permanent increase in lapse rates do not affect the assets side at projection start and cause the insurance liabilities to rise or fall conditional on the product specifics. Moreover, shocks to lapse have secondary impacts during projection that are subject to conditions such as:

- product features like minimum interest rate guarantees;
- the capital market situation like equity performance or the risk-free yield curve;
- characteristics of the company's CFP model like management actions or dynamic policyholder modeling.

A related application, where equity-linked insurance contracts with American-style surrender options and minimum interest rate guarantees are priced, can be found in Part III of this thesis. Here, the surrender decision is modeled as time-dependent and with respect to, for example, equity returns, the short rate, and minimum interest rate guarantees.

5 Risk Aggregation

5.1 Nested Simulations

General Remarks

As already stated in Section 1.1, the aim of Solvency II is to protect the insureds by demanding from the insurance companies that their risk of insolvency within a one-year time period falls below the threshold of 0.5%. The companies satisfy this condition by keeping an available capital as a reserve that is at least equal to the SCR. In the internal model approach, the SCR is defined as the 99.5% value-at-risk of the full loss probability distribution forecast over a one-year risk horizon, compare Section 1.2. This 99.5% value-at-risk is the risk measure for a company's overall risk position here. For the quantification of an overall risk position, all single risks a company is exposed to, compare Section 1.3, must be aggregated.

The *nested simulations approach* proposed by Bauer et al. (2012) to calculate SCRs and used by e.g. Gordy & Juneja (2010) to derive risk measures of derivative portfolios is a risk aggregation method which is intuitively easy to understand. Such a simulation-based approach is an obvious solution to valuation problems to which no closed-formula solutions are known. On the way to calculating an insurer's SCR, pricing BELs belongs to these kinds of problems. The simulation tools in which the BEL cash flows are modeled are the CFP models. If a life insurer's cash flow profile is asymmetric, for example, because a profit sharing mechanism is in place, the BEL cash flows need to be simulated stochastically. Thereby, the randomness typically comes in by the stochastic modeling of (at least) the capital market risk factors, compare Section 2. Under a risk-neutral probability measure, any balance sheet item at time $t = 0$ can be valued by taking the expectation of its discounted future cash flows. Similarly, at time $t = 1$, any balance sheet item can be valued by taking the expectation of its discounted future cash flows conditional on the development in the first projection year, compare Section 1.4.

Simulation Setting

In the nested simulations approach, the full loss probability distribution forecast and SCR, see Equation (3), are generated by Monte Carlo simulation as follows. To obtain the probability distribution forecast, each real-world profit, see Equation (2), needs to be derived. Since each real-world profit can be decomposed into the discounted real-world available capital $\mathrm{AC}_1^i,\ i=1,\dots,R$, after the one-year time period and the base available capital AC_0, the derivation requires the simulation of R real-world scenarios $X^i,\ i=1,\dots,R$, and of the base scenario X^0. Due to the stochastic modeling, each outer scenario X^i is complemented by a inner scenarios $\left(\phi_t^j\left(X^i\right)\right)_{t\geq t_0},\ j=1,\dots,a$, so that in total $(R+1)a$ simulations are performed (e.g., $R=2^{17}=131,072$ and $a=1,000$). In each of these risk-neutral simulations, all cash flows are projected into the future to compute the insurer's MVA, BEL, and resulting available capital, see Equation (1).

Full Distribution Forecast

The left panel of Figure 7 shows the scenario split for the valuation at time $t=0$. As there is no risk in the model at projection start, the base scenario X^0 reflecting the current market conditions is unique. The a risk-neutral scenarios branch out at $t=0$ and yield a Monte Carlo paths over the projection horizon. The average values of an economic variable such as MVA, BEL, or the available capital over these paths finally provide estimates for the expected values at $t=0$.

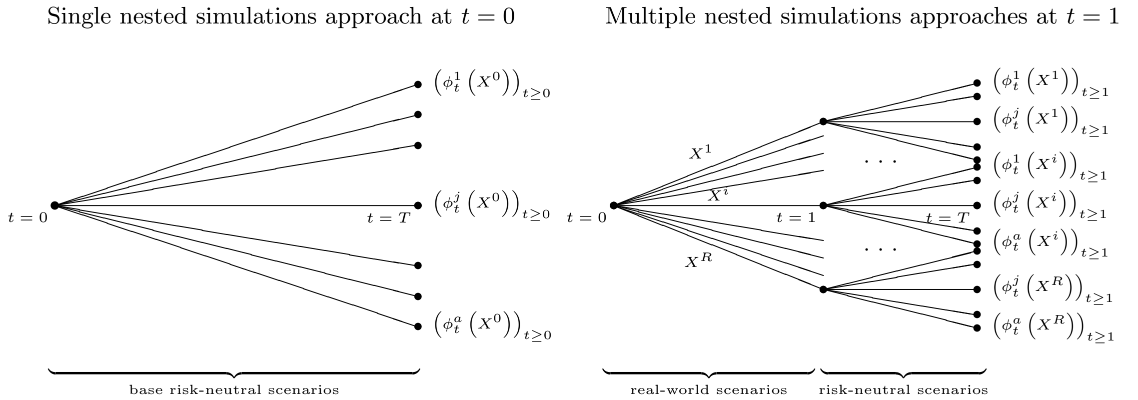

Figure 7: Nested simulation approaches in two set-ups.

The right panel of Figure 7 shows the scenario split for the valuations at time $t=1$. Here the R real-world scenarios represent the shocks to the risk factors in the first projection year. Per real-world scenario X^i, there are a risk-neutral scenarios branching out at $t=1$ yielding a Monte Carlo paths over the remaining projection horizon. The average values over these paths are now estimates for the conditional expected values at $t=1$.

According to Chapter 3 of Investment Committee of DAV (2015), crucial advantages of the nested simulations approach are the methodological consistency in the derivation of the estimates at $t=0$ and $t=1$, their equally high accuracy, and their independence from any mathematical assumptions. For these reasons, nested simulations approaches restricted to few wisely selected scenarios are suitable for out-of-sample validation of other risk aggregation methods. A major drawback is the tremendous computational burden which the insurance industry is currently not able to cope with.

5.2 Standard Formula

General Remarks

The *standard formula* stated by the technical specifications of EIOPA (2014) to be the standard approach to calculate SCRs under the Solvency II regime relies on sensitivity and deterministic scenario analyses, compare Section 3.1. It is therefore a computationally pretty cheap risk aggregation technique. Its core idea is to divide the calculation of the SCR into several hierarchical levels of modules and to aggregate their results from bottom to top, see slightly adjusted Figure 8 from the technical specifications.

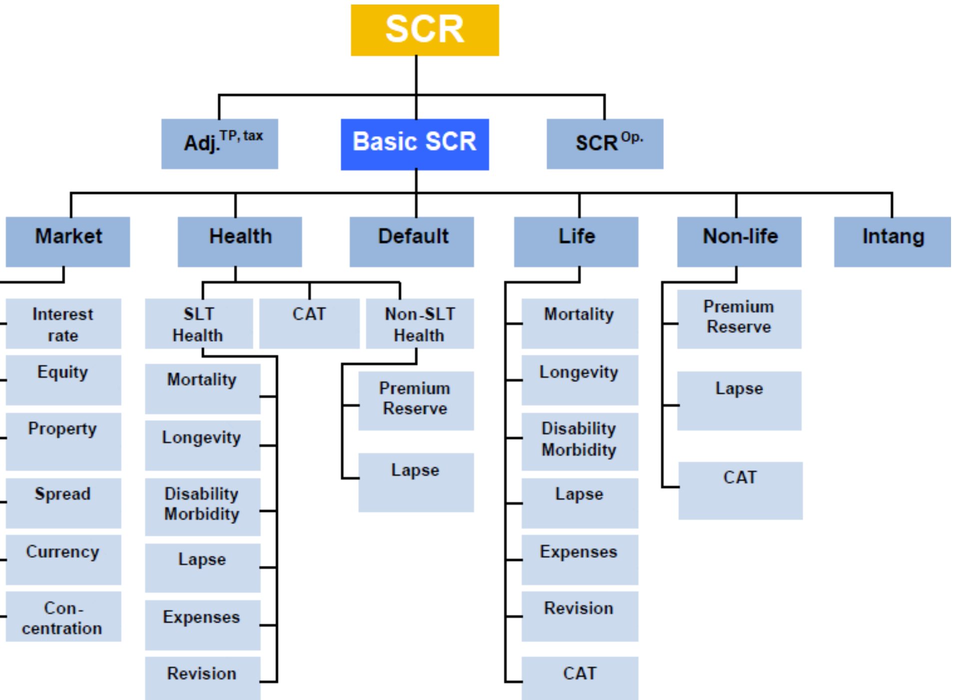

Figure 8: Risk classes of the modular standard formula approach.

On the lowest hierarchical level, the risk factor wise SCRs (interest rate, equity, ...) are determined. These are grouped on the next higher level to obtain the risk module wise SCRs (market, health, ...). Then, the obtained SCRs are passed on to the next level to compute the basic SCR. The last level of aggregation finally yields the overall SCR.

Risk Factor Wise SCRs

The *first level of aggregation* provides individual SCRs for each risk factor module (interest rate, equity, ...), usually by combining the contributions of the sub-modules via the square root formula, i.e.,

$$\mathrm{SCR}\left(X_l\right) = \sqrt{\sum_{g,h} \rho^l_{g,h} \cdot \text{Sub-SCR}^l_g \cdot \text{Sub-SCR}^l_h}, \tag{7}$$

where $\rho^l_{g,h}$ denotes the correlation between sub-modules g and h, or by running deterministic simulations with the risk factor shock being equal to the 1-in-200 loss stress (either the 0.5%- or 99.5%-quantile) of the one-dimensional real-world distribution, i.e.,

$$\mathrm{SCR}\,(X_l) = -\Delta \mid \text{1-in-200 loss risk factor shock}, \tag{8}$$

where "$-\Delta \mid$ 1-in-200 loss risk factor shock" is a loss greater than zero consistent with the notation in Equation (2). Besides (7) and (8), other module specific approaches exist.

Risk Module Wise SCRs

On the *second level of aggregation*, the SCRs for the risk modules (market, health, ...) are derived by aggregating the risk factor wise SCRs obtained on the first level. This is again achieved by the square root formula, i.e.,

$$\mathrm{SCR}^r\,(X_{r_1} + \ldots + X_{r_m}) = \sqrt{\sum_{i,j} \rho^r_{i,j} \cdot \mathrm{SCR}\,(X_i) \cdot \mathrm{SCR}\,(X_j)}. \tag{9}$$

Calculation rule (9) is an application of the formula for determining the quantile of the sum of multiple normally distributed random variables, see Chapter 4 of Investment Committee of DAV (2015). However, the underlying assumption that the loss function follows a normal distribution is daring in CFP models with asymmetric cash flow profiles. Its degree of validity essentially depends on the distributional properties of the one-dimensional loss functions. Moreover, the linearity of the loss function does not properly account for over-proportional losses under certain combined stresses. Because of these model deficiencies, the correlations are set conservatively.

For example, let an insurance company be only exposed to the interest rate and equity risks in the market module, and let their correlation be 0.5 according to the technical specifications by EIOPA (2014). On the first level of aggregation, let the SCRs be found to be 200 and 100, respectively. Then the SCR for the market module is given as

$$\mathrm{SCR}^{\mathrm{market}} = \sqrt{200^2 + 100^2 + 2 \cdot 0.5 \cdot 200 \cdot 100} = 264.56. \tag{10}$$

Except for the non-life module, where a two-stage approach is carried out (at first premium/reserve with lapse, then their result with catastrophe), all SCRs on the second hierarchical level are computed using formula (9) with predefined correlation matrices.

Basic SCR

Thereafter, a two-stage approach is performed on the *third level of aggregation.* Here, the modules "market", "health", "default", "life", and "non-life" are aggregated at first via the square root formula, and then the module "intangibles", representing the intangible asset risk which is valued separately, comes on top of that via another square root formula with correlation 1. The result is the so-called basic SCR, i.e.,

$$\text{Basic SCR} = \sqrt{\sum_{r,s} \rho_{r,s} \cdot \mathrm{SCR}^r \cdot \mathrm{SCR}^s} + \mathrm{SCR}^{\mathrm{intang.}}. \tag{11}$$

For the correlation matrix of the risk modules, see Table 2.

r \ s	**Market**	**Health**	**Default**	**Life**	**Non-life**
Market	1				
Health	0.25	1			
Default	0.25	0.25	1		
Life	0.25	0.25	0.25	1	
Non-life	0.25	0	0.5	0	1

Table 2: Correlation matrix of the third hierarchical level of aggregation (basic SCR).

Overall SCR

The *last level of aggregation* yields the overall SCR as the sum of the basic SCR, an adjustment for the risk absorbing effect of technical provisions and deferred taxes, and the capital requirement for operational risk, i.e.,

$$\text{SCR} = \text{Basic SCR} + \text{Adj.}^{\text{TP, tax}} + \text{SCR}^{\text{Op.}}. \tag{12}$$

The exceptionally low run time of the standard formula approach comes at the cost of the aforementioned mathematically strong and conservative assumptions which are typically not satisfied. Regarding validation, only capital requirements derived from simulations of combined 1-in-200 loss shocks of all risk factors are conceivable as benchmarks for the overall SCR.

5.3 Least-Squares Monte Carlo Method

General Remarks

The *least-squares Monte Carlo (LSMC) method* is a full internal model approach to derive SCRs under Solvency II. By combining Monte Carlo simulations with regression techniques, this method not only becomes computationally significantly cheaper than the nested simulations approach but also feasible for proxy modeling of life insurance companies. Prior to its formalization by e.g. Bauer & Ha (2015) and Krah et al. (2018) for risk aggregation, the LSMC method was spread in the area of finance by e.g. Carriere (1996), Tsitsiklis & Van Roy (2001) and Longstaff & Schwartz (2001) to price American and Bermudan options. While we will go into the details of LSMC for proxy modeling of life insurance companies in Parts I and II of this thesis, we will shed light on its option pricing application in Part III. This section shall introduce the method as a toolkit for calculating capital requirements alongside the risk aggregation techniques presented in Sections 5.1, 5.2 and 5.4. We divide the LSMC method into four steps. The first one addresses the simulation setting, the second and third steps respectively the proxy function calibration and validation, and the last one the full loss probability distribution forecast.

Simulation Setting

In the *first step*, the life insurer has to identify the risk factors $X_1, \ldots, X_d$ its business is exposed to. These are, for example, shocks to the risk-free interest rates movement, equity market value, etc. Furthermore, the budget for the Monte Carlo simulations needs to be determined conditional on the complexity of the CFP model, and be split reasonably between the fitting and validation computations, compare Sections 3.4 and 3.5. An important decision that has to be made involves the allocation of the computations to inner and outer scenarios, see Figure 9.

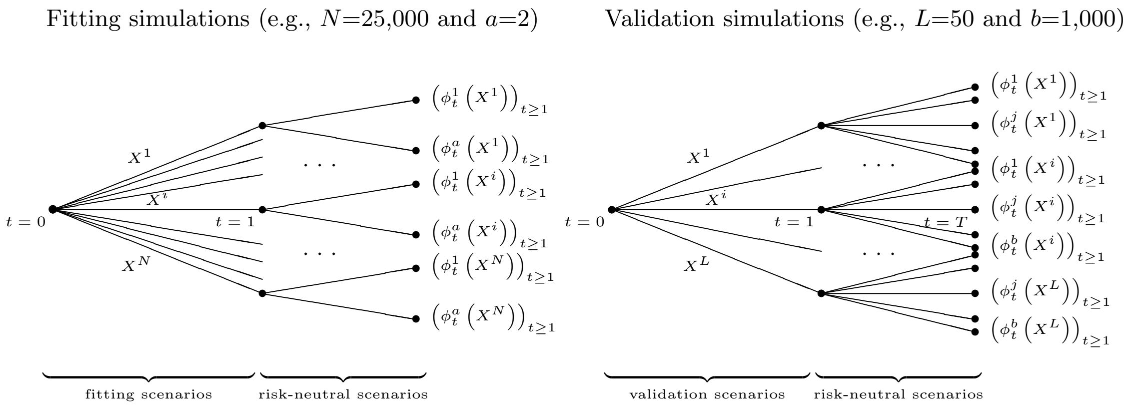

Figure 9: Scenario allocation in the LSMC method.

To ensure optimal usage of the simulation budgets, the fitting scenarios $X^1,\dots,X^N$ can be chosen to follow a Sobol low-discrepancy sequence, see e.g. Niederreiter (1992), whereas the validation scenarios $X^1,\dots,X^L$ should be selected manually according to certain paradigms. Like in the nested simulations approach, the Monte Carlo simulations provide the available capitals after the one-year risk horizon as the differences between the insurer's MVA and BEL per simulation. Finally, by taking the averages over the inner simulations, unique available capital, MVA and BEL values are obtained per outer scenario. These relationships define the fitting points $\left(x^i,y^i\right),\ i=1,\dots,N$, and validation points $\left(x^i,y^i\right),\ i=1,\dots,L$. For example, for $y^i=\mathrm{BEL}^i$ see Figure 10.

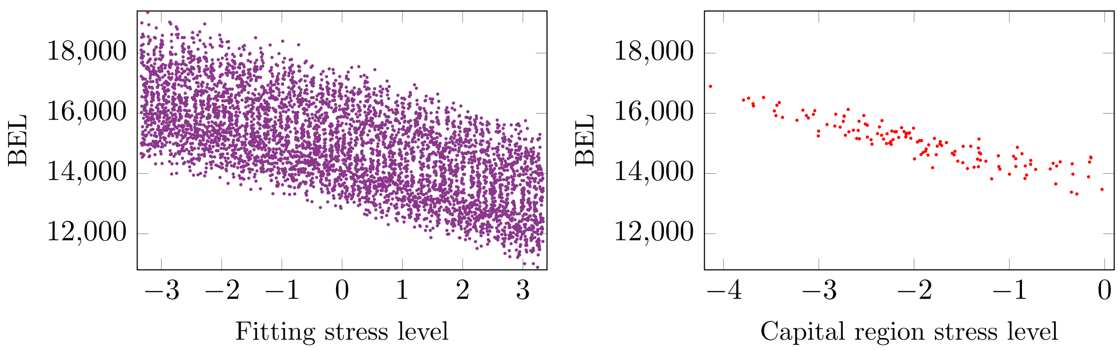

Figure 10: Fitting points and capital region points of BEL plotted in one risk factor dimension.

Proxy Function Calibration

The aim of the *second step* is to find a simple proxy model for the relationship between, for example, the available capital and risk factors in the CFP model. A transition from the "true" available capital $\mathrm{AC}(X)$ over the projection horizon to a *proxy model* or *proxy function* $\widehat{\mathrm{AC}}^{(K,N)}(X)$ conditional on outer scenario X involves two approximations in the LSMC method. Firstly, the conditional expected value is replaced by a linear combination of linearly independent basis functions $e_k(X),\ k=0,1,\dots,K-1$, i.e.,

$$\mathrm{AC}(X)\approx\widehat{\mathrm{AC}}^{(K)}(X)=\sum_{k=0}^{K-1}\beta_k e_k(X). \tag{13}$$

For instance, monomials or Legendre polynomials can act as basis functions but principally any functional form is conceivable. Secondly, vector $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{K-1})^{\mathrm{T}}$ is replaced by the ordinary least-squares (OLS) estimator $\widehat{\boldsymbol{\beta}}^{(\boldsymbol{N})} = \left(\widehat{\beta}_0^{(N)}, \ldots, \widehat{\beta}_{K-1}^{(N)}\right)^{\mathrm{T}}$. As a function of the fitting points $\left(x^i, y^i\right)$, $i = 1, \ldots, N$, where $y^i = \mathrm{AC}^i$, this estimator is typically given by minimization problem

$$\widehat{\boldsymbol{\beta}}^{(\boldsymbol{N})} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^K} \left\{ \sum_{i=1}^{N} \left(y^i - \sum_{k=0}^{K-1} \beta_k e_k\left(x^i\right) \right)^2 \right\}, \tag{14}$$

to which a closed-form solution exists that is easy to compute. Finally, plugging solution (14) into Equation (13) yields the proxy model

$$\widehat{\mathrm{AC}}^{(K)}(X) \approx \widehat{\mathrm{AC}}^{(K,N)}(X) = \sum_{k=0}^{K-1} \widehat{\beta}_k^{(N)} e_k(X). \tag{15}$$

In the life insurance business, a major challenge of the LSMC method consists of determining the functional form and basis functions of the proxy model. Due to the complex cash flow patterns and high risk factor dimensions there is no computationally feasible proxy model readily available. One way of solving this calibration challenge is to rely on adaptive model selection algorithms which automatically build up polynomial proxy functions term by term under further mathematical assumptions, see Part I.

Moreover, there are plenty of variations to the regression methodology that can be tailored to the statistical properties of the data sets, see Part II. As we will see, generalized linear models (GLMs) by Nelder & Wedderburn (1972), generalized additive models (GAMs) by e.g. Hastie & Tibshirani (1986), or feasible generalized least-squares (FGLS) regression, turn out to be well-suited regression methods. Amongst others, artificial neural networks also offer a wide range of options, see e.g. Hejazi & Jackson (2017) and Krah et al. (2020*b*).

An aspect which all regression techniques share is their ability to average out the high standard errors resulting from the few inner simulations per fitting scenario.

Proxy Function Validation

In the *third step*, the calibrated proxy function is validated with the aid of the validation points $\left(x^i, y^i\right)$, $i = 1, \ldots, L$. As opposed to the fitting values, the validation values are averages over many inner simulations and thus have by construction low standard errors. Therefore, they are supposed to come close to the target values of the proxy function. This makes them good reference values. Essentially, the validation constitutes an out-of-sample test for the proxy function.

It is pivotal to define certain *validation criteria* which the proxy function should meet in order to pass the validation procedure. As central ingredients serve summary statistics of the distribution of deviations between the proxy function predictions and validation values such as the (normalized) mean absolute error or maximum absolute error. Furthermore, graphical analyses can be used to verify the regression assumptions and to check the plausibility of the proxy function. For instance, a histogram of the residuals can be created, or the one-dimensional behavior of the proxy function can be compared risk factor wise to suitable validation values.

As long as a proxy function does not pass the validation procedure, its regression methodology in the calibration step has to be refined.

Full Distribution Forecast

After the proxy model has passed the validation procedure, it is used in the *last step* to derive the full loss probability distribution forecast. For this purpose, the real-world scenarios $X^1, \ldots, X^R$ (e.g., $R = 2^{18} = 262,144$) are drawn from the joint real-world distribution of the risk factors, compare Section 3.3. This distribution is, for example, modeled by a copula based on historical data and expert judgment, see e.g. Mai & Scherer (2012). We simply take it as given. All that needs to be done then is to plug the real-world scenarios one after the other into the proxy model to obtain predictions for the available capitals (1), to subtract from each one the base available capital to compute the profits (2), and to determine the SCR as the 99.5% value-at-risk of the resulting full loss probability distribution forecast (3), compare Figure 11.

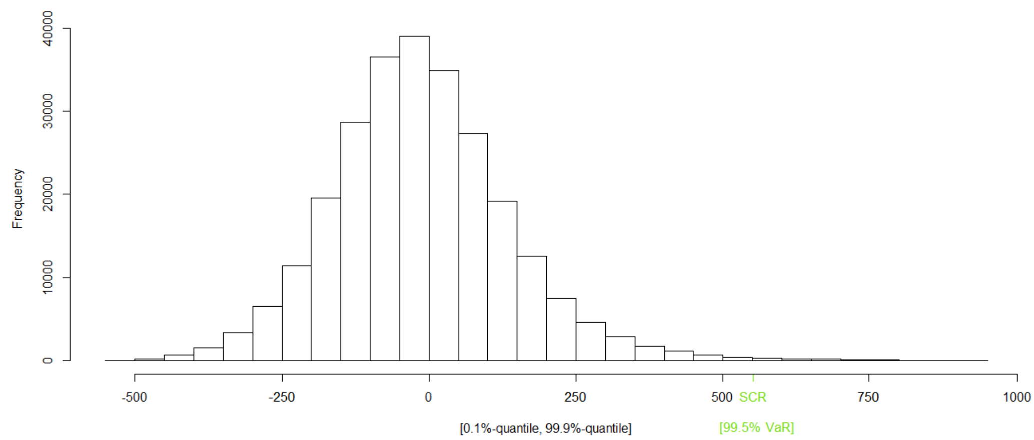

Figure 11: Histogram of real-world loss predictions.

5.4 Replicating Portfolios

General Remarks

Replicating portfolios are an internal model approach to derive SCRs under Solvency II. The central idea of this method is to replicate the cash flows of the available capital or BEL by portfolios consisting of financial instruments that can be valued efficiently. Since financial instruments only depend on capital market risks, actuarial risks cannot be captured by replicating portfolios. Therefore, this risk aggregation technique has to be supplemented by, for example, curve fitting, see e.g. Chapter 5 of Investment Committee of DAV (2015). For the mathematical background of replicating portfolios, see e.g. Natolski & Werner (2014) or Beutner et al. (2016). Like the LSMC method, replicating portfolios are currently applied by multiple large insurance companies. Moreover, the two approaches share four analogous steps: the definition of the simulation setting, the calibration and validation steps, and the full loss probability distribution forecast.

Simulation Setting

After having determined the capital market risk factors $X_1, \ldots, X_d$, the insurance company has to decide for a computation budget in the *first step* and allocate it to the derivations of the fitting and validation points. In this matter, the corresponding inner and outer scenarios need to be specified as well. A possible scenario allocation is depicted in Figure 12.

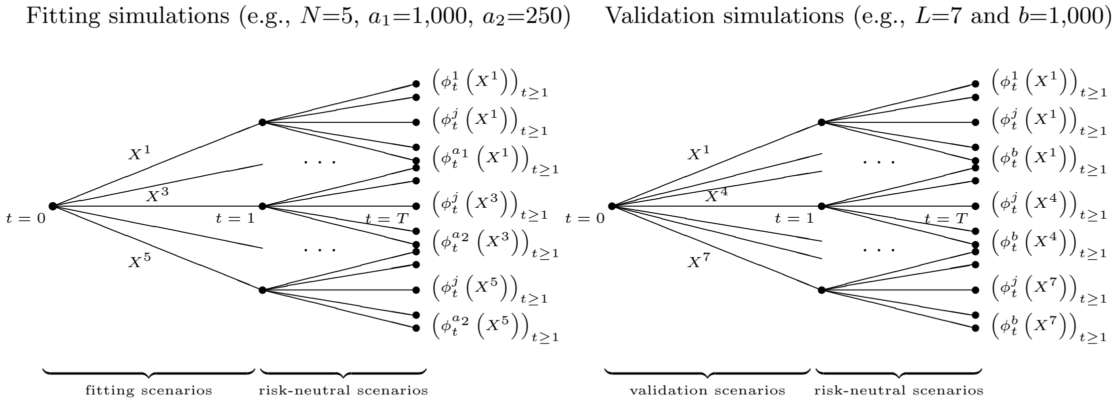

Figure 12: Scenario allocation in replicating portfolios.

The fitting scenarios $X^1, \ldots, X^N$ should be chosen such that they reflect all relevant market situations. An approach as in the LSMC method is therefore possible. Another way, which is in coincidence with the figure, is to take only the base scenario, representing the current market situation, and scenarios expected to yield risk capitals close to the 99.5% value-at-risk of the loss distribution, see Chapter 7 of Investment Committee of DAV (2015). Like the validation scenarios $X^1, \ldots, X^L$, each of the few fitting scenarios is then combined with several inner Monte Carlo simulations. While the validation points $\left(x^i, y^i\right),\ i = 1, \ldots, L$, are obtained as in the previous sections, that is, by averaging scenario wise over all inner simulations, the fitting points $\left(x^i, y^i\right),\ i = 1, \ldots, M$, with, for example, $M = a_1 N_1 + a_2 N_2$ and $N_1 + N_2 = N$ are defined pathwise this time. But scenario wise averages can be taken into account by additional constraints in the regression as well, see Equation (19) below. Validation scenarios that are equal to fitting scenarios need to be complemented by inner scenarios with different seeds.

Replicating Portfolio Calibration

The objective of the *second step* is to derive a *replicating portfolio* for the relationship between, for example, BEL and the risk factors in the CFP model. Transitioning from the "true" best estimate liability $\mathrm{BEL}\,(X)$ to the expectation associated with the cash flows of the replicating portfolio $\widehat{\mathrm{BEL}}^{(K,M)}(X)$ conditional on outer scenario X involves the following two approximations. Firstly, $\mathrm{BEL}\,(X)$ is replaced by a linear combination of linearly independent basis instruments $g_k(X),\ k = 0, 1, \ldots, K-1$, from the universe of financial instruments, i.e.,

$$\mathrm{BEL}\,(X) \approx \widehat{\mathrm{BEL}}^{(K)}(X) = \sum_{k=0}^{K-1} \beta_k g_k(X). \tag{16}$$

Typical basis instruments are coupon bonds for modeling constant policyholder and shareholder cash flows, swaps and swaptions for capturing interest rate guarantees and options, and index derivatives for modeling equity and property exposures. Secondly, the portfolio weight vector $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{K-1})^{\mathrm{T}}$ is replaced by $\widehat{\boldsymbol{\beta}}^{(\boldsymbol{M})} = \left(\widehat{\beta}_0^{(M)}, \ldots, \widehat{\beta}_{K-1}^{(M)}\right)^{\mathrm{T}}$. As a function of the fitting points $\left(x^i, y^i\right),\ i = 1, \ldots, M$, where $y^i = \mathrm{BEL}^i$, this estimator is given, for instance, by replication problem

$$\widehat{\boldsymbol{\beta}}^{(\boldsymbol{M})} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^K} \left\{ \sum_{i=1}^{M} \left(y^i - \sum_{k=0}^{K-1} \beta_k g_k\left(x^i\right) \right)^2 \right\}, \tag{17}$$

which is easy to compute. In Chapter 7 of Investment Committee of DAV (2015), the metric induced by the absolute-value norm is mentioned as a common alternative to the Euclidean metric. Regularization methods such as ridge regression by Hoerl & Kennard (1970) or the least absolute shrinkage and selection operator (LASSO) by Tibshirani (1996) can be applied to punish solutions with large portfolio weights such as long/short positions in options which mutually almost eliminate each other. Finally, plugging solution (17) into Equation (16) yields the replicating portfolio, i.e.,

$$\widehat{\mathrm{BEL}}^{(K)}(X) \approx \widehat{\mathrm{BEL}}^{(K,M)}(X) = \sum_{k=0}^{K-1} \widehat{\beta}_k^{(M)} g_k(X). \tag{18}$$

Often times, further information on the relationship between BEL and the risk factors are available. For example, the mean values over sets of risk-neutral scenarios from market consistent valuations or sensitivity analyses might be at disposal. These can be made use of by additional constraints. Then problem (17) becomes

$$\widehat{\boldsymbol{\beta}}^{(\boldsymbol{M})} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^K} \left\{ \sum_{i=1}^{M} \left(y^i - \sum_{k=0}^{K-1} \beta_k g_k\left(x^i\right) \right)^2 + \sum_{i=1}^{N} \eta_i \left(\bar{y}^i - \bar{g}^i\left(x^i\right)\right)^2 \right\}, \tag{19}$$

where $\bar{y}^i$ and $\bar{g}^i\left(x^i\right)$ are respectively the inner simulation averages over all y^i and $g^i\left(x^i\right) = \sum_{k=0}^{K-1} \widehat{\beta}_k^{(M)} g_k\left(x^i\right)$ per same outer scenario x^i.

Replicating Portfolio Validation

The validation of the calibrated replicating portfolio in the *third step* works in principle like the proxy function validation in the LSMC context except that the scenarios of the validation points $\left(x^i, y^i\right),\ i = 1, \ldots, L$, are now inserted into the calibrated replicating portfolio to obtain the model predictions.

Beyond its purpose to wave appropriate replicating portfolios through to the fourth step, the third step can also be used to identify the most suitable approximation techniques for calibration by comparison. This, by the way, holds for the LSMC method as well. In the case of replicating portfolios, the validation can give answers related to questions on which universe of financial instruments, optimization algorithm, loss metric, regularization method, or additional constraints to settle for.

If a replicating portfolio does not meet the specified validation criteria, it has to be refined based on the kinds of options provided in the previous paragraph.

Full Distribution Forecast

The *fourth step* can also be adopted from the LSMC method by replacing the final proxy function by the final replicating portfolio. For a histogram capturing a real-world loss probability distribution forecast based on $R = 2^{18} = 262,144$ scenarios following from the LSMC method, see again Figure 11. The histogram following from a replicating portfolio would look quite similar given it has a similar approximation quality as the final LSMC proxy function.

Part I

A Least-Squares Monte Carlo Framework in Proxy Modeling of Life Insurance Companies

Résumé

The Solvency II directive asks insurance companies relying on internal models to derive their solvency capital requirements from their full loss distributions over the coming year. While this is in general computationally infeasible in the life insurance business, an application of the least-squares Monte Carlo (LSMC) method offers a possibility to overcome this computational burden. We start with how LSMC for risk aggregation is related to its original use in American option pricing. Then we outline in detail the practical challenges a life insurer faces, a theoretical framework of the LSMC method for calculating capital requirements, the necessary steps on the way to a reliable proxy modeling and a concrete least-squares regression algorithm that can be implemented in the life insurance industry. Finally, we illustrate this algorithm and the advantages of the LSMC approach via a slightly disguised real-world application.

6 Introduction

Solvency II

The Solvency II directive passed by European Parliament & European Council (2009) has come into effect to reduce the risk of insolvency of the insurance and reinsurance companies from EU countries. As a result, a higher degree of financial stability shall be achieved and the policyholders' insurance sums are expected to become safer. More specifically, the directive demands from the insurers to keep certain amounts of reserves - the solvency capital requirements (SCRs) - as securities against adverse market developments. These reserves have to reflect the risks to which the insurers are exposed to.

The standards set out in the Solvency II directive mark the starting point for the recent developments of proxy modeling to calculate SCRs in the insurance sector. Article 122(2) of this directive states:

> "Where practicable, insurance and reinsurance undertakings shall derive the solvency capital requirement directly from the probability distribution forecast generated by the internal model of those undertakings, using the Value-at-Risk measure set out in Article 101(3)."

The crucial point of this quotation is that the insurers are asked to derive their full loss distributions in case they do not want to rely on the much simpler standard formula approach, compare Section 5.2. A common understanding is, that in order to obtain a reasonably accurate full distribution, several hundred thousand simulations are necessary. With the conventional nested simulations approach by Bauer et al. (2012), compare Section 5.1, for each of these simulations at least $1,000$ Monte Carlo valuations must be carried out. This leaves the insurance companies with the task to perform hundreds of millions of available capital calculations to generate their full loss distributions.

CFP Models

The mere fact that such an extensive calculation is anticipated in the directive shows that, in the years preceding the Solvency II directive, it was believed that by the time Solvency II would be introduced much lower hardware costs and increased valuation efficiency would allow European life insurance companies to perform that many simulations. Since the industry is currently still far away from such computational capacities, the companies face the challenge to find suitable approximation techniques. Among them are sophisticated techniques such as Least-Squares Monte Carlo (LSMC), which we focus on hereinafter, or replicating portfolios, compare Sections 5.3 and 5.4, respectively. Major parts of the following exposition on the LSMC method have already been published in Krah et al. (2018). The central idea of LSMC is to get along with comparatively few wisely selected Monte Carlo simulations and to derive an available capital proxy function by least-squares regression based on the limited simulation output.

The computational challenge arises primarily from the cash-flow-projection (CFP) models, compare Section 1.1. A CFP model replicates an insurer's contract portfolio and assets from the administration systems and projects their cash inflows and outflows into the future by satisfying regulatory standards like the profit sharing mechanism and implementing company-specific management actions. Besides being a full balance sheet projection tool, a CFP model is also a pricing machine of a complex derivative, processing premiums, benefits, costs and dividends. The more complex an insurer's CFP model is, the higher are the computational costs incurred for running the available capital Monte Carlo simulations.

It is our intention to shed some light on how the nested valuation problem can be overcome by applying the LSMC approach proposed by Cathcart (2012) and Bauer & Ha (2015). We divide the calculation process into four steps. In the first step, we allocate the available capacities to the simulations that will be used for generating the available capital fitting and validation data, and run the Monte Carlo simulations of the CFP model accordingly. Conditional on the fitting data, the proxy function for the available capital is then calibrated in the second step. This step is mathematically most demanding as it offers a great variety of regression methods which must be chosen from. A good method not only reliably generates proxy functions which capture the behavior of the CFP model well, but is also transparent, flexible with regard to the characteristics of the model and fast. In the third step, the proxy function is validated conditional on the validation data. Once the validation has been successful, the proxy function is passed on for evaluation to the last step, where it yields an approximation to the insurer's full loss distribution. Finally, the SCR is computed as the 99.5% value-at-risk of that loss distribution.

Outline

Before we give a current snapshot of why and how the companies in the life insurance sector can use an LSMC-based approach to make their full loss probability distribution forecasts, we offer in Section 7.1 a short excursion to the origins of the LSMC method in American option pricing, and describe in Sections 7.2–7.4 the basic idea behind the method to calculate capital requirements of insurance undertakings and relate the two applications to each other. Our main objective in the next section is to close the gap between theory and practice by designing a concrete algorithm along with a suitable theoretical framework. After some general remarks in Section 8.1, we give a detailed description of the simulation setting in Section 8.2, explain concepts for proxy function calibration and validation procedures in Sections 8.3 and 8.4, respectively, and demonstrate the actual application of the LSMC model to forecast the full loss distribution in Section 8.5. We highlight the practical considerations an insurer should make in the various necessary steps. Even though we do not intend to give a step-by-step worked example, we complete with a numerical illustration in which we define the approximation task in Section 9.1, place the focus on the calibration and validation procedures in Sections 9.2 and 9.3, respectively, the loss distribution forecast in Section 9.4 and the computation time in Section 9.5.

7 From American Option Pricing to SCR Calculation

7.1 American Option Pricing

Originally introduced in finance as a Monte Carlo alternative for pricing American or Bermudan options, the LSMC approach combines Monte Carlo methods with regression techniques, see e.g. Carriere (1996), Tsitsiklis & Van Roy (2001) and Longstaff & Schwartz (2001). The basic idea is to translate a set of simulated evolutions of the paths of underlying stock price(s) on a discrete time grid in suitable regression functions. While for standard European options it is straight forward to calculate the resulting payoffs, to average over them and consider the discounted average as an approximation for the option price, such a procedure is not feasible for American options due to the possibility of an early exercise.

Instead, we have to follow each path of the stock price closely and have to decide at each future time instantly if it is better to exercise the option now or to continue holding it. While the intrinsic value of the option is given by its actual payoff when immediately exercised, the continuation value depends on the future performance of the stock prices. Of course, the latter is not known at the current time. This, however, is exactly where the regression techniques come into the game. As the continuation value can be expressed as the discounted expectation of the future value of the American option conditional on the current stock price given it will not be exercised immediately, it can be expressed as a typically unknown function of the current stock price. Such a function can then be estimated, for instance, via linear regression, that is, with the help of the least-squares method, see e.g. Korn et al. (2010) for a detailed description of the full algorithm. Having obtained a regression function for the payoff at future time steps conditional on the current one, one can compare the intrinsic value of the option to the approximated continuation value by plugging the current stock price in the regression function. In this way, one can calculate the optimal exercise boundary of an American option backward from maturity to its initial time.

This LSMC technique for American options has the typical Monte Carlo advantage of beating tree methods or partial differential equation (PDE) based methods for multiple underlyings but tends to be slow in univariate applications.

7.2 Nested Valuation Problem

As outlined in the introduction, to calculate the capital requirement of an insurance company, its assets and liabilities in the CFP model have to be projected one year into the future under a large number of real-world scenarios. The obtained positions after one year have to be (re-)valued. For this, conditional expectations (under a pricing measure!) have to be calculated. To obtain a reliable estimate for those values, for each real-world scenario numerous stochastic simulations of the CFP model of the company need to be carried out in the conventional nested simulations Monte Carlo approach as described in Bauer et al. (2012). This approach with simulations in the simulations typically ends up in a nested valuation problem as it will for most life insurance companies exceed their computational capacities. Life insurance companies would need hundreds of millions of simulations for an acceptably accurate nested valuation, whereas due to time and hardware constraints they typically have the capacities to perform at most a few hundred thousand simulations.

7.3 Calculating Capital Requirements

Taking up the LSMC idea from American option pricing above, Bauer & Ha (2015) have extended the scope of the LSMC approach to the risk management activities of financial institutions such as in particular life insurance companies. They have suggested a way to overcome the nested valuation problem of calculating capital requirements in nested simulations approaches. Instead of simulating paths of stock price(s) for valuing options, they simulate paths of CFP models for valuing balance sheet items by Monte Carlo simulation. Similarly to estimating conditional continuation values, they estimate the available capital of a company as the difference between its assets and liabilities conditional on the real-world scenarios by ordinary least-squares regression.

Thereby, they also highlight the flexibility of the LSMC approach to switch between pricing and projection, i.e., to simulate under different probability measures over disjoint time intervals. Particularly, they cope with the nested valuation problem by introducing a hybrid probability measure: While the physical measure captures the one-year real-world scenarios under which the assets and liabilities shall be evaluated, the risk-neutral measure concerns the valuations in the CFP model after the one-year time period. Additional theoretical results can be found in Natolski & Werner (2016). Key to overcoming the nested valuation problem is that the LSMC technique gets along with only few stochastic simulations per real-world scenario. Where numerous simulations can be forgone in the LSMC approach compared to the nested simulations approach, the approximation of the available capital by ordinary least-squares comes in.

7.4 Least-Squares Regression

For a better understanding, suppose the number of stochastic simulations required to estimate the available capital under an arbitrary real-world scenario in the conventional approach is $\widetilde{N}$. Then, the LSMC approach takes advantage of the idea that the $\widetilde{N}$ required simulations do not necessarily have to be performed based on the single real-world scenario under which the available capital shall be derived. This means that the $\widetilde{N}$ simulations can be carried out based on *different* real-world scenarios. The effects of the different real-world scenarios just have to be excludable by ordinary least-squares regression. By implication, we can select and perform $\widetilde{N}$ simulations in the LSMC approach once for all relevant real-world scenarios together and thereby decrease the required computational capacities tremendously compared to the conventional approach.

Compare this to the American option pricing problem where we only simulate under the risk-neutral measure. However, each scenario at time t is only followed by one price path. Starting from maturity, the continuation values in the preceding period are simply calculated via a regression over the cross-sectional payments from not exercising the option. Thus, nested simulation at each time t for each single scenario is replaced by comparing the intrinsic option value with the current stock price inserted in the just obtained regression function. In the calculation of capital requirements, we do not even have to think about the exercise problem.

8 Least-Squares Monte Carlo Framework

8.1 General Remarks

In the following, we adopt the LSMC technique as described in Cathcart (2012) and Bauer & Ha (2015) and develop it further to meet the specific needs of proxy modeling in life insurance companies. Amongst others, Barrie & Hibbert (2011), Milliman (2013) and Bettels et al. (2014) have already given practical illustrations of the LSMC model in risk management.

We close the gap between theory and practice by developing a practical algorithm along with a suitable theoretical framework. Variants of this algorithm have already been applied successfully in many European countries and different life insurance companies under the Solvency II directive. Although this approach appears superior to other methods in the industry and the algorithm meets the demands for reliable derivations of SCRs, various active research streams search for refinements of the proxy functions, see the next part of this thesis for details.

In this section we look in detail at the necessary steps and ingredients on the way to a reliable proxy modeling using the LSMC framework. The particular steps are:

- a detailed description of the simulation setting and required task;
- a concept for a calibration procedure for the proxy function;
- a validation procedure for the obtained proxy function; and
- the actual application of the LSMC model to forecast the full loss distribution.

We would like to emphasize that only a carefully calibrated and rigorously validated proxy function can be used in the LSMC approach. Otherwise, there is the danger of, for example, using an insufficiently good approximation or an overfitted model.

8.2 Simulation Setting

8.2.1 Filtered Probability Space

We adopt the simulation framework of Bauer & Ha (2015) and modify it where necessary to describe the current state of implementation in the life insurance industry. Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a complete filtered probability space. We also assume the existence of a risk-neutral probability measure $\mathbb{Q}$ equivalent to the physical probability measure $\mathbb{P}$. Since $\mathbb{Q}$ is supposed to concern all tradable goods in the market, it is directly related to the insurer's risk factors on the assets side but not necessarily on the liabilities side. By $\mathbb{P}$, we capture the real-world scenario risk an insurer faces with regard to both its assets and liabilities.

We model the risk an insurer is exposed to in the first projection year by a vector $X = (X_1, \ldots, X_d)$, $X \in \mathbb{R}^d$, where each component represents the stress intensity of a financial or actuarial risk factor. We refer to X under $\mathbb{P}$ as an *outer scenario*. Conditional on X, we model the insurer's uncertainty under $\mathbb{Q}$ by a time-dependent market consistent Markov process $(\phi_t(X))_{t\geq 0}$ that reflects the developments of stochastic capital market variables over the projection horizon. We refer to Monte Carlo simulations $(\phi_t(X))_{t\geq 0}$ for the risk-neutral valuation as *inner scenarios*. Each outer scenario is assigned a set of inner scenarios. Under $\mathbb{Q}$, we can price any security by taking the expectation of its discounted cash flows. The discounting takes place with respect to the process $(B_t)_{t\geq 0}$ with $B_t = \exp\left(\int_{s=0}^{t} r_s \mathrm{d}s\right)$, where $r_t = r(\phi_t(X))$ denotes the instantaneous risk-free interest rate. For more details on outer and inner scenarios as well as stochastic and risk-neutral scenarios, see respectively Sections 1.4, 3.1 and 3.2.

The filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ of the σ-algebra $\mathcal{F}$ contains the inner scenario sets, that is, the sets with the financial and actuarial information over the projection horizon. The σ-algebra $\mathcal{F}_t$ captures solely the sets with the accumulated information up to time t.

8.2.2 Solvency Capital Requirement

An insurance company is interested in its risk profile to assess its risks associated with possible business strategies. With the knowledge of how its risks influence its profit, an insurer can steer the business using an appropriate balance between risks and profits. In risk management, a special focus is given to an insurer's loss distribution over a one-year horizon due to the Solvency II directive, compare European Parliament & European Council (2009).

We define an insurer's *available capital* at time t as its market value of assets minus liabilities and abbreviate it by AC_t. Furthermore, we define its one-year *profit* Δ as the difference between its discounted available capital after one year and its initial available capital, i.e.,

$$\Delta = B_1 \mathrm{AC}_1 - \mathrm{AC}_0, \tag{20}$$

where $B_1 \mathrm{AC}_1$ constitutes an *after risk* result and AC_0 the unique *before risk* result. The *solvency capital requirement* (SCR) is given as the value-at-risk of the insurer's one-year losses, that is, the negative of Δ, at confidence level φ. Formally, this is the φ-quantile of the loss distribution, i.e.,

$$\mathrm{SCR} = \mathrm{VaR}_{\varphi}\left(-\Delta\right) = \inf\left\{y \in \mathbb{R} \mid F_{-\Delta}^{\mathbb{P}}\left(y\right) \geq \varphi\right\}, \tag{21}$$

where $F_{-\Delta}^{\mathbb{P}}\left(y\right) = \mathbb{P}\left(-\Delta \leq y\right)$ denotes the cumulative distribution function of the loss under $\mathbb{P}$. Hence, if the initial available capital AC_0 is equal to SCR, the insurer is statistically expected to default only in $1 - \varphi$ business years. For instance, $\varphi = 99.5\%$ as set out in Article 101(3) of the Solvency II directive means the company will default only with probability 0.5%, see Figure 2. Additional information on the introduced balance sheet items can be found in Section 1.2.

8.2.3 Available Capital

Assume the projection of asset and liability cash flows occurs annually at the discrete times $t = 1, \ldots, T$. Let Z_t denote the net profit (dividends and losses for the shareholders after profit sharing) at time t and let T mark the projection end. At projection start, we can express the after risk available capital conditional on outer scenario X as

$$\mathrm{AC}\left(X\right) = E^{\mathbb{Q}}\left[\sum_{t=1}^{T} B_t^{-1} Z_t \mid X\right]. \tag{22}$$

In our modeling, we approximate the under Solvency II sought-for discounted available capital after a one-year horizon by the after risk available capital (22), i.e., $B_1 \mathrm{AC}_1 \approx \mathrm{AC}\left(X\right)$. We assume that an outer scenario $X = \left(X_1, \ldots, X_d\right)$ realizes immediately after projection start and not after one year. This consideration is important as an insurance company only knows its current CFP model but not its future ones such as after one year. Each summand $B_t^{-1} Z_t$ in the expectation on the right-hand side is a result of the characteristics of the underlying CFP model. Under the Markov assumption, such a summand takes on a value conditional on the present inner scenario component $\phi_t\left(X\right)$ in a simulation. Mathematically speaking, the summands can be interpreted as functionals z_t on the vector space of profit cash flows to $\mathbb{R}$. Each CFP model can be represented by an element $\left(z_1, \ldots, z_T\right)$ of a suitable function space. As long as the risk-free interest rate enters $\phi_t\left(X\right)$, we can write $z_t\left(\phi_t\left(X\right)\right) = B_t^{-1} Z_t$, $t = 1, \ldots, T$, so that, by linearity of expectation, Equation (22) becomes

$$\mathrm{AC}\left(X\right) = \sum_{t=1}^{T} E^{\mathbb{Q}}\left[z_t\left(\phi_t\left(X\right)\right) \mid X\right]. \tag{23}$$

For the sake of completeness, if we are interested in an economic variable other than the available capital such as the market value of assets or best estimate liability, we can proceed in principle as described above. Essentially, we would have to adjust the cash flows Z_t of the available capital in Equations (22) and (23) such that they represent the respective variable. Instead of the SCR, another economic variable such as the changes in assets or liabilities would be defined in Equation (21).

8.2.4 Fitting Points

Like Cathcart (2012) but unlike Bauer & Ha (2015), we specially construct a fitting space on which we define the outer scenarios that are fed into the CFP model solely for simulation purposes. We call these outer scenarios *fitting scenarios*. The fitting scenarios should be selected such that they cover the space of real-world scenarios sufficiently well. This is important as the proxy function of the available capital or another economic variable is derived based on these scenarios. See Sections 3.3 and 3.4 for more information on fitting and real-world scenarios. To formalize the distribution of the fitting scenarios, we introduce another physical probability measure $\mathbb{P}'$ besides $\mathbb{P}$. Apart from the physical measure, we adopt the filtered probability space from above so that we refer for the fitting scenarios to the space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}')$. We specify the fitting space as a d-dimensional cube

$$S_{\text{fit}} = \Pi_{l=1}^{d}[a_l, b_l] \subset \mathbb{R}^d, \tag{24}$$

where the intervals $[a_l, b_l]$, $l = 1, \ldots, d$, indicate the domains of the risk factors. The endpoints of these intervals correspond respectively to a very low and a very high quantile of the risk factor distribution. For an example of a risk factor domain, see Section 4.2.

The set of fitting scenarios x^i, $i = 1, \ldots, N$, and the number of inner scenarios a per fitting scenario need to be defined conditional on the run time of the CFP model in the given hardware architecture. Thereby, the minimum number of scenarios required to obtain reliable results needs to be guaranteed. Moreover, the calculation budget should be split reasonably between the fitting and validation computations. To allocate the fitting scenarios on S_{fit}, we need to specify the physical probability measure $\mathbb{P}'$ and suggest an appropriate allocation procedure.

Once the fitting scenarios have been defined, the inner scenarios $\left(\phi_t^j\left(x^i\right)\right)_{t\geq 1}$, $j = 1, \ldots, a$, per fitting scenario x^i, $i = 1, \ldots, N$, must be generated. An economic scenario generator (ESG) can be employed to take over this task. Essentially, an ESG simulates market consistent capital market variables over the projection horizon under $\mathbb{Q}$ and ensures risk neutrality. For an exemplary stochastic modeling of capital market variables, see Section 2.

When all required inner scenarios are available, the Monte Carlo simulations of the CFP model are performed. These simulations provide the results for the available capital, i.e.,

$$\left(y'\right)^{i,j} = \left(\mathrm{AC}'\left(x^i\right)\right)^j = \sum_{t=1}^{T} z_t \left(\phi_t^j\left(x^i\right)\right),\ j = 1, \ldots, a,\ i = 1, \ldots, N. \tag{25}$$

The averages of these results over the inner scenarios yield the *fitting values* per fitting scenario, i.e.,

$$y^i = \mathrm{AC}\left(x^i\right) = \frac{1}{a}\sum_{j=1}^{a}\left(y'\right)^{i,j},\ i = 1, \ldots, N. \tag{26}$$

Hereinafter, we call the points $\left(x^i, y^i\right),\ i = 1, \ldots, N$, consisting of the fitting scenarios and fitting values *fitting points*.

8.2.5 Practical Implementation

We now outline how the mathematical framework presented above can actually be implemented in the industry.

The first step of a life insurance company is to identify the risks its business is exposed to. Some of these risks are not quantifiable and are hence excluded from the modeling in the CFP models. However, if a risk is significant for the company and can also be implemented in the CFP model, it will usually be considered. Some quantifiable risks may not be relevant for a company: A company underwriting solely risk insurance policies is not exposed to the longevity risk for instance.

In Table 3, we list typical risks which can be implemented in the CFP model. In this example, $d = 17$ is the number of risk factors being quantifiable and relevant for the company. While the first nine risk factors $X_1, \ldots, X_9$ are shocks on the capital market, the remaining eight $X_{10}, \ldots, X_{17}$ constitute actuarial risks, compare Section 1.3. We model the risk factors as either additive or multiplicative stresses. For each risk factor, the *base value* is usually zero. Except for the mortality catastrophe stress, all actuarial risk factors in the table can be modeled as symmetric multiplicative or additive stresses.

Component	Risk Factor Description
X_1	Risk-free interest rates movement
X_2	Change in interest rate volatility
X_3	Change in equity volatility
X_4	Shock on volatility adjustment (if used by the company)
X_5	Credit default
X_6	Credit spread widening
X_7	Currency exchange rate risk
X_8	Shock on equity market value
X_9	Shock on property market value
X_{10}	Lapse stress on best estimate assumptions
X_{11}	Mortality catastrophe stress with a one-off increase in mortality
X_{12}	Mortality level stress on best estimate assumptions
X_{13}	Mortality trend stress on best estimate assumptions
X_{14}	Mortality volatility stress on best estimate assumptions
X_{15}	Longevity level stress on best estimate assumptions
X_{16}	Morbidity stress on best estimate assumptions
X_{17}	Expenses stress on best estimate assumptions

Table 3: Risk factors in the CFP model.

Some of the risk factors can depend on vectors of underlying random factors. For example, the historic interest rate movements cannot be explained reasonably well with a one-factor model, so companies use two-factor or three-factor models for the implementation of X_1. They can also include risk-free rates in different currencies which increase the number of dimensions even further. Equally, the equity shock X_8 can be the outcome of several shocks of indices if the company is exposed to different types of equity. The spread widening risk X_6 can also be multidimensional. For additional information on the modeling of shocks in CFP models, see Section 4.

As far as the number of outer scenarios is concerned, it depends on the dimensionality of the fitting space and calculation capacities. Numbers ranging from $N = 5,000$ to $N = 100,000$ fitting scenarios have been seen and tested in the industry. For the inner scenarios, the natural choice would be $a = 1$ as this would permit greater diversification among the fitting scenarios (and is in line with the original LSMC approach!). However, in the numerical examples we set $a = 2$ based on the observation that the benefits from the method of antithetic variates overcompensate the drawbacks from the reduction of the fitting scenarios. For the method of antithetic variates, see e.g. Chapter 4.2 in Glasserman (2004).

For the practical implementations, it is important that the fitting space S_{fit} in Equation (24) is a cube. This allows using low-discrepancy sequences which is a powerful tool ensuring optimal usage of the scenario budget in Monte Carlo simulations. By far the most widely used ones are Sobol low-discrepancy sequences which are easy to implement. They have a big advantage in that they make sure that each new addition of fitting points will be optimally placed in a certain sense. For details and the exact definition of what it means that a sequence is low-discrepancy, see Niederreiter (1992). Since a Sobol sequence is defined on the cube $\Pi_{l=1}^{d}[0,1]$, we perform a linear transformation of the dimensions to map it to our fitting space S_{fit}.

It is worth noting that some risk factors require an ESG for the inner scenarios conditional on the fitting scenarios to be generated. The financial models determining the dynamics of, for instance, interest rates, equity, property and credit risk are implemented in ESGs. The first four risk factors $X_1, \ldots, X_4$ from Table 3 are always modeled directly in the ESG, the next three ones $X_5, \ldots, X_7$ can be modeled in such a way as well. The remaining risk factors are modeled directly as input for the CFP model.

As the final step, the Monte Carlo simulations of the CFP model conditional on the inner and outer scenarios are performed leading to the non-averaged values in Equation (25) for the available capital. After averaging, we get the fitting values in Equation (26) per fitting scenario and thus the fitting points which enter the regression.

8.3 Proxy Function Calibration

8.3.1 Two Approximations

Now we describe how the proxy functions for the CFP model are practically obtained. As stated before, Cathcart (2012) and Bauer & Ha (2015) have transferred the LSMC approach from American option pricing to the field of capital requirement calculations. A more practical outline of this approach was given by Koursaris in Barrie & Hibbert (2011). Instead of approximating conditional continuation values, they approximate conditional profit functions by an LSMC technique. Both application fields, option pricing and capital requirement calculation, have in common that the proxy functions shall predict aggregate future cash flows conditional on current states. Overall, Cathcart (2012) and Bauer & Ha (2015) stay rather theoretical by focusing on stylized portfolios so that their considerations regarding the calibration of well-suited proxy functions are not fully applicable to practice. Therefore, we adopt their theoretical framework and complement it by a concrete calibration algorithm.

Due to limited computational capacities and a finite set of basis functions, we have to make two approximations with the aim of evaluating the economic variable in Equation (23) conditional on outer scenario X. Firstly, we replace the conditional expected value over the projection horizon with a linear combination of linearly independent basis functions $e_k(X) \in L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}')$, $k = 0, 1, \ldots, K-1$, i.e.,

$$\mathrm{AC}(X) \approx \widehat{\mathrm{AC}}^{(K)}(X) = \sum_{k=0}^{K-1} \beta_k e_k(X), \tag{27}$$

with $e_0(X) = 1$ and intercept β_0. Secondly, we replace vector $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{K-1})^{\mathrm{T}}$ with the ordinary least-squares estimator $\widehat{\boldsymbol{\beta}}^{(\boldsymbol{N})} = \left(\widehat{\beta}_0^{(N)}, \ldots, \widehat{\beta}_{K-1}^{(N)}\right)^{\mathrm{T}}$. In dependence of the fitting points (x^i, y^i), $i = 1, \ldots, N$, derived by the Monte Carlo simulations, $\widehat{\boldsymbol{\beta}}^{(\boldsymbol{N})}$ is given by

$$\widehat{\boldsymbol{\beta}}^{(\boldsymbol{N})} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^K} \left\{ \sum_{i=1}^{N} \left(y^i - \sum_{k=0}^{K-1} \beta_k e_k(x^i) \right)^2 \right\}. \tag{28}$$

By further approximating Equation (27) in terms of Equation (28), we obtain

$$\mathrm{AC}(X) \approx \widehat{\mathrm{AC}}^{(K)}(X) \approx \widehat{\mathrm{AC}}^{(K,N)}(X) = \sum_{k=0}^{K-1} \widehat{\beta}_k^{(N)} e_k(X), \tag{29}$$

which we can evaluate at any outer scenario X.

8.3.2 Convergence

According to Proposition 3.1 by Bauer & Ha (2015), the LSMC algorithm is convergent. We split their proposition into two parts and can adapt it such that we are able to apply their findings to our modified model setting. Since we have made the analogies and differences between their and our model setting clear above, the translation of their proposition into our propositions is simple. Therefore, we do not prove our propositions explicitly. For completeness, we state these main convergence results.

Proposition 1. $\widehat{\mathrm{AC}}^{(K)}(X) \to \mathrm{AC}(X)$ *in* $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}')$ *as* $K \to \infty$.

Proposition 2. $\widehat{\mathrm{AC}}^{(K,N)}(X) \to \widehat{\mathrm{AC}}^{(K)}(X)$ *as* $N \to \infty$ $\mathbb{Q}$-*almost surely.*

By conclusion, we arrive at $\widehat{\mathrm{AC}}^{(K,N)}(X) \to \mathrm{AC}(X)$ in probability and thus in distribution as $K, N \to \infty$. This means that a proxy function converges in probability to the true values of the economic variable if the basis functions are selected properly. Since $\mathbb{P}'$ can be replaced with $\mathbb{P}$, the propositions hold conditional on both the scenarios on the fitting space and the real-world scenarios. The convergence in distribution implies furthermore that the distribution of a proxy function evaluated at the real-world scenarios approximates the actual real-world distribution. A linear transformation of the available capital finally yields the special case that the estimate for the SCR converges to the actual one as $K, N \to \infty$. We take up these results once more later.

8.3.3 Adaptive Algorithm

The crucial ingredient for the regression step is the choice of the basis functions. We build up the set of basis functions according to an algorithm given in Krah (2015). To illustrate this algorithm, we adopt a flowchart from Krah (2015) and depict a slightly generalized version of it in Figure 13. Hereinafter, we explain a customized version comprehensively and discuss refinements. The core idea is to begin with a very simple proxy function and to extend it iteration by iteration until its goodness of fit can no longer be improved. We derive a proxy function for the available capital but the process can equally be run for other economic variables such as the best estimate liability or best estimate of guaranteed liabilities.

The procedure starts in the upper left side in Figure 13 with the specification of the basis functions for the start proxy function. Typically, this will be a constant function. Then, we perform the initial ordinary least-squares regression of the fitting values against the fitting scenarios ($k = 0$). With our choice of a constant function, the start proxy function becomes the average of all fitting values.

8.3.4 Model Selection Criterion

In the last step of the initialization, we determine a model selection criterion and evaluate it for the start proxy function. The model selection criterion serves as a relative measure for the goodness of fit of the proxy functions in our procedure. In the applications in the industry, one of the well-known information criteria such as the Akaike information criterion (AIC from (Akaike 1973)) or the Bayesian information criterion (BIC) is applied. For that it is assumed that the fitting values conditional on the fitting scenarios, or equivalently the errors, are normally distributed and homoscedastic. As we cannot guarantee this, the final proxy function has to pass an additional validation procedure before being accepted. The preference in the industry for AIC is based on its deep foundations, easiness to compute and compatibility with ordinary least-squares regression under the assumptions stated above. For a comparison of AIC, BIC, Mallows's C_p, adjusted R^2, cross-validation and of model selection criteria relying on F-tests within the same LSMC application, see e.g. Reichenwallner (2014).

AIC will help us to find the appropriate compromise between a too small and too large set of basis functions. In our case, it has the particular form of a suitably weighted sum of the calibration error and the number of basis functions:

$$\mathrm{AIC} = \underbrace{N\left(\log\left(2\pi\left(\widehat{\sigma}^{(N)}\right)^2\right)+1\right)}_{\text{calibration error}} + \underbrace{2\,(K+1)}_{\text{number of basis functions}} \tag{30}$$

Here, it is $\left(\widehat{\sigma}^{(N)}\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(y^i - \sum_{k=0}^{K-1}\widehat{\beta}_k^{(N)} e_k\left(x^i\right)\right)^2$ with the ordinary least-squares estimator $\widehat{\boldsymbol{\beta}}^{(\boldsymbol{N})}$ from Equation (28) and the fitting points $\left(x^i, y^i\right),\ i = 1,\ldots,N$. For the derivation of this particular form of AIC, see Krah (2015).

This form of AIC depends positively on both the calibration error and the number of basis functions. This permits a very intuitive interpretation of AIC. As long as the number of basis functions is comparatively small and the model complexity thus low, the proxy function *underfits* the CFP model. Under these circumstances, it makes sense to increase the model complexity appropriately to reduce the calibration error. As a consequence,

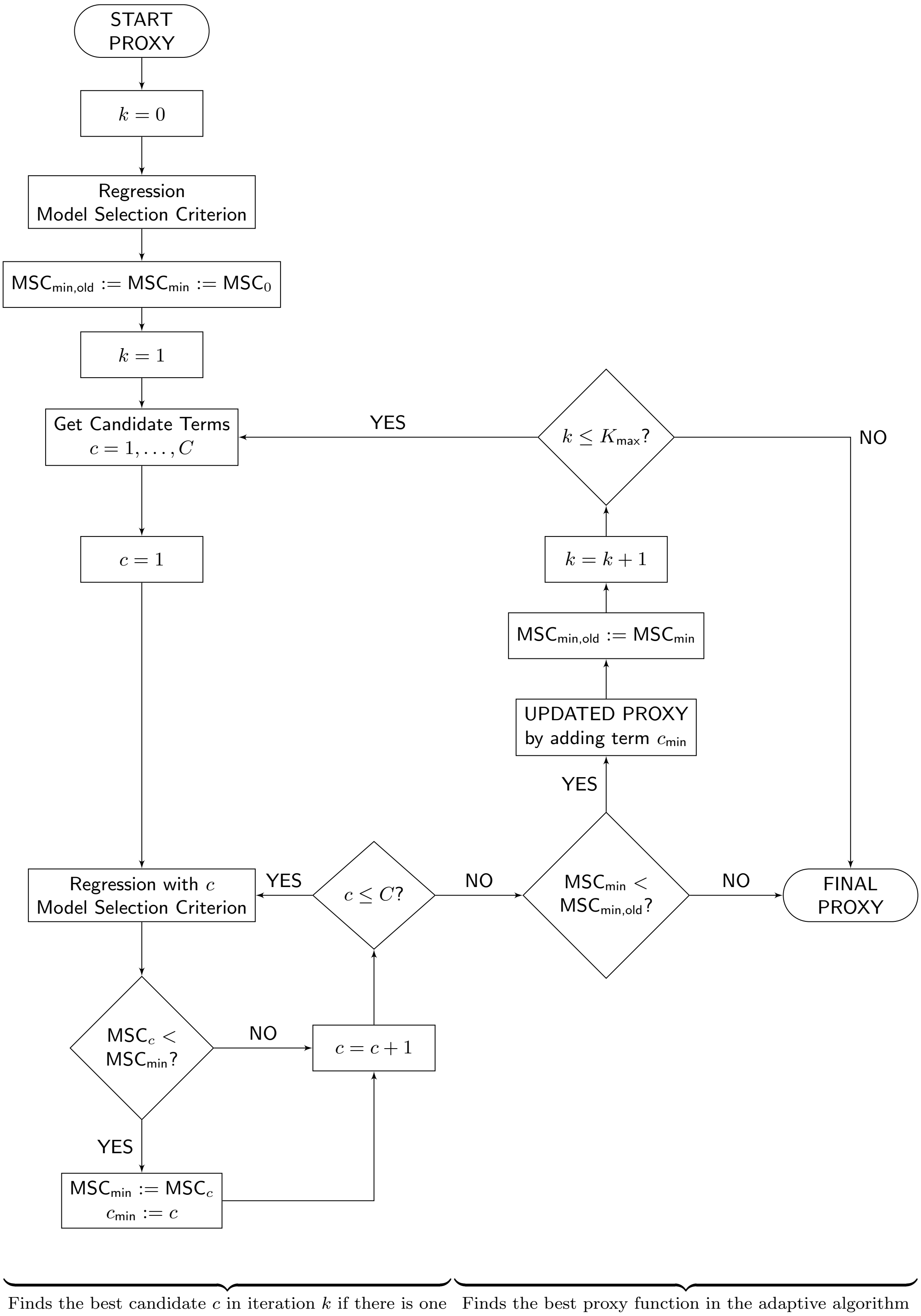

Figure 13: Flowchart of the calibration algorithm.

AIC decreases. We repeat this procedure unless AIC begins to stagnate or increase as such a behavior is a signal for a proxy function that *overfits* the CFP model. In this case, reductions in the calibration error are achieved if the proxy function approximates noise. Since such effects reduce the generalizability of the proxy function to new data, they should be avoided. This means, the task at hand is to find a proxy function with a comparatively low AIC score.

8.3.5 Iterative Procedure

After the initialization, the iterative nature of the adaptive algorithm comes into play. At the beginning of each iteration ($k = 1, \dots, K-1$), the candidate basis functions for the extension of the proxy function need to be determined according to a predefined principle. By providing the candidate basis functions, this principle also defines the function types which act as building blocks for the proxy function. Since we want the proxy function to be a multivariate polynomial function with the risk factors $X_1, \dots, X_d$ as the covariates, we let the principle generate monomial basis functions in these d risk factors. To keep the computational costs below an acceptable limit, we further restrict the candidate basis functions by conditioning them on the current proxy function term structures. The principle must be designed in a way that it derives candidate basis functions with possibly high explanatory power.

A principle which satisfies these properties and which we employ is the so-called *principle of marginality*. According to this principle, a monomial basis function may be selected if and only if all its partial derivatives are already part of the proxy function. For instance, after the initialization ($k = 1$), the proxy function term structure consists only of the constant basis function $e_0(X) = 1$. Since $1 = X_l^0,\ l = 1, \dots, d$, is the partial derivative of all linear monomials, the candidate basis functions of the first iteration are $\widetilde{e}_1^c(X) = X_c,\ c = 1, \dots, d$. To give another example, the monomial $X_1^2 X_2$ would become a candidate if the proxy function term structure had been extended by the basis functions $X_1 X_2$ and X_1^2 in the previous iterations. These basis functions would become candidates themselves if the linear monomials X_1 and X_2 had been selected before.

When the list of candidate basis functions is complete, the proxy function term structure is extended by the first candidate basis function $\widetilde{e}_k^1(X)$. An ordinary least-squares regression is performed based on the extended structure and AIC is calculated. If and only if this AIC score is smaller than the currently smallest AIC score of the present iteration, the latter AIC score is updated by the former one and the new proxy function memorized instead of the old one. This procedure is repeated separately with the remaining candidate basis functions $\widetilde{e}_k^c(X),\ c = 2, \dots, C$, one after the other. In Figure 13, this part of the algorithm can be viewed on the lower left side.

If there are candidates $\widetilde{e}_k^c(X),\ c = 1, \dots, C$, which let AIC decrease when being added to the proxy function, the one that lets AIC decrease most, say $\widetilde{c}$, is finally selected into the proxy function of the present iteration, i.e., $e_k(X) = \widetilde{e}_k^{\widetilde{c}}(X)$. The minimum AIC score of the entire adaptive algorithm is updated accordingly. The algorithm terminates and provides the final proxy function if either there is no candidate which lets AIC decrease or the prespecified maximum number of basis functions $K_{\max}$ could be exceeded in an additional iteration. Otherwise, the next iteration ($k+1$) starts by an update of the list of candidate basis functions.

8.3.6 Refinements

The adaptive algorithm illustrated in Figure 13 is an approach for the derivation of suitable proxy functions which has already proved its worth in practice when applied together with some refinements. The refinements are in particular needed if the algorithm has not generated a proxy function which passes the validation procedure. In this case, useful refinements can be restrictions that are made based on prior knowledge about the behavior of the CFP model. These restrictions might increase the accuracy of the proxy function or reduce the computational costs. For example, one can restrict the maximum powers to which the risk factors may be raised in the basis functions by defining joint or individual limits. Thereby, it could be reasonable to apply different rules for monomials with single and multiple risk factors. In addition, it is possible to specify a maximum allowed number of risk factors per monomial. The principle of marginality would have to be adjusted in accordance with these restrictions. Moreover, one could constrain the intercept or select some basis functions manually.

If there are risk factors that have no or negligible interactions with others, one can split the derivation of the proxy function into two or more parts to achieve better fits on each of these parts and finally merge the partial proxy functions. However, such an approach requires separate fitting points for each of these parts and thus an adjusted simulation setting in the first place. As long as such modifications do not become too time-consuming and stay transparent, they can serve as useful tools to increase the overall goodness of fit of the final proxy function or to eventually meet the validation criteria.

8.4 Proxy Function Validation

8.4.1 Validation Points

We cannot judge the quality of the derived proxy function only on the basis of the fitting points. In addition, we have to perform a validation on points with a lower standard error that have not entered the calibration procedure. Only if this works satisfactory, we can expect a good performance from the proxy function. To conduct the validation procedure, we adopt the filtered probability space and formalization of the available capital from the simulation setting. The idea of the validation is to check if the proxy function provides indeed approximately the available capital in Equation (23) conditional on the outer scenarios. Since the fitting values rely only on few inner scenarios per fitting scenario, they usually do not come close to (23) and are therefore not suitable for measuring the absolute goodness of fit of a proxy function. To measure the absolute goodness of fit, we derive validation points with sufficiently many inner scenarios per validation scenario.

To specify the *validation scenarios* x^i, $i = 1, \ldots, L$, considerations similar to those made above for the fitting scenarios are necessary. This means, we have to take into account the run time of the CFP model in the given hardware architecture while ensuring to choose enough validation scenarios for a reasonable validation. Moreover, the validation scenario budget must be harmonized with the fitting scenario budget. Additionally, the number of inner scenarios b per validation scenario needs to be set such that the resulting validation values approximate the expectation in Equation (23) sufficiently well. Lastly, the rather few validation scenarios have to be allocated in a way that the absolute goodness of fit of the proxy function is reliably measured. See Sections 3.4 and 3.5 for more details on fitting and validation scenarios.

Before the Monte Carlo simulations of the CFP model can be performed, the inner scenarios associated with the validation scenarios must be generated. Again, an ESG can take over this task. As a result from the simulations, we obtain the *validation values* $y^i,\ i = 1, \ldots, L$, analogously to the fitting values in Equation (26) by averaging the values in Equation (25), where L has to be substituted for N and b for a. In the end, we obtain the *validation points* $\left(x^i, y^i\right),\ i = 1, \ldots, L$.

The choice of validation scenarios is not an easy task. Since the number of validation scenarios L is limited, it is important to decide which scenarios give more insight into the quality of the fit. There exist different paradigms for their selection:

- Points known to be in the *capital region*, that is, scenarios producing a risk capital close to the SCR estimate from previous risk capital calculations;
- Quasi-random points from the entire fitting space;
- One-dimensional risks leading to a 1-in-200 loss in the one-dimensional distribution of this risk factor, that is, points which have only one coordinate changed and which ensure a good interpretability;
- Two- or three-dimensional stresses for risk factors with high interdependency, for example, interest rate and lapse; and
- Points with the same inner scenarios which can be used to more accurately measure a risk capital in scenarios which do not have the ESG relevant risk factors changed.

8.4.2 Practical Implementation

The number of validation scenarios obviously depends on the computational capacities of a company. Using more validation scenarios is always more accurate. It is important to strike a balance between the fitting and validation scenarios as well as between the number of outer and inner scenarios. In the industry, we have observed between $L = 15$ and $L = 200$ validation scenarios with, respectively, between $b = 1,000$ and $b = 16,000$ inner scenarios. This means that the fitting and validation computations can be split in half or the validation may get up to $3/4$ of the entire calculation budget. Depending on the insurer's complexity the choice needs to guarantee both reliable validations while remaining feasible.

Once we have selected the validation scenarios, we produce the corresponding bL inner scenarios by the same ESG which we have employed in the context of the fitting scenarios. After all Monte Carlo simulations have been run, we compute the validation values per validation scenario by averaging.

8.4.3 Out-of-Sample Test

To assess whether a proxy function reflects the CFP model sufficiently well, we specify three validation criteria. The first two criteria will be of quantitative nature whereas the third criterion will be more of qualitative nature. We let a proxy function pass the out-of-sample test if it fulfills at least two validation criteria. If a proxy function fails to meet exactly one validation criterion, in addition, a sound explanation needs to be given. If it fails more than one criterion, the proxy function calibration has to be refined and performed again.

The idea of our first validation criterion is to determine thresholds which the absolute deviations between the validation values and the proxy function conditional on the validation scenarios given by

$$\mathrm{dev}^i = \left| \frac{y^i - \widehat{\mathrm{AC}}^{(K,N)}\left(x^i\right)}{a^i} \right|, \ i = 1, \ldots, L, \tag{31}$$

should not exceed. Here, $\widehat{\mathrm{AC}}^{(K,N)}\left(x^i\right)$ is given by Equation (29). For measuring the relative deviation we use the *asset metric* a^i. This means that each absolute deviation is divided by the market value of assets a^i in scenario i. $|\cdot|$ denotes the absolute value. The asset metric is a more suitable denominator in the fraction above than the relative metric in the cases in which the approximated economic variable may take on very low values. This can occur for life business that is deeply in the money and of which the available capital can thus be very low. Another example can be a company selling exclusively term insurances, where the technical provisions can become small in comparison to the market value of assets.

We require that at least 90% of the validation points have deviations in Equation (31) not higher than 0.5% and that the remaining validation points have deviations of at maximum 1%. If a proxy function satisfies this condition, we say it satisfies the *first validation criterion*. Let the insurer have an equity-to-assets ratio of about 2%. Then, the deviations in Equation (31) for the available capital can be translated into the relative deviations as follows. If a validation point has a deviation with respect to the asset metric of between 0% and 0.5%, it has a relative deviation of between 0% and 25%. Accordingly, it has a relative deviation between 25% and 50% if its deviation with respect to the asset metric falls in between 0.5% and 1%.

For our second validation criterion, we define an overall measure for the goodness of fit of the proxy function by the normalized mean absolute error with respect to the asset metric, i.e.,

$$\mathrm{mae} = \frac{\sum_{i=1}^{L} \left| y^i - \widehat{\mathrm{AC}}^{(K,N)}\left(x^i\right) \right|}{\sum_{i=1}^{L} \left|a^i\right|}. \tag{32}$$

We say the *second validation criterion* is met if $\mathrm{mae} \leq 0.5\%$.

Further graphical analyses serve as a verification of the results. First, to check if the fitting values are at least roughly normally distributed, we create a histogram of the fitting values. Then, for each risk factor, we separately plot the fitting values together with the curve of the proxy function to see if the proxy function follows the behavior of the fitting values. Thereby, we vary the proxy function only in the respective risk factor and set all other risk factors equal to their base values. In the evaluation, we have to be aware of the fact that the proxy function usually also includes mixed monomials which might have no effect if one of the monomials' risk factors is equal to the base value. Additionally, for each risk factor, we separately create plots with selected validation points and the curve of the proxy function. These plots help us to verify if the proxy function behaves similarly to the validation points as well. For each risk factor, we thereby select only those validation points that are not stressed in the components different from the currently considered risk factor. If a proxy function follows both the fitting and validation points and shows a behavior that is consistent with our knowledge of the CFP model, we say it satisfies the *third validation criterion*.

8.5 Full Distribution Forecast

8.5.1 Solvency Capital Requirement

After the proxy function has been successfully validated, it can finally be used to produce the full loss distribution forecast. Based on this forecast, we are not only able to calculate the SCR as a value-at-risk but also to derive statistical figures related to other risk measures such as the expected shortfall. We take the set of *real-world scenarios* x^i, $i = 1, \ldots, R$, that needs to be drawn from the joint real-world distribution of the risk factors $\mathbb{P}$ as given, compare Section 3.3. A possibility to model the joint real-world distribution of the risk factors is the use of copulas, see e.g. Mai & Scherer (2012). In our context, $\mathbb{P}$ is typically modeled with the aid of historical data and expert judgment.

We obtain the *real-world values* of the available capital by evaluating the proxy function in Equation (29) at the real-world scenarios, i.e.,

$$\widehat{y}^i = \widehat{\mathrm{AC}}^{(K,N)}\left(x^i\right) = \sum_{k=0}^{K-1} \widehat{\beta}_k^{(N)} e_k\left(x^i\right),\ i = 1, \ldots, R. \tag{33}$$

Accordingly, we compute the real-world values of the profit in Equation (20), i.e.,

$$\widehat{\Delta}^i = \widehat{y}^i - \widehat{y}^0,\ i = 1, \ldots, R, \tag{34}$$

where $\widehat{y}^0 = \mathrm{AC}\left(x^0\right)$, with x^0 being the stress-neutral base scenario, denotes the estimate for the initial available capital.

We recall the definition of the SCR from the simulation setting in Equation (21) and replace the theoretical expressions with their empirical counterparts, i.e.,

$$\widehat{\mathrm{SCR}} = \widehat{\mathrm{VaR}}_\varphi\left(-\widehat{\Delta}\right) = \inf\left\{y \in \left\{-\widehat{\Delta}^1, \ldots, -\widehat{\Delta}^R\right\} \mid \widehat{F}^{\mathbb{P}}_{-\widehat{\Delta}}(y) \geq \varphi\right\}, \tag{35}$$

where $\widehat{\Delta}$ represents the real-world values of the profit and $\widehat{F}^{\mathbb{P}}_{-\widehat{\Delta}}(y) = \mathbb{P}\left(-\widehat{\Delta} \leq y\right)$ is the empirical distribution function of the loss under $\mathbb{P}$. By using the identity $\widehat{F}^{\mathbb{P}}_{-\widehat{\Delta}}(y) = \frac{1}{R}\sum_{i=1}^R \mathbf{1}_{-\widehat{\Delta}^i \leq y}$ for the empirical distribution function, Equation (35) becomes

$$\widehat{\mathrm{SCR}} = \inf\left\{y \in \left\{-\widehat{\Delta}^1, \ldots, -\widehat{\Delta}^R\right\} \mid \sum_{i=1}^R \mathbf{1}_{-\widehat{\Delta}^i \leq y} \geq \varphi R\right\}. \tag{36}$$

Hence, the SCR has to be greater than or equal to $\lceil \varphi R \rceil$ real-world losses. We estimate thus the SCR as the $\lfloor (1-\varphi) R \rfloor$ highest real-world loss.

8.5.2 Convergence

As already stated above, Proposition 1 and 2 for the convergence of the two approximations of the available capital also imply convergence in probability and distribution. Since $\mathbb{P}$ can be substituted for $\mathbb{P}'$, the convergence results are valid under both physical probability measures. The following two corollaries formalize the convergence in distribution.

Corollary 1. *$\widehat{F}^{\mathbb{P}}_{\widehat{\mathrm{AC}}^{(K,N)}}(y) = \mathbb{P}\left(\widehat{\mathrm{AC}}^{(K,N)} \leq y\right) \to \mathbb{P}\left(\mathrm{AC} \leq y\right) = F^{\mathbb{P}}_{\mathrm{AC}}(y)$ as $K, N \to \infty$ for $y \in \mathbb{R}$.*

Corollary 2. $\left(\widehat{F}^{\mathbb{P}}_{\widehat{\mathrm{AC}}^{(K,N)}}\right)^{-1}(\varphi) \to \left(F^{\mathbb{P}}_{\mathrm{AC}}\right)^{-1}(\varphi)$ *as* $K, N \to \infty$ *for all continuity points* $\varphi \in (0,1)$ *of* $\left(F^{\mathbb{P}}_{\widehat{\mathrm{AC}}}\right)^{-1}$.

Since Corollaries 1 and 2 hold as well for linear transformations of the available capital, we conclude that $\widehat{\mathrm{SCR}} \to \mathrm{SCR}$ as $K, N \to \infty$. Hence, the estimate for the SCR in Equations (35) and (36) converges to the theoretical SCR in Equation (21).

8.5.3 Practical Implementation

To model the joint real-world distribution of the risk factors, we can use a fully specified Gaussian copula and, for example, $R = 2^{17} = 131{,}072$ real-world scenarios. Then, we compute the corresponding real-world values of the available capital and profit according to Equations (33) and (34), respectively, to obtain a full probability distribution forecast of the loss.

The SCR is defined as the 99.5% value-at-risk of the loss distribution under the Solvency II directive. This is the reason why we set $\varphi = 99.5\%$ in Equations (35) and (36). As already mentioned above, this can be interpreted as a target for the insurance company to survive 199 out of 200 business years. Eventually, we calculate an estimate for the SCR by evaluating Equation (36). Independent of the data, we are able to characterize this estimate under the assumption $R = 131{,}072$ as the $\lfloor (1-\varphi)\, R \rfloor = 655$ th highest real-world loss.

Numerical calculations and examples illustrating our full approach can also be found in e.g. Bettels et al. (2014).

9 Numerical Illustration

9.1 Approximation Task

In this numerical example, we demonstrate how an actual application of our proposed LSMC-based approach might look like in practice and illustrate the convergence of the adaptive LSMC algorithm. We take the CFP model and real-world distribution of a German life insurer as given and stick to conveniently scaled best estimate liability (BEL) data that have already served as illustrations in Krah (2015). The exemplary insurer is exposed to $d = 14$ relevant capital market and actuarial risks from Table 3. By numbering these risks consecutively, i.e., $X_l,\ l = 1, \ldots, d$, we overwrite the notation of Table 3 for this example and therefore mainly disguise the meaning of the risks for keeping the anonymity of the exemplary insurer.

Let it be our task to find a proxy function for the insurer's BEL conditional on the specified risk factors to derive its SCR under Solvency II as the 99.5% value-at-risk of the corresponding loss distribution. A polynomial proxy function of the insurer's market value of assets can thereby be assumed to be known so that the available capital results can easily be extracted by taking the difference of the two functions. Given the complexity of the CFP model and the insurer's computational capacities, it is reasonable to run the CFP model for $N = 25{,}000$ fitting scenarios with each of these outer scenarios entailing $a = 2$ inner scenarios. While the fitting scenarios are defined such that they follow a linear transformation of the Sobol sequence, the components of the inner scenarios are partly generated by a suitable ESG and partly modeled directly as input in the CFP model.

9.2 Proxy Function Calibration

For the proxy function calibration, we apply the adaptive algorithm depicted in the flowchart of Figure 13 and set the maximum number of basis functions at $K_{\max} = 100$. As the start proxy function, we use a constant function. In Table A1, the start proxy function term structure is represented by the first row ($k = 0$) containing the basis function $e_0(X) = \prod_{l=1}^{14} X_l^{r_0^l} = 1$ since $r_0^1 = ... = r_0^{14} = 0$. Then, the initial regression is performed and the selected model selection criterion, here AIC, evaluated. For the initial AIC score in this example, see the AIC entry in the first row of Table A1. It is a high value indicating the obvious fact that a constant poorly describes the possible changes of the insurer's BEL.

Now that the initialization has been completed, the iterative part of the algorithm relying on the principle of marginality is carried out. As mentioned in the previous section, in the first iteration ($k = 1$), the candidate basis functions are just the linear functions of all risk factors. The second row of Table A1 indicates that the proxy function is extended by candidate $e_1(X) = X_8$ in this iteration, meaning that risk factor X_8, the credit default stress, has the highest explanatory power in terms of AIC among the candidates. For the AIC score corresponding to the updated proxy function term structure $e_0(X) + e_1(X)$, see the AIC entry in the second row. In the second iteration ($k = 2$), in addition to the linear functions of the remaining risk factors, the quadratic function of the credit default stress becomes a candidate basis function. However, as we can see in the third row of Table A1, risk factor X_6, the equity market value stress, is selected next as this risk factor complements the existing proxy function best in this iteration. The algorithm continues this way until iteration $k = 61$ in which no further candidate basis function lets AIC decrease anymore. The sequence in which the basis functions are selected into the proxy function and the corresponding course of the AIC scores are reported in Table A1. The coefficients denoted in this table belong to the final proxy function emerging from iteration $k = 60$. We can see that, except for the risks X_5 and X_9, all risk factors contribute to the explanation of BEL. With $K = 61$ and $e_k(X) = \prod_{l=1}^{14} X_l^{r_k^l}$, $k = 0, \dots, K-1$, the general expression of Equation (29) for the proxy function conditional on any outer scenario X becomes in this example

$$\widehat{\mathrm{BEL}}^{(K,N)}(X) = \sum_{k=0}^{K-1} \left(\widehat{\beta}_k^{(N)} \prod_{l=1}^{14} X_l^{r_k^l} \right). \tag{37}$$

9.3 Proxy Function Validation

We perform the proxy function validation based on two different sets of validation points to highlight the impact of these choices and depart for reasons of simplification slightly from the extensive validation procedure described in the previous section. In conjunction with the calculation budget for the fitting computations, we let Set 1 comprise $L = 51$ and Set 2 comprise $L = 56$ validation scenarios with each validation scenario entailing $b = 1,000$ inner scenarios. While Set 1 contains the stress-neutral base point, 26 properly transformed multi-dimensional stress points and 24 one-, two- or higher dimensional points, Set 2 contains only four equidistant one-dimensional stresses for each risk factor. Among the two sets of validation points, Set 1 can be viewed as the more sophisticated one.

The column "Out-of-Sample MSE" on the right-hand side of Table A1 contains the evolution of the mean squared errors associated with Set 1 and Set 2 over the iterations.

Typically, the mean squared errors decrease together with AIC over the iterations. However, since the sets of validation points are small and incur random fluctuations, they cannot fully reflect the goodness of fit of a proxy function. Therefore, we do not expect the mean squared errors to decrease monotonously over all iterations. Nevertheless, Table A1 shows well the trend of the diminishing impact of each additional basis function in explaining the dependencies of the BEL.

The plots of the one-dimensional curves of the proxy function together with the respective one-dimensional validation points confirm the good approximation quality of the proxy function. For exemplary plots with respect to risk factors X_1 and X_8, see Figures 14 and 15.

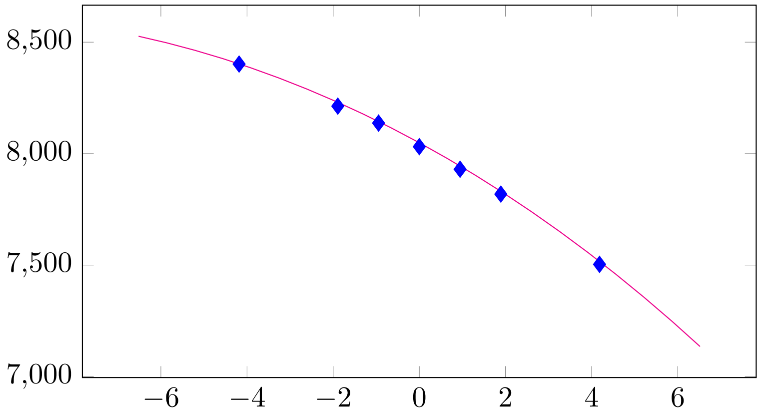

Figure 14: One-dimensional curve of the proxy function and validation points with respect to X_1.

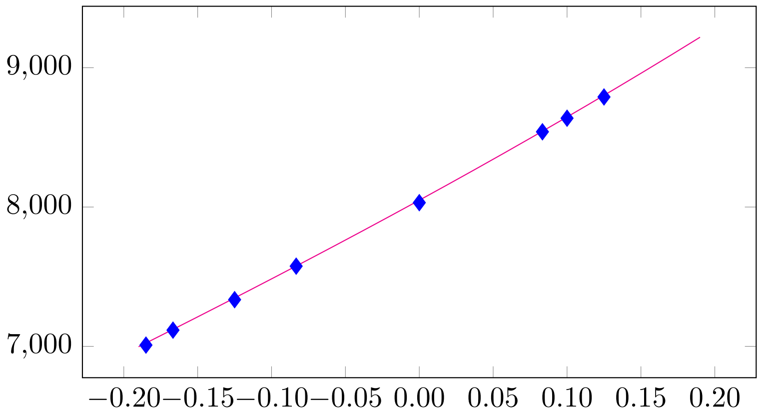

Figure 15: One-dimensional curve of the proxy function and validation points with respect to X_8.

Furthermore, by looking at descriptive statistics on the relative deviations between the validation values and proxy function predictions (e.g., quantiles, mean, and median), the proxy function admits a satisfying performance on the validation sets. In order to keep the anonymity of the insurer, we do not report exact numbers here.

Whether the proxy function is successfully validated by these simplified out-of-sample tests and can thus be passed on to the last step of the LSMC framework, or whether the calibration needs to be repeated with some refinements depends on the exact validation criteria which we have not further specified in this example.

9.4 Full Distribution Forecast

As the next step, we consider the forecast of the full distribution of the losses. Taking up the copula approach of Section 8.5, we have generated $R = 2^{17} = 131,072$ real-world scenarios from the joint distribution of the risk factors. The evaluation at each of these scenarios is simple, as it amounts to the evaluation of two polynomial proxy functions (the derived one of BEL and the given one of the market value of assets) and taking their difference. After ordering the resulting losses by size, we are able to calculate empirically the SCR as the corresponding 99.5%-quantile.

For our setting, we have furthermore performed a comparison between the SCR determined with the LSMC approach and a partial nested simulations approach, that is, one using an explicit valuation with the CFP model, compare Section 5.1. Krah et al. (2018) show in Chapter 5 that for a related company the scenarios leading to the 99.5%-quantile in the LSMC approach and nested simulations approach are similar. This means that, based on the LSMC algorithm, we can define a set of, for instance, 50 real-world scenarios leading to losses close to the SCR and feed them into the CFP model. Such a valuation is similar to the out-of-sample validation which is why we use again $1,000$ inner simulations per outer scenario. Finally, comparing the 50 LSMC-based losses with the 50 nested simulations losses unveils potential differences in the SCR estimation for the two approaches. This difference turns out to be only about 3%, which corresponds to a 0.5% difference in the available capital.

9.5 Computation Time

Given the required task, the number of real-world scenarios, the duration of one simulation in the CFP model and the computational capacities of the German life insurer, we can easily determine good estimates for the overall computation times of the two opposing approaches. To simplify the calculation of the computation times, we merge the two validation sets 1 and 2 to one validation set with $L = 51+56 = 107$ scenarios. Furthermore, it takes $\tau_{\text{MC}} = 55$ seconds to run one simulation of the CFP model and the simulations can be allocated to up to $\nu_{\text{CPU}} = 476$ CPUs.

For the nested simulations approach, we assign the same number of inner simulations to each real-world scenario as we did in the LSMC-based approach to each validation scenario. The computation time of the latter approach is mainly driven by two factors: the time needed to carry out the Monte Carlo simulations of the CFP model and the time needed to calibrate the proxy function. The calibration takes $\tau_{\text{calib.}} = 45$ min in our example and can be further reduced by applying more efficient implementation techniques (e.g., recompute only subblocks of the design matrix in the candidate loop, and use parallelization). In contrast, the validation involves only the fast computation of selected figures in the out-of-sample test and the full probability distribution forecast requires only the fast evaluation of the resulting proxy function at the real-world scenarios as well as the subsequent ordering of the available capital estimates. Together, these calculations take only a few seconds. In the nested simulations approach, the computation time is solely driven by one factor: the time for the Monte Carlo simulations. The ordering of the directly resulting available capital estimates for the value-at-risk computation is negligible here as well. The computation times are then, respectively,

$$\tau_{\text{LSMC}} = (aN + bM)\,\tau_{\text{MC}}/\nu_{\text{CPU}} + \tau_{\text{calib.}} < 6 \text{ h},$$
$$\tau_{\text{nest.stoch.}} = bR\tau_{\text{MC}}/\nu_{\text{CPU}} \approx 25 \text{ weeks},$$

where $N = 25,000$, $a = 2$, $L = 107$, $b = 1,000$ and $R = 131,072$. These are pure run times for the projections, on top of this 30–40% additional run time is needed for splitting the runs, saving and merging the results per scenarios as well as reading out the figures. Hence, the LSMC-based approach is even feasible for the exemplary life insurer if the calibration and validation procedures have to be repeated several times until the specified validation criteria are finally met, whereas the full nested simulations approach is not feasible at all.

10 Conclusion

Summary

In the foregoing sections, we have in detail derived the theoretical foundations of the LSMC approach for proxy modeling in the life insurance business. We have described all the necessary steps to a reliable implementation of the LSMC method in practice. They mainly consist of a detailed description of the simulation setting and the required task, a concept for a calibration procedure for the proxy function, a validation procedure for the obtained proxy function, and the actual application of the LSMC model to forecast the full loss distribution. In addition, we have presented a slightly disguised real-world application of the LSMC approach for illustration.

Outlook

The most intensive research is ongoing in the area of proxy modeling. In the standard approach presented above, ordinary monomial basis functions are used. Instead, various orthogonal polynomial basis functions such as Laguerre, Legendre, Hermite or Chebyshev polynomials can be employed, see e.g. Teuguia et al. (2014). To model non-smooth behaviors or other significant patterns of the underlying CFP model, the set of possible basis functions could be extended by other function types such as rational, algebraic, transcendental, composite or piecewise functions, where transcendental functions include exponential, logarithmic, power, periodic and hyperbolic functions. In cases where a function type is not compatible with the principle of marginality, an adjustment thereof or a new principle is required.

Additionally, other model selection criteria than AIC can be implemented. Examples are non-parametric cross-validation, the Bayesian information criterion (BIC), Mallows's C_p or the Takeuchi information criterion (TIC). Mallows's C_p is named after Mallows (1973) and has been shown by Boisbunon et al. (2014) to be equivalent to AIC under the normal distribution assumption. TIC has been introduced by Takeuchi (1976) as a generalization of AIC, which arises from TIC when the true distribution of the phenomenon is contained in the assumed parametric distribution family. As already stated above, for a comparison of AIC, BIC, Mallows's C_p, adjusted R^2, cross-validation and of model selection criteria relying on F-tests within the same LSMC application, see e.g. Reichenwallner (2014).

Besides these modifications, the ordinary least-squares regression method can be replaced by other regression techniques such as ridge, robust or feasible generalized least-squares regression. Moreover, generalized linear models or generalized additive models are options. Even stochastic alternatives replacing the entire adaptive algorithm such as artificial neural networks or decision tree learning are conceivable. A deep analysis of deterministic regression variants that are combinable with the adaptive algorithm can be found in the next part of this thesis. Additionally, an overview of results obtained by stochastic alternatives is given therein.

Part II

Machine Learning in Least-Squares Monte Carlo Proxy Modeling of Life Insurance Companies

Résumé

Under the Solvency II regime, life insurance companies with internal models are asked to derive their solvency capital requirements from their full loss distributions over the coming year. Since the industry is currently far from being endowed with sufficient computational capacities to fully simulate these distributions, the insurers have to rely on suitable approximation techniques such as the least-squares Monte Carlo (LSMC) method. The key idea of LSMC is to run only a few wisely selected simulations and to process their output further to obtain a risk-dependent proxy function of the loss. In this part of the thesis, we present and analyze various adaptive machine learning approaches that can take over the proxy modeling task. The studied approaches range from ordinary and generalized least-squares regression variants over GLM and GAM methods to MARS and kernel regression routines. We justify the compatibility of their regression ingredients in a theoretical discourse. Furthermore, we illustrate the approaches in slightly disguised real-world experiments and perform comprehensive out-of-sample tests.

11 Introduction

General Remarks

To keep this part self-contained, we repeat the most relevant concepts from the previous part before we get to its edge again. We not only summarize the central themes so far but also refine them where conducive to the research in this part. Thereby, we also point out how this part is related to the previous one. Mainly, the repetitions concern this and the subsequent section.

LSMC Framework under Solvency II

By the Solvency II directive of European Parliament & European Council (2009), life insurance companies are asked to derive their solvency capital requirements (SCRs) from their full loss probability distributions over the coming year if they do not want to rely on the much simpler standard formula. In order to obtain reasonably accurate full loss distributions via a nested simulations approach as described in Bauer et al. (2012), their cash-flow-projection (CFP) models would need to be simulated several hundred thousand times. But the insurers are currently far from being endowed with sufficient computational capacities to perform such expensive simulation tasks. By applying well-suited approximation techniques such as the least-squares Monte Carlo (LSMC) approach by Cathcart (2012) and Bauer & Ha (2015), the insurers are able to overcome these computational hurdles. For example, they can implement the LSMC framework formalized by Krah et al. (2018), see the first part of this thesis, and applied by e.g. Bettels et al. (2014). The central idea of this framework is to carry out a comparatively small number of wisely chosen Monte Carlo simulations and to feed the simulation results into a supervised machine learning algorithm that translates the results into a proxy function of the insurer's loss (output) with respect to the underlying risk factors (input). To guarantee a certain approximation quality, the proxy function has to pass a subsequent validation procedure before it can finally be used for the full loss probability distribution forecast.

Machine Learning Calibration Algorithm

Apart from the calibration and validation steps, we adopt the LSMC framework from Krah et al. (2018) and the first part of this thesis without any changes. Therefore, we neither repeat the simulation setting nor the procedure for the full loss distribution forecast and SCR calculation here in detail. The purpose of this part is to introduce various deterministic machine learning methods that can be applied in the calibration step of the LSMC framework and other high-dimensional variable selection applications, to point out their similarities and differences and to compare their out-of-sample performances in a slightly disguised real-world LSMC example. The majority of this part of the thesis has already been published in Krah et al. (2020*a*). We describe the data basis used for calibration and validation in Section 12.1, the structure of the calibration algorithm in Section 12.2 and the validation approach in Section 12.3. Our focus lies on out-of-sample performance rather than computational efficiency as the latter becomes only relevant if the former gives reason for it. We analyze a very realistic data basis with 15 risk factors and validate the proxy functions based on a very comprehensive and computationally expensive nested simulations test set comprising the SCR estimate.

The main idea of our approach is to combine classical regression methods with an adaptive algorithm, in which the proxy functions are built up of basis functions in a stepwise fashion. In a four risk factor LSMC example, Teuguia et al. (2014) applied a full model approach, forward selection, backward elimination and a bidirectional approach as, for example, discussed in Hocking (1976) with orthogonal polynomial basis functions. They stated that only forward selection and the bidirectional approach were feasible when the number of risk factors or the polynomial degree exceeded 7 as then the resulting other models exploded. Life insurance companies covering a wide range of contracts in their portfolio are typically exposed to even more risk factors like, for instance, 15. In complex business regulation frameworks such as in Germany, they furthermore often require polynomial degrees of at least 4. In these cases, even the standard forward selection and bidirectional approaches become infeasible as the sets of candidate terms from which the basis functions are chosen will explode then as well. We therefore follow the suggestion of Krah et al. (2018) to implement the so-called principle of marginality, an iteration-wise update technique of the set of candidate terms that lets the algorithm get along with comparatively few carefully selected candidate terms.

Regression Methods & Model Selection Criteria

Our main contribution is to identify, explain and illustrate a collection of regression methods and model selection criteria from the variety of regression design options that provide suitable proxy functions in the LSMC framework when applied in combination with the principle of marginality. After some general remarks in Section 13.1, we describe ordinary least-squares (OLS) regression in Section 13.2, generalized linear models (GLMs) by Nelder & Wedderburn (1972) in Section 13.3, generalized additive models (GAMs) by Hastie & Tibshirani (1986) and Hastie & Tibshirani (1990) in Section 13.4, feasible generalized least-squares (FGLS) regression in Section 13.5, multivariate adaptive regression splines (MARS) by Friedman (1991) in Section 13.6, and kernel regression by Watson (1964) and Nadaraya (1964) in Section 13.7. At the end of each section, we recap the assumptions, properties and estimation algorithms in a short summary. While some regression methods such as OLS and FGLS regression or GLMs can immediately be applied in conjunction with numerous model selection criteria such as Akaike information criterion (AIC), Bayesian information criterion (BIC), Mallow's C_P or generalized cross-validation (GCV), other regression methods such as GAMs, MARS, kernel, ridge or robust regression require well thought-through modifications thereof or work only with non-parametric alternatives such as k-fold or leave-one-out cross-validation. For adaptive approaches of FGLS, ridge and robust regression in life insurance proxy modeling, see also Hartmann (2015), Krah (2015) and Nikolić et al. (2017), respectively.

In the theory sections, we present the models together with their assumptions, important properties and popular estimation algorithms and demonstrate how they can be embedded in the adaptive algorithm by proposing feasible implementation designs and combinable model selection criteria. While we shed light on the theoretical basic concepts of the models to lay the groundwork for the application and interpretation of the later following numerical experiments, we forgo describing in detail technical enhancements or peculiarities of the involved algorithms and instead refer the interested reader to further sources. Additionally we provide the practitioners with R packages containing useful implementations of the presented regression routines. We complement the theory sections by practice sections 14.1–14.7, throughout which we perform the same Monte Carlo approximation task to make the performance of the various methods comparable. We measure the approximation quality of the resulting proxy functions by means of aggregated validation figures on three out-of-sample test sets. Again, we summarize the results obtained with each routine at the end.

Stochastic Machine Learning Alternatives

Conceivable alternatives to the entire adaptive algorithm are also stochastic machine learning techniques such as artificial neural networks (ANNs), decision tree learning or support vector machines. In particular, the classical feed forward networks proposed by Hejazi & Jackson (2017) and applied in various ways by Kopczyk (2018), Castellani et al. (2018), Born (2018) and Schelthoff (2019) were shown to capture the complex nature of CFP models well. A major challenge here is to find reliable hyperparameters such as the numbers of hidden layers and nodes in the network, batch size, weight initializer probability distribution, learning rate or activation function. Since the random seed used in the training of a network can be crucial for finding the global optimum, it should be considered as a hyperparameter choice as well. Future research should thus be dedicated to hyperparameter search algorithms and, as a means of mitigation thereof, stabilization methods such as ensemble methods. A starting point for this kind of research, going beyond the scope of this thesis, can already be found in Krah et al. (2020*b*). As an alternative to feed forward networks, Kazimov (2018) suggested to use radial basis function networks albeit so far none of the tested approaches worked out well.

In decision tree learning, random forests and tree-based gradient boosting machines were considered by Kopczyk (2018) and Schoenenwald (2019). While random forests were outperformed by feed forward networks but did better than the least absolute shrinkage and selection operator (LASSO) by Tibshirani (1996) in the example of the former author, they generally performed worse than the adaptive approaches by Krah et al. (2018) with OLS regression in numerous examples of the latter author. The gradient boosting machines, requiring more parameter tuning and thus being more versatile and demanding, came overall very close to the adaptive approaches. The tree-based methods belong by definition to the aforementioned ensemble methods, a modeling concept transferrable to arbitrary regression techniques, mitigating random model artefacts through averaging.

Castellani et al. (2018) compared support vector regression (SVR) by Drucker et al. (1997) to ANNs and the adaptive approaches by Teuguia et al. (2014) in a seven risk factor example and found the performance of SVR placed somewhere between the other two approaches with the ANNs getting closest to the nested simulations benchmark. As some further non-parametric approaches, Sell (2019) tested least-squares support-vector machines (LS-SVM) by Suykens & Vandewalle (1999) and shrunk additive least-squares approximations (SALSA) by Kandasamy & Yu (2016) in comparison to ANNs and the adaptive approaches by Krah et al. (2018) with OLS regression. In his examples, SALSA was able to beat the other two approaches whereas LS-SVM was left far behind. The analyzed machine learning alternatives have in common that they require at least to some degree a fine-tuning of some model hyperparameters. Since this is often a non-trivial but crucial task for generating suitable proxy functions, finding efficient hyperparameter search algorithms should become a subject of future research.

12 Calibration & Validation in the LSMC Framework

12.1 Fitting & Validation Points

12.1.1 Outer Scenarios & Inner Simulations

Our starting point is the LSMC approach from Part I. LSMC proxy functions are calibrated conditional on the *fitting points* generated by the Monte Carlo simulations of the CFP model. Additional out-of-sample *validation points* serve as a means for an assessment of the goodness of fit. The explaining variables of a proxy function are financial and actuarial risks the insurance company is exposed to. Examples for these risks are changes in interest rates, equity, credit, mortality, morbidity, lapse and expense levels over the one-year period. The dependent variable is an economic variable like the available capital, loss of available capital or the best estimate liability over the one-year period. Figure 16 plots the fitting values of an exemplary economic variable with respect to a financial risk

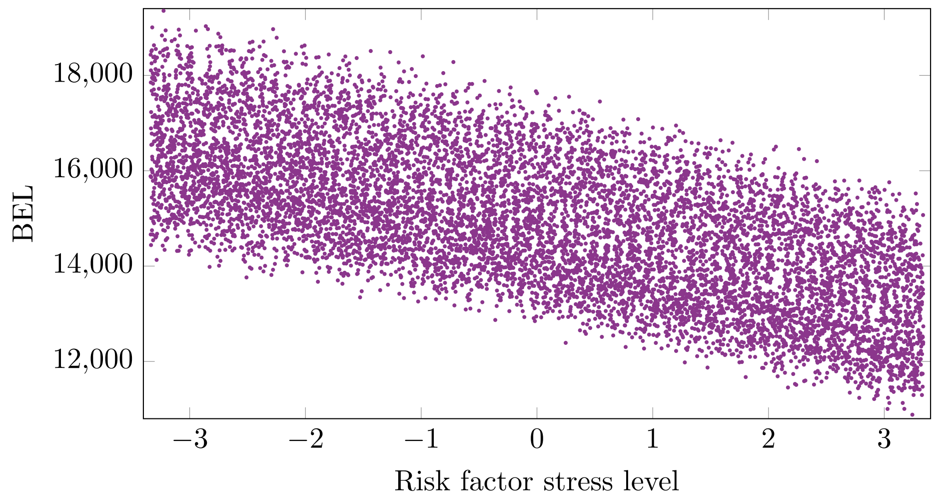

Figure 16: Fitting values of best estimate liability with respect to a financial risk factor.

factor. By an *outer scenario* we refer to a specific stress level combination of these risk factors, and by an *inner simulation* to a stochastic path of an outer scenario in the CFP model under the given risk-neutral probability measure. Each outer scenario is assigned the probability weighted mean value of the economic variable over the corresponding inner simulations. In the LSMC context, the fitting values are the mean values over only few inner simulations whereas the validation values are derived as the mean values over many inner simulations.

12.1.2 Different Trade-off Requirements

According to the law of large numbers, this construction makes the validation values comparatively stable while the fitting values are very volatile. Typically, the very limited fitting and validation simulation budgets are of similar sizes. Hence the few inner simulations in the case of the fitting points allow a great diversification among the outer scenarios whereas the many inner simulations in the case of the validation points let the validation values be quite close to their expectations but at the cost of only little diversification among the outer scenarios. These opposite ways to deal with the trade-off between the numbers of outer scenarios and inner simulations reflect the different requirements for the fitting and validation points in the LSMC approach. While the fitting scenarios should cover the domain of the real-world scenarios well to serve as a good regression basis, the validation values should approximate the expectations of the economic variable at the validation scenarios well to provide appropriate target values for the proxy functions.

12.2 Calibration Algorithm

12.2.1 Five Major Components

The calibration of the proxy function is performed by an adaptive algorithm that can be decomposed into the following five major components: (1) a set of allowed basis function types for the proxy function, (2) a regression method, (3) a model selection criterion, (4) a candidate term update principle, and (5) the number of steps per iteration and the directions of the algorithm. For illustration, a flowchart of the adaptive algorithm is depicted in Figure 13 of the previous part of this thesis. While components (1) and (5) enter the flowchart implicitly through the start proxy function, candidate terms and the order of the processes and decisions in the chart, components (2), (3) and (4) are explicitly indicated through the labels "Regression", "Model Selection Criterion" and "Get Candidate Terms".

Let us briefly recapitulate the state-of-the-art choices of components (1)–(5) in the insurance industry that we have introduced in Part I. As the function types for the basis functions (1), let only monomials be permitted. Let the regression method (2) be ordinary least-squares (OLS) regression and the model selection criterion (3) be Akaike information criterion (AIC) from Akaike (1973). Let the set of candidate terms (4) be updated by the principle of marginality to which we will return in greater detail below. Lastly, when building up the proxy function iteratively, let the algorithm make only one step per iteration in the forward direction (5) meaning that in each iteration exactly one basis function is selected which cannot be removed anymore (adaptive forward stepwise selection).

12.2.2 Iterative Procedure

The algorithm starts in the upper left side of Figure 13 with the specification of the start proxy basis functions. We specify only the intercept so that the first regression ($k = 0$) reduces to averaging over all fitting values. In order to harmonize the choices of OLS regression and AIC, we assume that the errors are normally distributed and homoscedastic because then the OLS estimator coincides with the maximum likelihood estimator. AIC is a relative measure for the goodness of fit of a proxy function and defined as twice the negative of the maximum log-likelihood plus twice the number of degrees of freedom. The smaller the AIC score, the better the fit, and thus the trade-off between a too complex (overfitting) and too simple model (underfitting).

At the beginning of each iteration ($k = 1, \dots, K - 1$), the set of candidate terms is updated by the *principle of marginality* which stipulates that a monomial basis function becomes a candidate if and only if all its derivatives are already included in the proxy function. The choice of a monomial basis is compatible to the principle of marginality. Using such a principle saves computational costs by selecting the basis functions conditionally on the current proxy function structure. In the first iteration ($k = 1$), all linear monomials of the risk factors become candidates as their derivatives are constant values which are represented by the intercept.

The algorithm proceeds on the lower left side of the flowchart with a loop in which all candidate terms are separately added to the proxy function term structure and tested with regard to their additional explanatory power. With each candidate, the fitting values are regressed against the fitting scenarios and the AIC score is calculated. If no candidate reduces the currently smallest AIC score, the algorithm is terminated, and otherwise, the proxy function is updated by the one which reduces AIC most. Then the next iteration $(k+1)$ begins with the update of the set of candidate terms, and so on. As long as no termination occurs, this procedure is repeated until the prespecified maximum number of terms $K_{\max}$ is reached.

12.3 Validation Figures

12.3.1 Validation Sets

Since it is the objective of this part of the thesis to propose suitable regression methods for the proxy function calibration in the LSMC framework, we introduce several validation figures serving as indicators for the approximation quality of the proxy functions. Primarily, we measure the out-of-sample performance of each proxy function based on three different validation sets by calculating five validation figures per set. In addition, we provide for the best performing proxy functions the five validation figures based on further reduced validation sets which allow us to draw conclusions for settings in which no extrapolation takes place.

The three validation sets are a Sobol set, a nested simulations set and a capital region set. Unlike the Sobol set, the nested simulations and capital region sets do not serve as feasible validation sets in the LSMC routine as they become known only after evaluating the proxy function for the real-world loss distribution forecast. Furthermore, they require massive computational capacities. Yet they can be regarded as the natural benchmark for the LSMC-based method and are thus very valuable for this analysis. Figure 17 plots

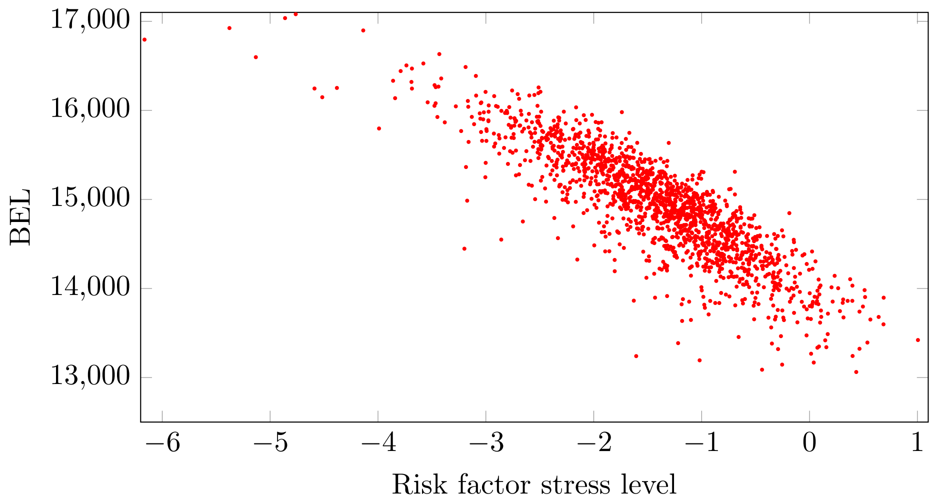

Figure 17: Nested simulation values of best estimate liability with respect to a financial risk factor.

the nested simulation values of an exemplary economic variable with respect to a financial risk factor. The Sobol set consists of, for example, between $L = 15$ and $L = 200$ Sobol validation points, of which the scenarios follow a Sobol sequence covering the fitting space uniformly. The fitting space is the cube on which the outer fitting scenarios are defined, and which has to cover the space of real-world scenarios used for the full loss distribution forecast sufficiently well. For interpretive reasons, sometimes the Sobol set is extended by points with, for example, one-dimensional risk scenarios or scenarios producing a risk capital close to the SCR (= 99.5% value-at-risk) in previous risk capital calculations.

The nested simulations set comprises, for example, between $L = 820$ and $L = 6,554$ validation points of which the scenarios correspond to, for example, the highest 2.5% to 5% losses from the full loss distribution forecast made by the proxy function that had been derived under the standard calibration algorithm choices described in Section 12.2. Like in the example of Chapter 5.2 in Krah et al. (2018), the order of these losses - which scenarios lead to which quantiles - following from the fourth and last step of the LSMC approach is very similar to the order following from the nested simulations approach. Therefore the scenarios of the nested simulations set are simply chosen by the order of the losses resulting from the LSMC approach. Several of these scenarios consist of stresses falling out of the fitting space. Compare Figures 16 and 17 which depict fitting and nested simulation values from the same proxy modeling task with respect to the same risk factor. Few points with severe outliers due to extreme stresses far beyond the fitting space should be excluded from the set. The capital region set is a subset of the nested simulations set containing the nested simulations SCR estimate, that is, the scenario leading to the 99.5% loss, and the, for example, 64 losses above and below, which makes in total $L = 129$ validation points.

12.3.2 Validation Figures

The five validation figures reported in our numerical experiments comprise two normalized mean absolute errors (MAEs), one with respect to the magnitude of the economic variable itself and one with respect to the magnitude of the corresponding market value of assets. They comprise further the mean error, that is, the mean of the residuals, as well as two validation figures based on the change of the economic variable from its base value (see the definition of the base value below): the normalized MAE with respect to the magnitude of the changes and the mean error of these changes. While the first three validation figures measure how well the proxy function reflects the economic variable in the CFP model, the latter two address the approximation effects on the SCR.

The smaller the normalized MAEs are, the better the proxy function approximates the economic variable. However, the validation values are afflicted with Monte Carlo errors so that the normalized MAEs serve only as meaningful indicators as long as the proxy functions do not become too precise. The mean errors should be preferably close to zero since they indicate systematic deviations of the proxy functions.

Let us write the absolute value as $|\cdot|$ and let L denote the number of validation points. Then we can express the MAE of the proxy function $\widehat{f}\left(x^i\right)$ evaluated at the validation scenarios x^i versus the validation values y^i as $\frac{1}{L}\sum_{i=1}^{L}\left|y^i - \widehat{f}\left(x^i\right)\right|$. After normalizing the MAE with respect to the mean of the absolute values of the economic variable or the market value of assets, i.e., $\frac{1}{L}\sum_{i=1}^{L}\left|d^i\right|$ with $d^i \in \left\{y^i, a^i\right\}$, we obtain the first two validation figures, i.e.,

$$\text{mae} = \frac{\sum_{i=1}^{L}\left|y^i - \widehat{f}\left(x^i\right)\right|}{\sum_{i=1}^{L}\left|d^i\right|}. \tag{38}$$

In the following, we will refer to (38) with $d^i = y^i$ as the MAE with respect to the *relative metric*, and to (38) with $d^i = a^i$ as the MAE with respect to the *asset metric*. The mean error is given by

$$\text{res} = \frac{1}{L}\sum_{i=1}^{L}\left(y^i - \widehat{f}\left(x^i\right)\right). \tag{39}$$

Let us refer by the base value y^0 to the validation value corresponding to the base scenario x^0 in which no risk factor has an effect on the economic variable. In analogy to (38) but only with respect to the relative metric, we introduce another normalized MAE by

$$\text{mae}^0 = \frac{\sum_{i=1}^{L}\left|\left(y^i - y^0\right) - \left(\widehat{f}\left(x^i\right) - \widehat{f}\left(x^0\right)\right)\right|}{\sum_{i=1}^{L}\left|y^i - y^0\right|}. \tag{40}$$

The corresponding mean error is given by

$$\text{res}^0 = \frac{1}{L}\sum_{i=1}^{L}\left(\left(y^i - y^0\right) - \left(\widehat{f}\left(x^i\right) - \widehat{f}\left(x^0\right)\right)\right). \tag{41}$$

In addition to these five validation figures, let us define the base residual which can be used as a substitute for (41) depending on personal taste. The base residual can easily be extracted from (39) and (41) by

$$\text{res}^{\text{base}} = y^0 - \widehat{f}\left(x^0\right) = \text{res} - \text{res}^0. \tag{42}$$

13 Machine Learning Regression Methods

13.1 General Remarks

As the main part of our work, we will compare various types of machine learning regression approaches for determining suitable proxy functions in the LSMC framework. The deterministic methods we present in this section range from ordinary and generalized least-squares regression variants over GLM and GAM methods to multivariate adaptive regression splines and kernel regression routines. The performance of the newly derived proxy functions when applied to the validation sets is one way of how to judge the different methods. Their compatibility with the principle of marginality and a suitable model selection criterion such as AIC to compare iteration-wise the candidate models inside the approaches is another way.

Our aim in the calibration step below is to estimate the conditional expectation of the economic variable $Y(X)$ under the risk-neutral probability measure $\mathbb{Q}$ given an outer scenario X. In this part, we use the notation $Y(X)$ as the economic variable does not necessarily have to be the available capital but can instead be, for example, the best estimate liability or market value of assets. The proxy modeling of $Y(X)$ involves two approximations because the sets of basis functions and fitting points are finite. The d-dimensional fitting scenarios are always generated under the physical probability measure $\mathbb{P}'$ on the fitting space which itself is a subspace of $\mathbb{R}^d$.

13.2 Ordinary Least-Squares (OLS) Regression

13.2.1 Classical Linear Regression Model

In iteration $K-1$ of the adaptive forward stepwise algorithm (as given in Section 12.2), the ordinary least-squares (OLS) approximation consists of a linear combination of suitable linearly independent basis functions $e_k(X) \in L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{P}'\right),\ k = 0, 1, \ldots, K-1$, i.e.,

$$Y(X) \overset{K<\infty}{\approx} f(X) = \sum_{k=0}^{K-1} \beta_k e_k(X). \tag{43}$$

We call $f(X)$ the linear predictor of $Y(X)$ or the *systematic component.*

With the fitting points $\left(x^i, y^i\right),\ i = 1, \ldots, N$, and uncorrelated errors ϵ^i (the *random components*) having the same variance $\sigma^2 > 0$ (*homoscedastic errors*), we obtain the classical linear regression model

$$y^i = \sum_{k=0}^{K-1} \beta_k e_k\left(x^i\right) + \epsilon^i, \tag{44}$$

where $e_0\left(x^i\right) = 1$ and β_0 is the intercept. Then the OLS estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ of the coefficients, minimizing the residual sum of squares, is given by

$$\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^K}{\arg\min} \left\{ \sum_{i=1}^{N} \left(y^i - \sum_{k=0}^{K-1} \beta_k e_k\left(x^i\right) \right)^2 \right\}. \tag{45}$$

The residuals corresponding to the OLS solution are $\widehat{\epsilon}^i = y^i - \sum_{k=0}^{K-1} \widehat{\beta}_{\mathrm{OLS},k} e_k\left(x^i\right)$. By substituting $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ for $\boldsymbol{\beta}$ in (43), we arrive at the proxy function $\widehat{f}(X)$ for the economic variable $Y(X)$ conditional on any outer scenario X, i.e.,

$$Y(X) \overset{K,N<\infty}{\approx} \widehat{f}(X) = \sum_{k=0}^{K-1} \widehat{\beta}_{\mathrm{OLS},k} e_k(X). \tag{46}$$

13.2.2 OLS Estimator & Closed-form Solution

If we use the notation $z_{ik} = e_k\left(x^i\right)$, we can replace the minimization problem (45) by the closed-form expression of the OLS estimator in which $Z = \left(z_{ik}\right)_{\substack{i=1,\ldots,N \\ k=0,\ldots,K-1}}$ denotes the design matrix and $\mathbf{y} = \left(y^1, \ldots, y^N\right)^{\mathrm{T}}$ the response vector, i.e.,

$$\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = \left(Z^{\mathrm{T}} Z\right)^{-1} Z^{\mathrm{T}} \mathbf{y}. \tag{47}$$

The system $\left(Z^{\mathrm{T}} Z\right) \widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = Z^{\mathrm{T}} \mathbf{y}$ equivalent to (47) is in practice often solved via a QR or singular value decomposition of Z to increase numerical stability. For a practical implementation, see, for example, function *lm(·)* in R package *stats* of R Core Team (2018).

The sample variance is obtained by $s_{\mathrm{OLS}}^2 = \frac{1}{N-K} \widehat{\boldsymbol{\epsilon}}^{\mathrm{T}} \widehat{\boldsymbol{\epsilon}}$ where $\widehat{\boldsymbol{\epsilon}} = \mathbf{y} - Z \widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ is the residual vector. With $\mathbf{z} = \left(e_0(X), \ldots, e_{K-1}(X)\right)^{\mathrm{T}}$, Equation (46) becomes in matrix notation

$$Y(X) \overset{K,N<\infty}{\approx} \widehat{f}(X) = \mathbf{z}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}. \tag{48}$$

13.2.3 Gauss-Markov Theorem, ML Estimation & AIC

We formulate the Gauss-Markov theorem in our setting conditional on the fitting scenarios and in line with Hayashi (2000) under the assumptions of strict exogeneity $E\left[\boldsymbol{\epsilon} \mid Z\right] = \mathbf{0}$ (A1), a spherical error variance $\mathrm{Var}\left[\boldsymbol{\epsilon} \mid Z\right] = \sigma^2 I_N$, where I_N is the N-dimensional identity matrix (A2), and no multicollinearity, that is, linearly independent basis functions (A3).

Gauss-Markov theorem. *The OLS estimator is the best linear unbiased estimator (BLUE) of the coefficients in the classical linear regression model (44) under Assumptions (A1)-(A3).*

Akaike information criterion (AIC) needs to be evaluated at the maximum likelihood (ML) estimators of the coefficients and variance of the errors. For this purpose, we have to make an assumption about the distribution of the economic variable, or equivalently the errors. In order to make AIC and OLS regression easily combinable we assume in addition to (A1), (A2) and (A3) that the errors are normally distributed conditional on the fitting scenarios (A4) because then Proposition 1.5 by Hayashi (2000) states the following.

Theorem 1. *The ML coefficient estimator coincides with the OLS coefficient estimator and the ML estimator of the error variance $\widehat{\sigma}^2$ can be expressed as $\frac{N-K}{N}$ times the OLS sample variance s^2_{OLS}, i.e., $\widehat{\sigma}^2 = \frac{1}{N}\widehat{\boldsymbol{\epsilon}}^{\mathrm{T}}\widehat{\boldsymbol{\epsilon}}$, under Assumptions (A1)-(A4).*

Furthermore, the OLS estimator is the efficient estimator under these assumptions according to Greene (2002).

According to Krah et al. (2018), AIC has the form of a suitably weighted sum of the calibration error and number of basis functions under Assumption (A4), i.e.,

$$\begin{aligned} \mathrm{AIC} &= -2l\left(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}, \widehat{\sigma}^2\right) + 2\left(K+1\right) \\ &= N\left(\log\left(2\pi\widehat{\sigma}^2\right) + 1\right) + 2\left(K+1\right). \end{aligned} \tag{49}$$

More generally, the calibration error corresponds to twice the negative of the log-likelihood $l\left(\cdot\right)$ of the model and the number of basis functions corresponds to the degrees of freedom of the model. The smaller the AIC score is, the better is the fitted model supposed to approximate the underlying data. AIC penalizes both a small log-likelihood and a high model complexity and helps thus to select a possibly simple model with a possibly high goodness of fit. However, since AIC is only a relative measure of the goodness of fit, the final proxy function has to pass an additional out-of-sample validation procedure in the LSMC algorithm.

13.2.4 Summary

The OLS regression algorithm in Section 13.2 requires the assumptions of strict exogeneity, homoscedastic errors and linearly independent basis functions for the coefficient estimator to be the best linear unbiased estimator by Gauss-Markov theorem. The OLS estimator minimizes the residual sum of squares by definition and has a closed-form expression. To evaluate AIC properly at the OLS estimator, the errors also have to be normally distributed according to Theorem 1.

13.3 Generalized Linear Models (GLMs)

13.3.1 Systematic & Random Components plus Link Function

Nelder & Wedderburn (1972) developed the class of generalized linear models (GLMs) as a generalization of the classical linear model in (44). A GLM consists of a random component, systematic component and link function and is derived by ML estimation. According to Chapter 2.2 of McCullagh & Nelder (1989), in a GLM the economic variable $Y(X)$ comes from a distribution of the exponential family conditional on outer scenario X, for instance, the normal, gamma, or inverse gaussian distribution.

With canonical parameter θ, the canonical form of this *random component* is given by density function

$$\pi(y \mid \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right), \tag{50}$$

where $a(\phi)$, $b(\theta)$ and $c(y,\phi)$ have particular functional forms. For example, a normally distributed economic variable with mean μ and variance σ^2 is given by $a(\phi) = \phi$, $b(\theta) = \frac{\theta^2}{2}$ and $c(y,\phi) = -\frac{1}{2}\left(\frac{y^2}{\sigma^2} + \log\left(2\pi\sigma^2\right)\right)$ with $\theta = \mu$ and $\phi = \sigma^2$ because then (50) becomes $\pi(y \mid \theta, \phi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$. The equivalence between the distribution assumption of the economic variable and raw errors $\epsilon = y - \mu$ persists in GLMs.

For a random variable Y from a distribution of the exponential family, the canonical parameter θ is related to the expected value while the dispersion parameter ϕ only affects the variance, i.e.,

$$E\left[Y\right] = \mu = b'\left(\theta\right), \quad \operatorname{Var}\left[Y\right] = b''\left(\theta\right) a(\phi) =: V\left[\mu\right] a(\phi), \tag{51}$$

whereby we refer to $V[\mu]$ as the variance function. As a simplification we consider only equal prior weights, that is, we set $a(\phi) = \phi$ to be constant over all observations.

The *systematic component* of a GLM coincides with the linear predictor $\eta = f(X)$ of the linear model in (43).

However, the first equality in (43) does not generally hold anymore. Instead a monotonic *link function* $g(\cdot)$ relates now the economic variable to the linear predictor, in literature usually formalized by $g(\mu) = \eta$, here by

$$g(\underbrace{Y(X)}_{=\,\mu}) \overset{K<\infty}{\approx} \underbrace{f(X)}_{=\,\eta} = \sum_{k=0}^{K-1} \beta_k z_k = \mathbf{z}^{\mathrm{T}} \boldsymbol{\beta} \tag{52}$$

with $\mathbf{z} = \left(e_0\left(X\right), \ldots, e_{K-1}\left(X\right)\right)^{\mathrm{T}}$. When the link function is the identity function as in the normal model the extension disappears, i.e., $\mu = \eta$.

Applying a link function is especially appealing when the range of the linear predictor may deviate substantially from that of the economic variable. For instance, an economic variable capturing service times that follow a gamma distribution can only be positive but the linear predictor may also take on negative values. With, for example, $g\left(\cdot\right) = \log\left(\cdot\right)$ such a potential inconsistency can be eliminated.

Another popular choice are the *canonical link functions* $\widetilde{g}\left(\cdot\right)$ which express the canonical parameter $\theta = \theta(X)$ with respect to the expected value $\mu = Y(X)$ if the variance is known, i.e., $\widetilde{g}\left(\mu\right) = \theta$, hence due to (52) also $\theta \overset{K<\infty}{\approx} f(X)$ with $\widetilde{g}\left(\cdot\right)$. For instance, the canonical link functions are $g\left(\mu\right) = \mathrm{id}\left(\mu\right)$ for the normal, $g\left(\mu\right) = \frac{1}{\mu}$ for the gamma, and $g\left(\mu\right) = \frac{1}{\mu^2}$ for the inverse gaussian distribution.

13.3.2 GLM Estimator & ML Estimation

The log-likelihood of a single observation is given by $l^i\left(\boldsymbol{\beta},\phi^i\right)=\log\pi(y^i\mid\theta^i,\phi^i)$ with the dependence $\theta^i=\theta^i\left(\mu^i\left(\eta^i\left(\boldsymbol{\beta},x^i\right)\right)\right)$ due to the equality $\mu=b'\left(\theta\right)$ and (52). With constant dispersion $a(\phi^i)=\phi^i=\phi,\ i=1,\ldots,N$, it follows $l\left(\boldsymbol{\beta},\phi\right)=\sum_{i=1}^N\log\pi\left(y^i\mid\theta^i,\phi\right)$ for the log-likelihood function.

The GLM estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{GLM}}$ of the coefficients is given as the ML maximizer, i.e.,

$$\widehat{\boldsymbol{\beta}}_{\mathrm{GLM}}=\underset{\boldsymbol{\beta}\in\mathbb{R}^K}{\arg\max}\left\{\sum_{i=1}^{N}\left(\frac{y^i\theta^i-b(\theta^i)}{\phi}+c(y^i,\phi)\right)\right\}. \tag{53}$$

While for the Poisson or binomial distribution the dispersion is simply 1, for the other distributions from the exponential family the dispersion ϕ is unknown. Assuming constant dispersion/equal prior weights (A5) lets the factors $a(\phi^i)$ disappear in the first-order ML condition. Therefore, we will omit the dispersion in the IRLS algorithm described below. Once $\widehat{\boldsymbol{\beta}}_{\mathrm{GLM}}$ is known, ϕ can be estimated with the aid of the Pearson residual chi-squared statistic. Using unequal prior weights might be beneficial, however it is not clear how they should be selected in the adaptive algorithm. Furthermore, they would make the estimation procedure more complicated.

13.3.3 GLM Estimation via IRLS Algorithm

Under Assumption (A5), there generally does not exist a closed-form solution for the GLM coefficient estimator (53). In Chapter 2.5, McCullagh & Nelder (1989) apply Fisher's scoring method, a standard approach in log-likelihood maximization, to obtain an approximation to the GLM estimator, i.e.,

$$\widehat{\boldsymbol{\beta}}^{(\boldsymbol{t}+\boldsymbol{1})}=\widehat{\boldsymbol{\beta}}^{(\boldsymbol{t})}+I^{-1}\frac{\partial l}{\partial\boldsymbol{\beta}}. \tag{54}$$

Here, $\widehat{\boldsymbol{\beta}}^{(\boldsymbol{t})}$ is the coefficient estimator in iteration t, $\frac{\partial l}{\partial\boldsymbol{\beta}}$ the score function, and $I=E\left[-\frac{\partial^2 l}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}}\right]$ the Fisher information matrix (equal to the negative of the expected value of the Hessian matrix) with the expectation being taken with respect to the random component. While $\frac{\partial l}{\partial\boldsymbol{\beta}}$ depends on the regressors and response values, I depends only on the regressors due to the expectation operator. Both have to be evaluated at $\widehat{\boldsymbol{\beta}}^{(\boldsymbol{t})}$.

McCullagh & Nelder (1989) justify how Fisher's scoring method can be cast in the form of the iteratively reweighted least squares (IRLS) algorithm. As an alternative, they suggest the Newton-Raphson method, which coincides with Fisher's scoring method if canonical link functions are used since the actual value of the Hessian matrix equals its expected value then.

The IRLS algorithm works for canonical link functions in our context as follows. Let the dependent variable in the iterative procedure be

$$\widehat{s}^i\left(\widehat{\boldsymbol{\beta}}^{(\boldsymbol{t})}\right)=\widehat{\eta}^i_{(t)}+\left(y^i-\widehat{\mu}^i_{(t)}\right)\frac{d\eta}{d\mu}\left(\widehat{\mu}^i_{(t)}\right), \tag{55}$$

where $\widehat{\eta}^i_{(t)}=\widehat{f}\left(x^i\right)$ is the estimate for the linear predictor evaluated at fitting scenario x^i, compare (52), where $\widehat{\mu}^i_{(t)}=g^{-1}\left(\widehat{\eta}^i_{(t)}\right)$ derived from $\widehat{\eta}^i_{(t)}$ is the estimate for the economic variable, and $\frac{d\eta}{d\mu}\left(\widehat{\mu}^i_{(t)}\right) = g'\left(\widehat{\mu}^i_{(t)}\right)$ is the first derivative of the link function with respect to the economic variable evaluated at $\widehat{\mu}^i_{(t)}$. Let $\widehat{\mathbf{s}}^{(t)} = \left(\widehat{s}^1\left(\widehat{\boldsymbol{\beta}}^{(t)}\right), \ldots, \widehat{s}^N\left(\widehat{\boldsymbol{\beta}}^{(t)}\right)\right)^{\mathrm{T}}$ denote the vector of the dependent variable over all fitting points.

Moreover, let the (quadratic) weight in the iterative procedure be given by

$$\widehat{w}^i\left(\widehat{\boldsymbol{\beta}}^{(t)}\right) = \left(\frac{d\eta}{d\mu}\left(\widehat{\mu}^i_{(t)}\right)\right)^{-2} V\left[\widehat{\mu}^i_{(t)}\right]^{-1}, \tag{56}$$

where $V\left[\widehat{\mu}^i_{(t)}\right]$ is the variance function from above evaluated at $\widehat{\mu}^i_{(t)}$. Then the (quadratic) weight matrix is defined by $W^{(t)} = \mathrm{diag}\left(w^1\left(\widehat{\boldsymbol{\beta}}^{(t)}\right), \ldots, w^N\left(\widehat{\boldsymbol{\beta}}^{(t)}\right)\right)$.

IRLS algorithm. *Perform the following iterative approximation procedure with, for example, an initialization of* $\widehat{\mu}^i_{(0)} = y^i + 0.1$ *and* $\widehat{\eta}^i_{(0)} = g\left(\widehat{\mu}^i_{(0)}\right)$ *as proposed by Dutang (2017) until convergence:*

$$\begin{aligned} \widehat{\boldsymbol{\beta}}^{(t+1)} &= \underset{\boldsymbol{\beta}\in\mathbb{R}^K}{\arg\min}\left\{\sum_{i=1}^{N} w^i\left(\widehat{\boldsymbol{\beta}}^{(t)}\right)^{-1}\left(\widehat{s}^i\left(\widehat{\boldsymbol{\beta}}^{(t)}\right) - \sum_{k=0}^{K-1}\beta_k z_{ik}\right)^2\right\} \\ &= \left(Z^{\mathrm{T}} W^{(t)} Z\right)^{-1} Z^{\mathrm{T}} W^{(t)} \widehat{\mathbf{s}}^{(t)}. \end{aligned} \tag{57}$$

After convergence, we set $\widehat{\boldsymbol{\beta}}_{\mathrm{GLM}} = \widehat{\boldsymbol{\beta}}^{(t+1)}$.

For example, Green (1984) proposes to solve system $\left(Z^{\mathrm{T}} W^{(t)} Z\right)\widehat{\boldsymbol{\beta}}^{(t+1)} = Z^{\mathrm{T}} W^{(t)} \widehat{\mathbf{s}}^{(t)}$ equivalent to (57) via a QR decomposition to increase numerical stability. For a practical implementation of GLMs using the IRLS algorithm, see, for example, function *glm(·)* in R package *stats* of R Core Team (2018).

By inserting (55), (56) and the GLM estimator into (57) and by using (52), we arrive at the property

$$\widehat{\boldsymbol{\beta}}_{\mathrm{GLM}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^K}{\arg\min}\left\{\sum_{i=1}^{N} V\left[\widehat{\mu}^i_{\mathrm{GLM}}\right]\left(y^i - \widehat{\mu}^i_{\mathrm{GLM}}\right)^2\right\}, \tag{58}$$

that is, the GLM estimator minimizes the squared sum of raw residuals scaled by the estimated individual variances of the economic variable. The higher the individual variance is, the more weight gets the point in the regression. The Pearson residuals are defined as the raw residuals divided by the estimated individual standard deviations, i.e.,

$$\widehat{\epsilon}^i = \frac{y^i - \widehat{\mu}^i_{\mathrm{GLM}}}{\sqrt{V\left[\widehat{\mu}^i_{\mathrm{GLM}}\right]}}. \tag{59}$$

For example, in the normal model from above with mean μ and variance σ^2, we have $b(\theta) = \frac{\theta^2}{2}$ and thus constant estimated individual variances across all observations $V[\mu] = b''(\theta) = 1$ so that no actual weighting takes place.

13.3.4 AIC & Dispersion Estimation

Since AIC depends on the ML estimators, it is combinable with GLMs in the adaptive algorithm. Here, it has the form

$$\mathrm{AIC} = -2l\left(\widehat{\boldsymbol{\beta}}_{\mathrm{GLM}}, \widehat{\phi}\right) + 2\,(K+p)\,, \tag{60}$$

where K is the number of coefficients and p indicates the number of the additional model parameters associated with the distribution of the random component. For instance, in the normal model, we have $p = 1$ due to the error variance/dispersion.

A typical estimate of the dispersion in GLMs is the Pearson residual chi-squared statistic divided by $N - K$ as described by Zuur et al. (2009) and implemented, for example, in function *glm(·)* belonging to R package *stats*, i.e.,

$$\widehat{\phi} = \frac{1}{N-K}\sum_{i=1}^{N}\left(\widehat{\epsilon}^{i}\right)^{2}, \tag{61}$$

with $\widehat{\epsilon}^{i}$ given by (59). Even though this is not the ML estimator, it is a good estimate because, if the model is specified correctly, the Pearson residual chi-squared statistic divided by the dispersion is asymptotically χ^2_{N-K} distributed and the expected value of a chi-squared distribution with $N - K$ degrees of freedom is $N - K$.

13.3.5 Summary

The GLM algorithm in Section 13.3 is a generalization of the OLS regression algorithm insofar as the errors are now permitted to come from an arbitrary distribution of the exponential family and the economic variable is related to the linear predictor by a monotonic link function. The GLM estimator maximizes the log-likelihood and can be derived by an IRLS algorithm. Without more ado, the GLM estimator can be fed into AIC.

13.4 Generalized Additive Models (GAMs)

13.4.1 Richly Parameterized GLM with Smooth Functions

The class of generalized additive models (GAMs) was introduced by Hastie & Tibshirani (1986) and Hastie & Tibshirani (1990) to unite the properties of GLMs and additive models. While GAMs inherit from GLMs the random component (50) and link function (52), they inherit from the additive models by Friedman & Stuetzle (1981) the linear predictor with the smooth functions.

By following Wood (2006), in the adaptive algorithm we apply GAMs of the form

$$g(\underbrace{Y(X)}_{=\,\mu}) \overset{K<\infty}{\approx} \underbrace{f(X)}_{=\,\eta} = \beta_0 + \sum_{k=1}^{K-1} h_k\,(z_k), \tag{62}$$

where $z_k = e_k\,(X)$, β_0 is the intercept and $h_k\,(\cdot)\,,\ k = 1, \ldots, K-1$, are the smooth functions to be estimated. In addition to the smooth functions, GAMs can also include simple linear terms of the basis functions as they appear in the linear predictor of GLMs.

Such an approach would be more parsimonious but also less straightforward. A smooth function $h_k(\cdot)$ can be written as a basis expansion

$$h_k(z_k) = \sum_{j=1}^{J} \beta_{kj} b_{kj}(z_k), \tag{63}$$

with coefficients β_{kj} and known basis functions $b_{kj}(z_k)$, $j = 1, \ldots, J$, which should not be confused with their arguments, namely the first-order basis functions $z_k = e_k(X)$, $k = 0, \ldots, K-1$. The slightly adapted Figure 18 from Wood (2006) depicts an exemplary approximation of y by a GAM with a basis expansion in one dimension z_k without an intercept. The solid colorful curves represent the pure basis functions $b_{kj}(z_k)$, $j = 1, \ldots, J$, the dashed colorful curves show them after scaling with the coefficients $\beta_{kj} b_{kj}(z_k)$, $j = 1, \ldots, J$, and the black curve is their sum (63).

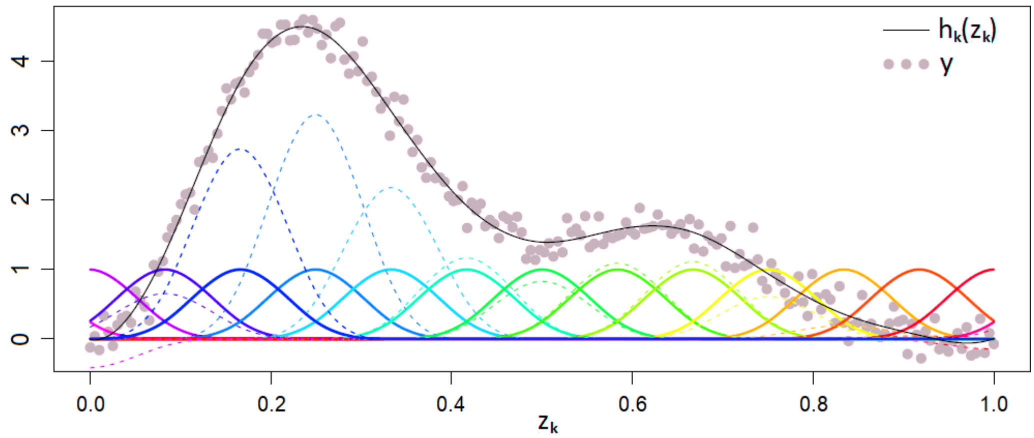

Figure 18: GAM with a basis expansion in one dimension.

Typical examples for basis functions are thin plate regression splines, duchon splines, penalized cubic regression splines or Eilers and Marx style P-splines. See, for example, function *gam(·)* in R package *mgcv* of Wood (2018) for a practical implementation of GAMs admitting these types of basis functions and using the PIRLS algorithm, which we present below.

In vector notation, we can write $\boldsymbol{\beta} = \left(\beta_0, \boldsymbol{\beta}_1^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_{K-1}^{\mathrm{T}}\right)^{\mathrm{T}}$ with $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kJ})^{\mathrm{T}}$ and $\mathbf{a} = \left(1, \mathbf{b}_1(z_1)^{\mathrm{T}}, \ldots, \mathbf{b}_{K-1}(z_{K-1})^{\mathrm{T}}\right)^{\mathrm{T}}$ with $\mathbf{b}_k(z_k) = (b_{k1}(z_k), \ldots, b_{kJ}(z_k))^{\mathrm{T}}$, hence (62) becomes

$$g(\underbrace{Y(X)}_{=\mu}) \overset{K<\infty}{\approx} \underbrace{f(X)}_{=\eta} = \mathbf{a}^{\mathrm{T}} \boldsymbol{\beta}. \tag{64}$$

This parameterization is a richer version of (52) so that a GAM having a random component from the exponential family (50) can be viewed as a richly parameterized GLM. In order to make the smooth functions $h_k(\cdot)$, $k = 1, \ldots, K-1$, identifiable, identifiability constraints $\sum_{i=1}^{N} h_k(z_{ik}) = 0$ with $z_{ik} = e_k(x^i)$ can be imposed. According to Wood (2006) this can be achieved by modification of the basis functions $b_{kj}(\cdot)$ with one of them being lost.

13.4.2 GAM Estimator & Penalization

Let the deviance corresponding to observation y^i be $D^i(\boldsymbol{\beta}) = 2\left(l^i_{\text{sat}} - l^i(\boldsymbol{\beta}, \phi)\right)\phi$ where $D^i(\boldsymbol{\beta})$ is independent of dispersion ϕ, where $l^i_{\text{sat}} = \max_{\boldsymbol{\beta}^i} l^i\left(\boldsymbol{\beta}^i, \phi\right)$ is the saturated log-likelihood and $l^i(\boldsymbol{\beta}, \phi)$ the log-likelihood. Then the model deviance can be written as $D(\boldsymbol{\beta}) = \sum_{i=1}^N D^i(\boldsymbol{\beta})$. It is a generalization of the residual sum of squares for ML estimation. For instance, in the normal model the unit deviance is $\left(y^i - \mu^i\right)^2$.

For given smoothing parameters $\lambda_k > 0,\ k = 1, \ldots, K-1$, the GAM estimator $\widehat{\boldsymbol{\beta}}_{\text{GAM}}$ of the coefficients is defined as the minimizer of the penalized deviance

$$\widehat{\boldsymbol{\beta}}_{\text{GAM}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^{(K-1)J+1}}{\arg\min} \left\{ D(\boldsymbol{\beta}) + \sum_{k=1}^{K-1} \lambda_k \int h''_k(z_k)^2\, \mathrm{d}z_k \right\}, \text{ where} \tag{65}$$

$$\int h''_k(z_k)^2\, \mathrm{d}z_k = \boldsymbol{\beta}_k^{\mathrm{T}} \left(\int \mathbf{b}''_k(z_k)\, \mathbf{b}''_k(z_k)^{\mathrm{T}}\, \mathrm{d}z_k \right) \boldsymbol{\beta}_k = \boldsymbol{\beta}_k^{\mathrm{T}} \mathcal{S}_k \boldsymbol{\beta}_k$$

are the smoothing penalties. The smoothing parameters λ_k control the trade-off between a too wiggly model (overfitting) and a too smooth model (underfitting). The larger the λ_k values are, the more pronounced is the wiggliness of the basis functions reflected by their second derivatives in the minimization problem (65), and the higher is thus the penalty associated with the coefficients and the smoother is the estimated model.

Similarly to how we have defined the GAM estimator as the minimizer of the penalized deviance, we could have defined the GLM estimator (53) as the minimizer of the unpenalized deviance.

13.4.3 GAM Estimation via PIRLS Algorithm

Buja et al. (1989) proposed to estimate GAMs by a backfitting procedure which can be shown to be the Gauss-Seidel iterative method for solving a set of normal equations associated with the additive model. Their backfitting procedure works for any scatterplot smoother so that the random component does no longer have to come from the exponential family, in fact, non-parametric models such as running-mean, running-line or kernel smoothers are possible as well. However, their suggestions to select the degree of smoothness through, for instance, graphical analyses or cross-validation are for practitioners still difficult to implement.

Therefore, GAMs have recently been increasingly defined in the form of (62) with basis expansions (63) of which the degree of smoothness is controlled by the smoothing penalties (65). A major advantage of this definition is its compatibility with information criteria and other model selection criteria such as generalized cross-validation. Besides, the resulting penalty matrix favors numerical stability in the PIRLS algorithm.

Since the saturated log-likelihood is a constant for a fixed distribution and set of fitting points, we can turn the minimization problem (65) into the maximization task of the penalized log-likelihood, i.e.,

$$\widehat{\boldsymbol{\beta}}_{\text{GAM}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^{(K-1)J+1}}{\arg\max} \left\{ l(\boldsymbol{\beta}, \phi) - \frac{1}{2} \sum_{k=1}^{K-1} \lambda_k \boldsymbol{\beta}_k^{\mathrm{T}} \mathcal{S}_k \boldsymbol{\beta}_k \right\}. \tag{66}$$

Wood (2000) points out that Fisher's scoring method can be cast in a penalized version of the iteratively reweighted least squares (PIRLS) algorithm when being used to approximate the GAM coefficient estimator (66). This derivation is very similar to the one of the

IRLS algorithm in the GLM context with the constant dispersion ϕ disappearing in the first-order condition. We formulate the PIRLS algorithm based on Marx & Eilers (1998) who indicate the iterative solution explicitly.

Let $\widehat{\boldsymbol{\beta}}^{(t)}$ now be the GAM coefficient approximation in iteration t. Then the vector of the dependent variable $\widehat{\mathbf{s}}^{(t)} = \left(\widehat{s}^1\left(\widehat{\boldsymbol{\beta}}^{(t)}\right), \ldots, \widehat{s}^N\left(\widehat{\boldsymbol{\beta}}^{(t)}\right)\right)^{\mathrm{T}}$ and the weight matrix given by $W^{(t)} = \operatorname{diag}\left(w^1\left(\widehat{\boldsymbol{\beta}}^{(t)}\right), \ldots, w^N\left(\widehat{\boldsymbol{\beta}}^{(t)}\right)\right)$ have the same form as in the IRLS algorithm, see (55) and (56). Additionally, let $S = \operatorname{blockdiag}\left(0, \lambda_1\mathcal{S}_1, \ldots, \lambda_{K-1}\mathcal{S}_{K-1}\right)$ with $S_{11} = 0$ belonging to the intercept be the penalty matrix.

PIRLS algorithm. *Perform the following iterative approximation procedure with, for example, an initialization of* $\widehat{\mu}^i_{(0)} = y^i + 0.1$ *and* $\widehat{\eta}^i_{(0)} = g\left(\widehat{\mu}^i_{(0)}\right)$ *in analogy to the IRLS algorithm until convergence:*

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^{(t+1)} &= \underset{\boldsymbol{\beta}\in\mathbb{R}^{(K-1)J+1}}{\arg\min} \left\{\sum_{i=1}^{N} w^i\left(\widehat{\boldsymbol{\beta}}^{(t)}\right)^{-1}\left(\widehat{s}^i\left(\widehat{\boldsymbol{\beta}}^{(t)}\right) - \beta_0 - \sum_{k=1}^{K-1}\sum_{j=1}^{J}\beta_{kj}b_{kj}\left(z_{ik}\right)\right)^2 + \sum_{k=1}^{K-1}\lambda_k\boldsymbol{\beta}_k^{\mathrm{T}}\mathcal{S}_k\boldsymbol{\beta}_k\right\} \\ &= \left(Z^{\mathrm{T}}W^{(t)}Z + S\right)^{-1} Z^{\mathrm{T}}W^{(t)}\widehat{\mathbf{s}}^{(t)}. \end{aligned} \tag{67}$$

After convergence, we set $\widehat{\boldsymbol{\beta}}_{\mathrm{GAM}} = \widehat{\boldsymbol{\beta}}^{(t+1)}$.

13.4.4 Smoothing Parameter Selection, AIC & GCV

The smoothing parameters λ_k can be selected such that they minimize a suitable model selection criterion, for the sake of consistency, preferably the one used in the adaptive algorithm for basis function selection. The GAM estimator (66) does not exactly maximize the log-likelihood, therefore AIC has another form for GAMs than for GLMs. The degrees of freedom need to be adjusted with respect to the smoothing effects of the penalties on the coefficients. The reasoning behind this adjustment is that high smoothing parameters restrict the coefficients more than low smoothing parameters and so that they need to be associated with less effective degrees of freedom.

Hastie & Tibshirani (1990) propose a widely used version of AIC for GAMs, which uses effective degrees of freedom df in place of the number of coefficients $(K-1)J+1$. This is

$$\mathrm{AIC} = -2l\left(\widehat{\boldsymbol{\beta}}_{\mathrm{GAM}}, \widehat{\phi}\right) + 2\left(\mathrm{df} + p\right), \tag{68}$$

where

$$\mathrm{df} = \operatorname{tr}\left(\left(I+S\right)^{-1} I\right). \tag{69}$$

The expression $I = Z^{\mathrm{T}}WZ$ for the Fisher information matrix with the weight matrix W evaluated at the GAM estimator is obtained as a by-product when casting Fisher's scoring method in the form of the PIRLS algorithm.

Without the penalty matrix S, we have $\mathrm{df} = \operatorname{tr}\left(I^{-1}I\right) = (K-1)J+1$. If we follow Wood (2006) by denoting the unpenalized GAM estimator by $\widehat{\boldsymbol{\beta}}^0_{\mathrm{GAM}}$ and the so-called shrinkage matrix by $F = \left(Z^{\mathrm{T}}WZ + S\right)^{-1} Z^{\mathrm{T}}WZ$ with $\widehat{\boldsymbol{\beta}}_{\mathrm{GAM}} = F\widehat{\boldsymbol{\beta}}^0_{\mathrm{GAM}}$, we arrive at the equality $\mathrm{df} = \operatorname{tr}\left(F\right)$ revealing the shrinkage effects on the effective degrees of freedom. After convergence of the PIRLS algorithm, the dependent variable is constant, i.e., $\widehat{\boldsymbol{s}} = \widehat{\boldsymbol{s}}^{(t)}$, and the hat matrix H satisfies $\left(\widehat{\eta}^1, \ldots, \widehat{\eta}^N\right)^{\mathrm{T}} = Z\widehat{\boldsymbol{\beta}}_{\mathrm{GAM}} = H\widehat{\boldsymbol{s}}$ so that $H = Z\left(Z^{\mathrm{T}}WZ + S\right)^{-1} Z^{\mathrm{T}}W$.

Due to the cyclic property of the trace, the effective degrees of freedom can also be written as $\mathrm{df} = \mathrm{tr}\,(H)$.

For GAMs, an estimate of the dispersion $\widehat{\phi}$ is obtained similarly to GLMs by (61). The parameter p is defined as in (60). For a refinement of (68) accounting for the uncertainty of the smoothing parameters and tending to select models less prone to overfitting, see Wood et al. (2016).

Another popular and effective smoothing parameter selection criterion invented by Craven & Wahba (1979) is generalized cross-validation (GCV), i.e.,

$$\mathrm{GCV} = \frac{N D\left(\widehat{\boldsymbol{\beta}}_{\mathrm{GAM}}\right)}{\left(N - \mathrm{df}\right)^2}, \tag{70}$$

with the model deviance $D\left(\widehat{\boldsymbol{\beta}}_{\mathrm{GAM}}\right)$ evaluated at the GAM estimator and the effective degrees of freedom defined just like for AIC.

13.4.5 Adaptive Forward Stagewise Selection & Performance

In situations where the economic variable depends on many risk factors and where large sample sizes are required, the adaptive forward stepwise algorithm depicted in Figure 13 can become computationally infeasible with GAMs as opposed to, for instance, GLMs. In iteration k, a GAM has $(K-1)J+1$ coefficients which need to be estimated while a GLM has only K coefficients. This difference in the estimation effort is increased further due to the iterative nature of the IRLS and PIRLS algorithms. Moreover, GAMs involve the task of optimal smoothing parameter selection, which scales the estimation effort for GAMs up once more tremendously.

Wood (2000) has found a way to make smoothing parameter selection more efficient. Furthermore, Wood et al. (2015) and Wood et al. (2017) have developed practical GAM fitting methods for large data sets. These methods involve, for example, iterative update schemes, requiring only subblocks of the design matrix to be recomputed, and parallelization. The suitable application of these methods in the adaptive algorithm is beyond the scope of this analysis though since our focus does not lie on computational performance.

Besides parallelizing the candidate loop on the lower left side of Figure 13, we achieve the necessary performance gains in GAMs by replacing the stepwise algorithm by a stagewise algorithm. This means that in each iteration, a predefined number L or proportion of candidate terms is selected simultaneously until a termination criterion is fulfilled. Thereby we select in one stage those basis functions which reduce the model selection criterion of our choice most when added separately to the current proxy function term structure. When there are not at least as many basis functions as targeted, the algorithm is terminated after the ones leading to a reduction of the model selection criterion have been selected.

13.4.6 Summary

The GAM algorithm in Section 13.4 acts as a generalization of the GLM algorithm and brings in the additive models with the smooth functions as the new component. The GAM estimator maximizes the penalized log-likelihood and can be derived by a PIRLS algorithm. The penalization takes place with respect to smoothing parameters controlling the trade-off between a too wiggly and too smooth model. To evaluate AIC at the GAM estimator, the degrees of freedom are generalized such that they account for the smoothing.

As an alternative to AIC, generalized cross-validation GCV is introduced. The smoothing parameters are selected such that they minimize the model selection criterion. For reasons of computational efficiency, adaptive forward stagewise selection is suggested.

13.5 Feasible Generalized Least-Squares (FGLS) Regression

13.5.1 Generalized Linear Regression Model

The linear predictor of the generalized linear regression model has the same form (44) as in the OLS case. But while the errors were assumed to be uncorrelated and to have the same unknown variance $\sigma^2 > 0$ in the classical linear regression model, now they are assumed to have the covariance matrix $\Sigma = \sigma^2 \Omega$ where Ω is positive definite and known and $\sigma^2 > 0$ is unknown.

We transform the generalized linear regression model according to Hayashi (2000) to obtain a model (*) which satisfies Assumptions (A1), (A2) and (A3) of the classical linear regression model. As Ω is by construction symmetric and positive definite, there exists an invertible matrix H such that $\Omega^{-1} = H^{\mathrm{T}} H$. The matrix H is not unique but this is not important since any choice of H such as, for example, the Cholesky matrix works. The generalized response vector $\mathbf{y}^*$, design matrix Z^* and error vector $\boldsymbol{\epsilon}^*$ are then given by

$$\mathbf{y}^* = H\mathbf{y}, \quad Z^* = HZ, \quad \boldsymbol{\epsilon}^* = \mathbf{y}^* - Z^*\boldsymbol{\beta} = H\left(\mathbf{y} - Z\boldsymbol{\beta}\right) = H\boldsymbol{\epsilon}. \tag{71}$$

Strict exogeneity (A1) is satisfied by the transformed regression model (*) as $E\left[\boldsymbol{\epsilon}^* \mid Z^*\right] = HE\left[\boldsymbol{\epsilon} \mid Z\right] = \mathbf{0}$, the error variance is spherical (A2) because of $\Sigma^* = \mathrm{Var}\left[\boldsymbol{\epsilon}^* \mid Z^*\right] = H\mathrm{Var}\left[\boldsymbol{\epsilon} \mid Z\right] H^{\mathrm{T}} = H\left[\sigma^2\Omega\right] H^{\mathrm{T}} = H\left[\sigma^2\left(H^{\mathrm{T}}H\right)^{-1}\right] H^{\mathrm{T}} = \sigma^2 I_N$ with the N-dimensional identity matrix I_N and the no-multicollinearity assumption (A3) holds as Ω is positive definite.

13.5.2 GLS Estimator & Closed-form Solution

In analogy to the OLS estimator, the generalized least-squares (GLS) estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ of the coefficients is given as the minimizer of the generalized residual sum of squares, i.e.,

$$\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^K}{\arg\min} \left\{ \sum_{i=1}^{N} \left(\epsilon^{*,i}\right)^2 \right\}. \tag{72}$$

The closed-form expression of the GLS estimator is

$$\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = \left(Z^{*,\mathrm{T}} Z^*\right)^{-1} Z^{*,\mathrm{T}} \mathbf{y}^* = \left(Z^{\mathrm{T}}\Omega^{-1}Z\right)^{-1} Z^{\mathrm{T}}\Omega^{-1}\mathbf{y}, \tag{73}$$

and the proxy function becomes

$$\widehat{f}\left(X\right) = \mathbf{z}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}, \tag{74}$$

where $\mathbf{z} = \left(e_0\left(X\right), \ldots, e_{K-1}\left(X\right)\right)^{\mathrm{T}}$. The scalar σ^2 can be estimated in analogy to OLS regression by $s^2_{\mathrm{GLS}} = \frac{1}{N-K}\widehat{\boldsymbol{\epsilon}}^{*,\mathrm{T}}\widehat{\boldsymbol{\epsilon}}^*$ where $\widehat{\boldsymbol{\epsilon}}^* = \mathbf{y}^* - Z^*\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ is the residual vector.

13.5.3 Gauss-Markov-Aitken Theorem & ML Estimation

We formulate the Gauss-Markov-Aitken theorem conditional on the fitting scenarios in line with Huang (1970) and Hayashi (2000) under the assumptions of strict exogeneity (A1), no multicollinearity (A3) and a covariance matrix $\Sigma = \sigma^2\Omega$ of which Ω is positive definite and known (A6).

Gauss-Markov-Aitken theorem. *The GLS estimator is the BLUE of the coefficients in the generalized regression model (44) under Assumptions (A1), (A3) and (A6).*

In order to make AIC and GLS regression combinable, we assume additionally to (A1), (A3) and (A6) that the economic variable, or equivalently the errors, are jointly normally distributed conditional on the fitting scenarios (A7). The transformation (*) transfers to the ML function of the generalized regression model so that we can state the following theorem in analogy to Theorem 1, see e.g. Hartmann (2015).

Theorem 2. *The ML coefficient estimator coincides with the GLS coefficient estimator and the ML estimator of the scalar $\widehat{\sigma}^2$ can be expressed as $\frac{N}{N-K}$ times s^2_{GLS}, i.e., $\widehat{\sigma}^2 = \frac{1}{N}\widehat{\boldsymbol{\epsilon}}^{*,\mathrm{T}}\widehat{\boldsymbol{\epsilon}}^{*}$, under Assumptions (A1), (A3), (A6) and (A7).*

13.5.4 FGLS Estimator & Unknown Covariance Matrix

In the LSMC framework, Ω is unknown. If a consistent estimator $\widehat{\Omega}$ exists, we can apply feasible generalized least-squares (FGLS) regression, of which the estimator

$$\widehat{\boldsymbol{\beta}}_{\mathrm{FGLS}} = \left(Z^{\mathrm{T}}\widehat{\Omega}^{-1}Z\right)^{-1} Z^{\mathrm{T}}\widehat{\Omega}^{-1}\mathbf{y} \tag{75}$$

has asymptotically the same properties as the GLS estimator (73).

Greene (2002) remarks that the asymptotic efficiency of the FGLS estimator does not carry over to finite samples. In small sample studies with no severe deviations from the homoscedasticity assumption, the OLS estimator has been shown to be sometimes more efficient than the FGLS estimator. However, in cases where the deviations from this assumption were more severe, the FGLS estimator has been shown to outperform the OLS estimator.

With $\mathbf{z} = \left(e_0\left(X\right), \ldots, e_{K-1}\left(X\right)\right)^{\mathrm{T}}$ the FGLS proxy function is given in analogy to (48) and (74) as

$$\widehat{f}\left(X\right) = \mathbf{z}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{FGLS}}. \tag{76}$$

13.5.5 FGLS Estimation via ML Algorithm

For the estimation of Ω we will in the following set $\sigma^2 = 1$ which can be done without loss of generality and then consider $\Sigma = \Omega$. Hereby, any specification of $\sigma^2 > 0$ would be possible as the GLS and FGLS coefficient estimators are invariant to scalings of Ω and $\widehat{\Omega}$, respectively. In addition to (A1), (A3) and (A7), we assume that the elements of the covariance matrix Σ are twice differentiable functions of parameters $\boldsymbol{\alpha} = \left(\alpha_0, \ldots, \alpha_{M-1}\right)^{\mathrm{T}}$ with $K + M \leq N$ so that we can write $\Sigma = \Sigma\left(\boldsymbol{\alpha}\right)$ (A8). The following result is the basis of the iterative ML algorithm for deriving the regression coefficients and covariance matrix.

Theorem 3. *The generalized regression model (44) under Assumptions (A1), (A3), (A7) and (A8) has the following first-order ML conditions:*

$$\widehat{\boldsymbol{\beta}}_{\mathrm{ML}} = \left(Z^{\mathrm{T}}\widehat{\Sigma}^{-1}Z\right)^{-1} Z^{\mathrm{T}}\widehat{\Sigma}^{-1}\mathbf{y}, \tag{77}$$

$$\frac{\partial l}{\partial \alpha_m} = \frac{1}{2}\mathrm{tr}\left(\frac{\partial \Sigma^{-1}}{\partial \alpha_m}\Sigma\right)_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}_{\mathrm{ML}}} - \frac{1}{2}\widehat{\boldsymbol{\epsilon}}^{\mathrm{T}}\left(\frac{\partial \Sigma^{-1}}{\partial \alpha_m}\right)_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}_{\mathrm{ML}}}\widehat{\boldsymbol{\epsilon}} = 0, \tag{78}$$

where $m = 0, \ldots, M-1$, $\widehat{\Sigma} = \Sigma\left(\widehat{\boldsymbol{\alpha}}_{\mathrm{ML}}\right)$ *and* $\widehat{\boldsymbol{\epsilon}} = \mathbf{y} - Z\widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$.

The system in (77) and (78) is then solved iteratively, for example, according to Magnus (1978). We start the procedure with $\boldsymbol{\beta}^{(0)}$ and use PORT optimization routines as described in Gay (1990) and implemented in function *nlminb(·)* belonging to R package *stats* of R Core Team (2018). In these routines, $\widehat{\boldsymbol{\alpha}}^{(t+1)}$ can also be initialized, for instance, by random numbers from the standard normal distribution.

ML algorithm. *Perform the following iterative approximation procedure with, for example, an initialization of* $\widehat{\boldsymbol{\beta}}^{(0)} = \widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ *until convergence:*

1. *Calculate the residual vector* $\widehat{\boldsymbol{\epsilon}}^{(t+1)} = \mathbf{y} - Z\widehat{\boldsymbol{\beta}}^{(t)}$.

2. *Substitute* $\widehat{\boldsymbol{\epsilon}}^{(t+1)}$ *into the* M *equations in* M *unknowns* α_m *given by (78) and solve them. If an explicit solution exists, set* $\widehat{\boldsymbol{\alpha}}^{(t+1)} = \boldsymbol{\alpha}\left(\widehat{\boldsymbol{\epsilon}}^{(t+1)}\right)$. *Otherwise, select the maximum likelihood solution* $\widehat{\boldsymbol{\alpha}}^{(t+1)}$ *iteratively, for example, by using PORT optimization routines.*

3. *Calculate*

$$\widehat{\Sigma}^{(t+1)} = \Sigma\left(\widehat{\boldsymbol{\alpha}}^{(t+1)}\right),$$

$$\widehat{\boldsymbol{\beta}}^{(t+1)} = \left(Z^{\mathrm{T}}\left(\widehat{\Sigma}^{(t+1)}\right)^{-1}Z\right)^{-1} Z^{\mathrm{T}}\left(\widehat{\Sigma}^{(t+1)}\right)^{-1}\mathbf{y}. \tag{79}$$

   Continue with the next iteration.

After convergence, we set $\widehat{\boldsymbol{\beta}}_{\mathrm{ML}} = \widehat{\boldsymbol{\beta}}^{(t+1)}$ *and* $\widehat{\boldsymbol{\alpha}}_{\mathrm{ML}} = \widehat{\boldsymbol{\alpha}}^{(t+1)}$.

Some further regularity conditions guarantee the consistency of the ML estimators and therefore lead to the following result.

Theorem 4. *The FGLS coefficient estimator can be derived as the ML coefficient estimator by the ML algorithm under Assumptions (A1), (A3), (A7) and (A8) and some further regularity conditions stated in Theorem 5 by Magnus (1978).*

13.5.6 Heteroscedasticity & Breusch-Pagan Test

Besides Assumption (A8) about the structure of the covariance matrix, we assume that the errors are uncorrelated with possibly different variances (*heteroscedastic errors*), i.e., $\Sigma = \mathrm{diag}\left(\sigma_1^2, \ldots, \sigma_N^2\right)$. We model each variance σ_i^2, $i = 1, \ldots, N$, by a twice differentiable function in dependence of parameters $\boldsymbol{\alpha} = \left(\alpha_0, \ldots, \alpha_{M-1}\right)^{\mathrm{T}}$ and a suitable set of linearly independent basis functions $e_m(X) \in L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{P}'\right)$, $m = 0, 1, \ldots, M-1$, with $\mathbf{v}^i = \left(e_0\left(x^i\right), \ldots, e_{M-1}\left(x^i\right)\right)^{\mathrm{T}}$, i.e.,

$$\sigma_i^2 = \sigma^2 V\left[\boldsymbol{\alpha}, \mathbf{v}^i\right], \tag{80}$$

where $V\left[\boldsymbol{\alpha}, \mathbf{v}^i\right]$ is referred to as the variance function in analogy to $V[\mu]$ for GLMs and GAMs. Without loss of generality, we set again $\sigma^2 = 1$. Like in minimization problem (58) of GLMs, the individual variances determine the regression weights in generalized least-squares problem (72).

Hartmann (2015) has already applied FGLS regression with different variance models in the LSMC framework. In her numerical examples, variance models with multiplicative heteroscedasticity led to the best performance of the proxy function in the validation. Therefore, we restrict our analysis on these kinds of structures, compare e.g. Harvey (1976), i.e.,

$$V\left[\boldsymbol{\alpha}, \mathbf{v}^i\right] = \exp\left(\mathbf{v}^{i,\mathrm{T}}\boldsymbol{\alpha}\right). \tag{81}$$

The conceivable alternatives applied by Hartmann (2015) are variance models with additive heteroscedasticity, see e.g. Glejser (1969), heteroscedasticity with respect to powers, see e.g. Carroll & Ruppert (1988), or heteroscedasticity with respect to the conditional expectation.

We should only apply FGLS regression as a substitute of OLS regression if heteroscedasticity prevails. If the variance function has the structure

$$V\left[\boldsymbol{\alpha}, \mathbf{v}^i\right] = h\left(\mathbf{v}^{i,\mathrm{T}}\boldsymbol{\alpha}\right), \tag{82}$$

where the function $h(\cdot)$ is twice differentiable and the first element of $\mathbf{v}^i$ is $\mathrm{v}_0^i = 1$, the Breusch-Pagan test by Breusch & Pagan (1979) can be used to diagnose heteroscedasticity under the assumption of normally distributed errors. We use it in the numerical computations to check if heteroscedasticity still prevails during the iterative procedure.

13.5.7 Variance Model Selection & AIC

Like the proxy function, the variance function (81) has to be calibrated to apply FGLS regression, which means that the variance function has to be composed of suitable basis functions. Again, such a composition can be found with the aid of a model selection criterion. We stick to AIC but have to take care of the fact that the covariance matrix now contains M unknown parameters instead of only one as in the OLS case (the same variance for all observations). Under Assumption (A7), AIC is given as

$$\begin{aligned} \mathrm{AIC} &= -2l\left(\widehat{\boldsymbol{\beta}}_{\mathrm{FGLS}}, \widehat{\Sigma}\right) + 2\left(K+M\right) \qquad (83)\\ &= N\log\left(2\pi\right) + \log\left(\det\widehat{\Sigma}\right) + \left(\mathbf{y} - Z\widehat{\boldsymbol{\beta}}_{\mathrm{FGLS}}\right)^{\mathrm{T}}\widehat{\Sigma}^{-1}\left(\mathbf{y} - Z\widehat{\boldsymbol{\beta}}_{\mathrm{FGLS}}\right) + 2\left(K+M\right). \end{aligned}$$

When using a variance model with multiplicative heteroscedasticity, AIC becomes

$$\mathrm{AIC} = N\log\left(2\pi\right) + \left(\sum_{i=1}^{N}\mathbf{v}^{i,\mathrm{T}}\right)\widehat{\boldsymbol{\alpha}} + \sum_{i=1}^{N}\exp\left(-\mathbf{v}^{i,\mathrm{T}}\widehat{\boldsymbol{\alpha}}\right)\left(\widehat{\epsilon}^i\right)^2 + 2\left(K+M\right). \tag{84}$$

As an alternative or complement, the basis functions of the variance model can be selected with respect to their correlations with the absolute values of the final OLS residuals or based on graphical residual analysis.

A difficulty of variance model selection poses its potential interdependency with proxy function selection because the basis functions minimizing the model selection criterion when being added to the proxy function might depend on the selected basis functions of the variance model and vice versa. There are multiple ways to tackle the interdependency difficulty, compare Hartmann (2015), of which we implement two variants with rather short run times and promising out-of-sample validation performances.

Our type I variant starts with the derivation of the proxy function by the standard adaptive OLS regression approach and then selects the variance model adaptively from the set of proxy basis functions of which the exponents sum up to at most two. The type II variant builds on the type I algorithm by taking the resulting variance model as given in its adaptive proxy basis function selection procedure with FGLS regression in each iteration.

13.5.8 Summary

The FGLS regression algorithm in Section 13.5 is another generalization of the OLS regression algorithm insofar as the errors are here allowed to have any positive definite covariance matrix. For the GLS estimator to be the best linear unbiased estimator by Gauss-Markov-Aitken theorem, the assumptions of strict exogeneity, linearly independent basis functions and a known covariance matrix are required. The GLS estimator minimizes the generalized residual sum of squares. When the covariance matrix is unknown but can be estimated consistently, the FGLS estimator serves as a substitute for the GLS estimator that has asymptotically the same properties. If furthermore the errors are jointly normally distributed, the FGLS estimator can be derived by a maximum likelihood algorithm and fed into AIC according to Theorem 3. Suitable implementations are multiplicative heteroscedasticity, adaptive variance model selection procedures and Breusch-Pagan test for heterogeneity diagnosis.

13.6 Multivariate Adaptive Regression Splines (MARS)

13.6.1 OLS Regression/GLM with Hinge Functions

The multivariate adaptive regression splines (MARS) were introduced by Friedman (1991). The classical MARS model is a form of the classical linear regression model (44), where the basis functions $e_k\left(x^i\right)$ are so-called hinge functions. Therefore, the theory of OLS regression applies in this context. Since GLMs (52) are generalizations of the classical linear regression model, they can also be applied in conjunction with MARS models. In this case we speak of generalized MARS models.

We describe the standard MARS algorithm in the LSMC routine along the lines of Chapter 9.4 of Hastie et al. (2017). The building blocks of MARS proxy functions are reflected pairs of piecewise linear functions with knots t as depicted in Figure 19, i.e.,

$$\begin{aligned}(X_l - t)_+ &= \max\left(X_l - t, 0\right),\\ (t - X_l)_+ &= \max\left(t - X_l, 0\right),\end{aligned} \tag{85}$$

where the $X_l,\ l = 1, \dots, d$, represent the risk factors which together form the outer scenario $X = (X_1, \dots, X_d)^{\mathrm{T}}$.

For each risk factor, reflected pairs with knots at each fitting scenario stress $x_l^i,\ i = 1, \dots, N$, are defined. All pairs are united in the following collection serving as the initial candidate term set of the MARS algorithm, i.e.,

$$C_1 = \left\{(X_l - t)_+, (t - X_l)_+\right\}_{t \in \left\{x_l^1, x_l^2, \dots, x_l^N\right\} \ \mid\ l = 1, \dots, d}. \tag{86}$$

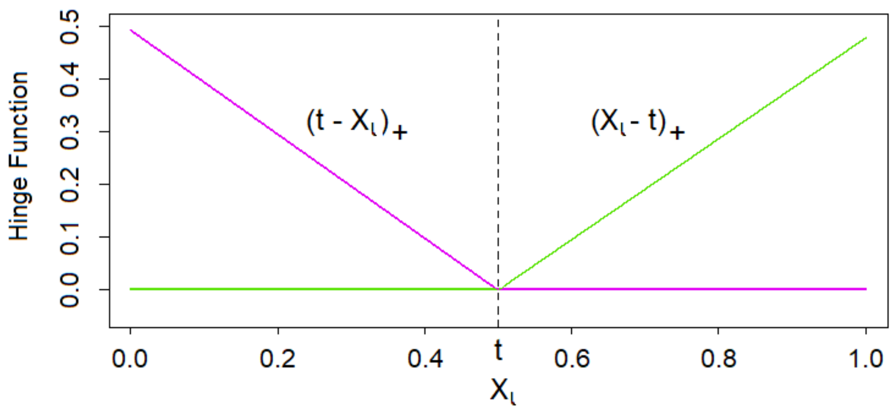

Figure 19: Reflected pair of piecewise linear functions with a knot at t.

We call the elements of such a collection hinge functions and write them as functions $h(X)$ over the entire input space $\mathbb{R}^d$. The initial set C_1 contains in total $2dN$ basis functions.

The adaptive basis function selection algorithm now consists of two parts, the forward and the backward part. An especially fast MARS algorithm was developed by Friedman (1993) and is implemented, for example, in function *earth(·)* of R package *earth* provided by Milborrow (2018).

13.6.2 Adaptive Forward Stepwise Selection & Forward Pass

The forward pass of the MARS algorithm can be viewed as a variation of the adaptive forward stepwise algorithm depicted in Figure 13. The start proxy function consists only of the intercept, that is, $h_0(X) = 1$. In the classical MARS model, the regression method of choice is the standard OLS regression approach with the estimator (45), where in each iteration a reflected pair of hinge functions is selected instead of $e_k\left(x^i\right)$. Similarly, the regression method of choice in the generalized MARS model is the IRLS algorithm (57). Let us denote the MARS coefficient estimator by $\widehat{\boldsymbol{\beta}}_{\text{MARS}}$. As the model selection criterion serves the residual sum of squares, or equivalently, the negative of R squared. Note that the theory about AIC and ML estimation cannot be transferred hereto without any adjustments since the knots in the hinge functions act as additional degrees of freedom.

After each iteration, the set of candidate terms is extended by the products of the last two selected hinge functions with all hinge functions in C_1 that depend on risk factors of which the last two selected hinge functions do not depend on. Let the reflected pair selected in the first iteration ($k = 1$) be

$$\begin{aligned} h_1(X) &= \left(X_{l_1} - t_1\right)_+, \\ h_2(X) &= \left(t_1 - X_{l_1}\right)_+. \end{aligned} \tag{87}$$

Furthermore, let $C_{1,-} = C_1 \backslash \{h_1(X), h_2(X)\}$. Then, the set of candidate terms is updated at the beginning of the second iteration ($k = 2$) such that

$$\begin{aligned} C_2 = C_{1,-} &\cup \left\{(X_l - t)_+ h_1(X), (t - X_l)_+ h_1(X)\right\}_{t \in \left\{x_l^1, x_l^2, \ldots, x_l^N\right\} \,\mid\, l=1,\ldots,d,\; l \neq l_1} \\ &\cup \left\{(X_l - t)_+ h_2(X), (t - X_l)_+ h_2(X)\right\}_{t \in \left\{x_l^1, x_l^2, \ldots, x_l^N\right\} \,\mid\, l=1,\ldots,d,\; l \neq l_1}. \end{aligned} \tag{88}$$

The second set C_2 contains thus $2(dN-1)+4(d-1)N$ basis functions. Often, the order of interaction is limited to improve the interpretability of the proxy functions. Besides the maximum allowed number of terms, a minimum threshold for the decrease in the residual sum of squares can be employed as a termination criterion in the forward pass. Typically, the proxy functions generated in the forward pass overfit the data since model complexity is only penalized conservatively by stipulating a maximum number of basis functions and a minimum threshold.

13.6.3 Backward Pass & GCV

Due to the overfitting tendency of the proxy function generated in the forward pass, a backward pass is executed afterwards. Apart from the direction and slight differences, the backward pass is similar to the forward pass. In each iteration, the hinge function of which the removal causes the smallest increase in the residual sum of squares is removed and the backward model selection criterion for the resulting proxy function is evaluated. By this backward procedure, we generate the "best" proxy functions of each size in terms of the residual sum of squares. Out of all these best proxy functions, we finally select the one which minimizes the backward model selection criterion. As a result, the final proxy function will not only contain reflected pairs of hinge functions but also single hinge functions of which the complements have been removed. Optionally, the backward pass can also be omitted or alternatives which include combinations with forward steps could be implemented.

Let the number of basis functions in the MARS model be K, the number of knots be T and the smoothing parameter be c. With the effective degrees of freedom $\mathrm{df} = K + cT$, the standard choice for the backward model selection criterion is GCV, i.e.,

$$\mathrm{GCV} = \frac{ND\left(\widehat{\boldsymbol{\beta}}_{\mathrm{MARS}}\right)}{(N-\mathrm{df})^2}, \tag{89}$$

compare the definition in (70) for GAMs. For cases in which no interaction terms are allowed, Friedman & Silverman (1989) give a mathematical argument for using $c = 2$. For the other cases, Friedman (1991) concludes from a wide variety of simulation studies that a parameter of $c = 3$ is fairly effective. Across all these studies, $2 \leq c \leq 4$ was found to give the best value of c. Alternatively, but with significantly higher computational costs, c could be estimated by resampling techniques such as bootstrapping by Efron (1983) or cross-validation by Stone (1974).

13.6.4 Summary

The classical and generalized MARS algorithms in Section 13.6 are special cases of respectively the OLS regression algorithm and GLM algorithm, in which the basis functions are hinge functions and variable selection is carried out subsequently in a forward and backward pass. While in the forward pass the proxy functions are built up with respect to the residual sum of squares as the model selection criterion, in the backward pass they are cut back with respect to GCV where the degrees of freedom are modified to account for the knots in the hinge functions.

13.7 Kernel Regression

13.7.1 One-dimensional LC & LL Regression

Kernel regression is a type of locally weighted OLS regression where the weights vary with the input variable (the *target scenario*). This non-parametric regression approach using a kernel as the weighting function goes back to Nadaraya (1964) and Watson (1964).

We start with local constant (LC) and local linear (LL) regression in one dimension along the lines of Chapter 6 of Hastie et al. (2017). Thereby we carve out the idea of kernel regression which generalizes very naturally to more dimensions.

Let the target scenario be denoted by $x_0 \in \mathbb{R}$ and let the *univariate kernel* with given *bandwidth* $\lambda > 0$ be

$$K_\lambda\left(x_0, x^i\right) = D\left(\frac{\left|x^i - x_0\right|}{\lambda}\right), \tag{90}$$

where $D(\cdot)$ is the specified kernel function. While, for example, the Epanechnikov (see the green shaded areas of Figure 20 inspired by Hastie et al. (2017)), tri-cube and uniform kernels are commonly used kernel functions with bounded support, the gaussian kernel is one with infinite support. Moreover, the kernels can be defined with different orders, often the second order kernels are used, see e.g. Li & Racine (2007).

The LC kernel estimator or Nadaraya-Watson kernel smoother is given at each x_0 as the kernel-weighted average over the fitting values y^i, i.e.,

$$\widehat{f}_{\mathrm{LC}}\left(x_0\right) = \widehat{\beta}_{\mathrm{LC}}\left(x_0\right) = \frac{\sum_{i=1}^N K_\lambda\left(x_0, x^i\right) y^i}{\sum_{i=1}^N K_\lambda\left(x_0, x^i\right)}. \tag{91}$$

It is a continuous function since the weights die off smoothly with increasing distance from x_0. This locally constant function varies over the domain of target scenarios x_0 and therefore needs to be estimated separately at all of them. Due to the asymmetry of the

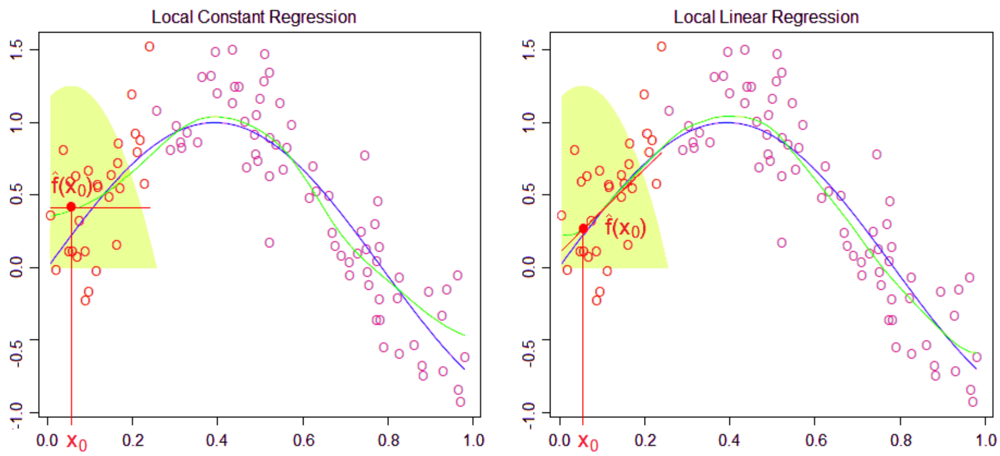

Figure 20: LC and LL kernel regression using the Epanechnikov kernel with $\lambda = 0.2$ in one dimension.

kernels at the boundaries of the domain, the LC kernel estimator (91) can be severely biased in that region, see the left panel of Figure 20.

We can overcome this problem by fitting locally linear functions instead of locally constant functions, see the right panel of Figure 20. At each target x_0, the LL kernel estimator is defined as the minimizer of the kernel-weighted residual sum of squares, i.e.,

$$\widehat{\boldsymbol{\beta}}_{\mathrm{LL}}(x_0) = \underset{\boldsymbol{\beta}(x_0)\in\mathbb{R}^2}{\arg\min}\left\{\sum_{i=1}^{N} K_\lambda\left(x_0, x^i\right)\left(y^i - \beta_0(x_0) - \beta_1(x_0)\, x^i\right)^2\right\}, \tag{92}$$

with $\boldsymbol{\beta}(x_0) = (\beta_0(x_0), \beta_1(x_0))^{\mathrm{T}}$. If we omit the linear term in (92) by setting $\beta_1 = 0$, the intercept $\widehat{\beta}_{\mathrm{LL},0}(x_0)$ of the LL kernel estimator becomes the LC kernel estimator (91). The proxy function at x_0 is given by

$$\widehat{f}_{\mathrm{LL}}(x_0) = \widehat{\beta}_{\mathrm{LL},0}(x_0) + \widehat{\beta}_{\mathrm{LL},1}(x_0)\, x_0. \tag{93}$$

Again the minimization problem (92) must be solved separately for all target scenarios so that the coefficients of the proxy function vary across their domain. For each target scenario x_0, (92) is a weighted least-squares (WLS) problem with weights $K_\lambda\left(x_0, x^i\right)$. Its solution is the WLS estimator

$$\widehat{\boldsymbol{\beta}}_{\mathrm{LL}}(x_0) = \left(Z^{\mathrm{T}} W(x_0)\, Z\right)^{-1} Z^{\mathrm{T}} W(x_0)\, \mathbf{y}, \tag{94}$$

with $\mathbf{y}$ as the response vector, $W(x_0) = \operatorname{diag}\left(K_\lambda\left(x_0, x^1\right), \ldots, K_\lambda\left(x_0, x^N\right)\right)$ as the weight matrix and Z as the design matrix containing row-wise the vectors $\left(1, x^i\right)^{\mathrm{T}}$. We call H the hat matrix if $\widehat{\mathbf{y}} = H\mathbf{y}$ such that $\widehat{\mathbf{y}} = \left(\widehat{f}_{\mathrm{LL}}\left(x^1\right), \ldots, \widehat{f}_{\mathrm{LL}}\left(x^N\right)\right)^{\mathrm{T}}$ contains the proxy function values at their target scenarios.

When we use proxy functions in LL regression that are composed of polynomial basis functions with exponents greater than one, we could also speak of local polynomial regression.

13.7.2 Multidimensional LC & LL Regression

We generalize LC regression to $\mathbb{R}^K$ based on Chapter 2 of Li & Racine (2007) by expressing the kernel with respect to the basis function vector $\mathbf{z} = (e_0(X), \ldots, e_{K-1}(X))^{\mathrm{T}}$ following from the adaptive forward stepwise selection with OLS regression and small $K_{\max}$. At each target scenario vector $\mathbf{z}_0 \in \mathbb{R}^K$ with elements z_{0k}, basis function vector $\mathbf{z}^i \in \mathbb{R}^K$ with elements z_{ik} evaluated at fitting scenario x^i and given bandwidth vector $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_{K-1})^{\mathrm{T}}$, the *multivariate kernel* is defined as the product of univariate kernels, i.e.,

$$K_{\boldsymbol{\lambda}}\left(\mathbf{z}_0, \mathbf{z}^i\right) = \prod_{k=0}^{K-1} D\left(\frac{|z_{ik} - z_{0k}|}{\lambda_k}\right). \tag{95}$$

The LC kernel estimator or Nadaraya-Watson kernel smoother in $\mathbb{R}^K$ is defined at each $\mathbf{z}_0$ as

$$\widehat{f}_{\mathrm{LC}}(\mathbf{z}_0) = \widehat{\beta}_{\mathrm{LC}}(\mathbf{z}_0) = \frac{\sum_{i=1}^{N} K_{\boldsymbol{\lambda}}\left(\mathbf{z}_0, \mathbf{z}^i\right) y^i}{\sum_{i=1}^{N} K_{\boldsymbol{\lambda}}\left(\mathbf{z}_0, \mathbf{z}^i\right)}. \tag{96}$$

Since we let $e_0(X)$ represent the intercept so that $z_{i0} = z_{00} = 1$, the corresponding univariate kernel $D\left(\frac{|z_{i0} - z_{00}|}{\lambda_0}\right) = D(0)$ is constant over all fitting points, thus cancels in (96) and can be omitted in (95).

The LL kernel estimator in $\mathbb{R}^K$ is given as the multidimensional analogue of (92) at each $\mathbf{z}_0$, i.e.,

$$\widehat{\boldsymbol{\beta}}_{\mathrm{LL}}\left(\mathbf{z}_0\right) = \underset{\boldsymbol{\beta}(\mathbf{z}_0)\in\mathbb{R}^K}{\arg\min}\left\{\sum_{i=1}^{N} K_{\boldsymbol{\lambda}}\left(\mathbf{z}_0, \mathbf{z}^i\right)\left(y^i - \mathbf{z}^{i,\mathrm{T}}\boldsymbol{\beta}\left(\mathbf{z}_0\right)\right)^2\right\}, \tag{97}$$

where $\boldsymbol{\beta}\left(\mathbf{z}_0\right) = \left(\beta_0\left(\mathbf{z}_0\right), \ldots, \beta_{K-1}\left(\mathbf{z}_0\right)\right)^{\mathrm{T}}$ and the proxy function at $\mathbf{z}_0$ is given by

$$\widehat{f}_{\mathrm{LL}}\left(\mathbf{z}_0\right) = \mathbf{z}_0^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{LL}}\left(\mathbf{z}_0\right). \tag{98}$$

The LL kernel estimator can again be computed by WLS regression, i.e.,

$$\widehat{\boldsymbol{\beta}}_{\mathrm{LL}}\left(\mathbf{z}_0\right) = \left(Z^{\mathrm{T}}W\left(\mathbf{z}_0\right)Z\right)^{-1}Z^{\mathrm{T}}W\left(\mathbf{z}_0\right)\mathbf{y}, \tag{99}$$

where $W\left(\mathbf{z}_0\right) = \operatorname{diag}\left(K_{\boldsymbol{\lambda}}\left(\mathbf{z}_0, \mathbf{z}^1\right), \ldots, K_{\boldsymbol{\lambda}}\left(\mathbf{z}_0, \mathbf{z}^N\right)\right)$ is the weight matrix and Z the design matrix containing row-wise the vectors $\mathbf{z}^{i,\mathrm{T}}$. The hat matrix H satisfies $\widehat{\mathbf{y}} = H\mathbf{y}$ with $\widehat{\mathbf{y}} = \left(\widehat{f}_{\mathrm{LL}}\left(\mathbf{z}^1\right), \ldots, \widehat{f}_{\mathrm{LL}}\left(\mathbf{z}^N\right)\right)^{\mathrm{T}}$ containing the proxy function values at their target scenario vectors.

13.7.3 Bandwidth Selection, AIC & LOO-CV

The bandwidths λ_k in kernel regression can be selected similarly to the smoothing parameters in GAMs by minimization of a suitable model selection criterion. In fact, kernel smoothers can be interpreted as local non-parametric GLMs with identity link functions. More precisely, at each target scenario the kernel smoother can be viewed as a GLM (52) where the parametric weights $V\left[\widehat{\mu}^i_{\mathrm{GLM}}\right]$ in (58) are the non-parametric kernel weights $K_{\boldsymbol{\lambda}}\left(\mathbf{z}_0, \mathbf{z}^i\right)$ in (97). Since GLMs are special cases of GAMs and the bandwidths in kernel regression can be understood as smoothing parameters, kernel smoothers and GAMs are sometimes lumped together in one category. If the numbers N of the fitting points and K of the basis functions are large, from a computational perspective it might be beneficial to perform bandwidth selection based on a reduced set of fitting points.

Hurvich et al. (1998) propose to select the bandwidths $\lambda_1, \ldots, \lambda_{K-1}$ based on an improved version of AIC which works in the context of non-parametric proxy functions that can be written as linear combinations of the observations. It has the form

$$\mathrm{AIC} = \log\left(\widehat{\sigma}^2\right) + \frac{1 + \operatorname{tr}(H)/N}{1 - \left(\operatorname{tr}(H) + 2\right)/N}, \tag{100}$$

where $\widehat{\sigma}^2 = \frac{1}{N}\left(\mathbf{y} - \widehat{\mathbf{y}}\right)^{\mathrm{T}}\left(\mathbf{y} - \widehat{\mathbf{y}}\right)$ and H is the hat matrix.

As an alternative, leave-one-out cross-validation (LOO-CV) is suggested by Li & Racine (2004) for bandwidth selection. Let us refer to

$$\widehat{\boldsymbol{\beta}}_{\mathrm{LL},-j}\left(\mathbf{z}_0\right) = \underset{\boldsymbol{\beta}(\mathbf{z}_0)\in\mathbb{R}^K}{\arg\min}\left\{\sum_{i=1, i\neq j}^{N} K_{\boldsymbol{\lambda}}\left(\mathbf{z}_0, \mathbf{z}^i\right)\left(y^i - \mathbf{z}^{i,\mathrm{T}}\boldsymbol{\beta}\left(\mathbf{z}_0\right)\right)^2\right\} \tag{101}$$

as the leave-one-out LL kernel estimator and to $\widehat{f}_{\mathrm{LL},-j}\left(\mathbf{z}_0\right) = \mathbf{z}_0^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{LL},-j}\left(\mathbf{z}_0\right)$ as the leave-one-out proxy function at $\mathbf{z}_0$. The objective of LOO-CV is to choose the bandwidths $\lambda_1, \ldots, \lambda_{K-1}$ which minimize

$$\mathrm{CV} = \frac{1}{N}\sum_{i=1}^{N}\left(y^i - \widehat{f}_{\mathrm{LL},-i}\left(\mathbf{z}_0\right)\right)^2. \tag{102}$$

13.7.4 Adaptive Forward Stepwise OLS Selection

A practical implementation of kernel regression can be found, for example, in the combination of functions *npreg(·)* and *npregbw(·)* from R package *np* of Racine & Hayfield (2018).

In the other sections, basis function selection depends on the respective regression methods. Since the crucial process of bandwidth selection in kernel regression takes a very long time in the implementation of our choice, it would be infeasible to proceed here in the same way. Therefore, we derive the basis functions for LC and LL regression by adaptive forward stepwise selection based on OLS regression, by risk factor wise linear selection or a combination thereof. Thereby, we keep the maximum allowed number of terms $K_{\max}$ rather small as we aim to model the subtleties by kernel regression.

13.7.5 Summary

The kernel regression algorithm in Section 13.7 is a non-parametric local regression approach using a kernel as a weighting function. While at each target point the LC kernel estimator is given as the kernel-weighted average, the LL kernel estimator minimizes everywhere the kernel-weighted residual sum of squares. To evaluate AIC at a kernel estimator, a non-parametric version accounting for the bandwidths is presented. As an alternative to AIC, non-parametric leave-one-out cross-validation LOO-CV is introduced. The bandwidths are selected such that they minimize the chosen model selection criterion. For reasons of computational efficiency, the adaptive basis function selection procedures need to be performed prior to the kernel regression approach based on, for example, OLS regression.

14 Numerical Experiments

14.1 General Remarks

14.1.1 Data Basis

In our slightly altered real-world example, the life insurance company has a portfolio with a large proportion of traditional annuity business. In order to challenge the regression techniques, the traditional annuity business features by construction very high interest rate guarantees so that the insurer suffers huge losses in low interest rate environments. We let the insurance company be exposed to $d = 15$ relevant financial and actuarial risk factors. For the derivation of the fitting points, we run its CFP model conditional on $N = 25,000$ fitting scenarios with each of these outer scenarios entailing two antithetic inner simulations. For a subset of the resulting fitting values of the best estimate liability (BEL), see Figure 16, for summary statistics, the left column of Table 4, and for a histogram, the left panel of Figure 21.

The Sobol validation set is generated based on $L = 51$ validation scenarios with $1,000$ inner simulations, where the 51 scenarios comprise 26 Sobol scenarios, one base scenario, 15 one-dimensional risk scenarios and 9 scenarios that turned out to be capital region scenarios in the previous year risk capital calculations.

The nested simulations set which is due to its high computational costs not available in the regular LSMC approach reflects the highest 5% real-world losses and is based on

	Fitting Values	Nested Simulation Values
Minimum:	10,883	12,479
1st quartile:	13,824	14,515
Median:	14,907	14,940
Mean:	14,922	14,922
3rd quartile:	15,989	15,330
Maximum:	19,354	17,080
Std. deviation:	1,519	610
Skewness:	0.067	-0.081
Kurtosis:	2.478	3.214

Table 4: Summary statistics of fitting and nested simulation values of BEL.

$L = 1,638$ outer scenarios with respectively 4,000 inner simulations. From the 1,638 real-world scenarios, 14 exhibit extreme stresses far beyond the bounds of the fitting space and are therefore excluded from the analysis. For the remaining nested simulation values of BEL, see Figure 17, for summary statistics, the right column of Table 4, and for a histogram, the right panel of Figure 21. The corresponding data set still comprises 107 points with scenarios lying outside of the fitting space. Although these points distort the validation figures slightly, we deliberately keep them in the data set to present a more realistic picture of the regression techniques for prediction. As the univariate risk factor distributions are modeled by unbounded distributions, the fitting space covers only a subspace of the realizations of the insurer's real-world distribution. Therefore, it is from a practical perspective important that a proxy function also shows a reasonably good extrapolation behavior.

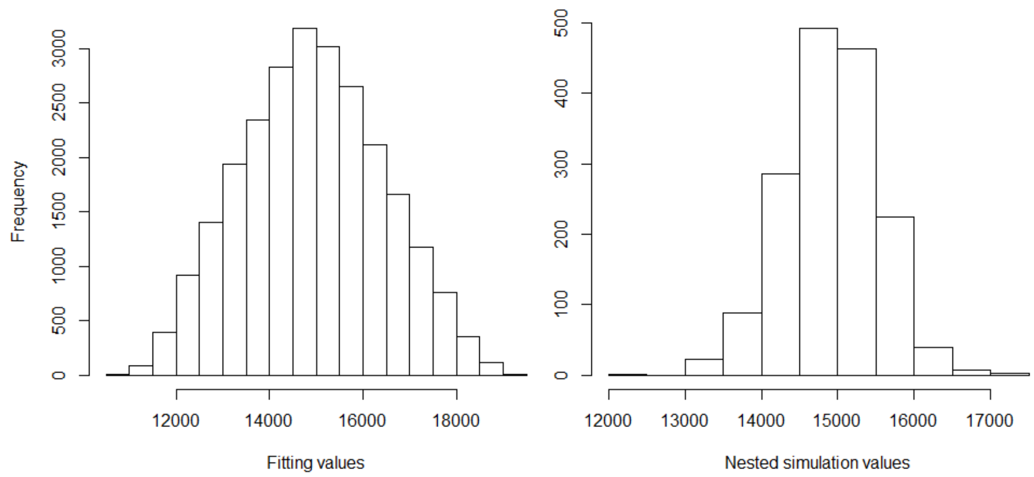

Figure 21: Histograms of fitting and nested simulation values of BEL.

The capital region set consists of the $L = 129$ nested simulations points which correspond to the nested simulations SCR estimate (= 99.5% highest loss) and the 64 losses above and below (= 99.3% to 99.7% highest losses). While one point with an extreme scenario is excluded from this data set, 24 points with scenarios lying outside of the fitting space are kept in it.

For the sake of completeness, we evaluate the validation figures for the four best models from Sections 14.2–14.5 based on further reduced nested simulations and capital region sets in Sections 14.2.6, 14.3.6, 14.4.8 and 14.5.7, and compare them to the results obtained based on the two full sets. The reduced sets are exclusively composed of points with scenarios lying inside the fitting space. Therefore, the corresponding validation figures demonstrate how the four models perform in settings where extrapolation is not required.

14.1.2 Validation Figures

We will output validation figure (38) with respect to the relative and asset metric, and figures (39), (40) and (41). In all tables of the appendix, figures (40) and (41) are evaluated on the Sobol set, i.e., v.mae^0, v.res^0, with respect to base value $y^0_{1{,}000} = 14{,}646.7$, which is a result of averaging over $1{,}000$ inner simulations. On the nested simulations set, i.e., ns.mae^0, ns.res^0, and capital region set, i.e., cr.mae^0, cr.res^0, these figures are everywhere computed with respect to base value $y^0_{16{,}000} = 14{,}661.1$, which is a result of averaging over $16{,}000$ inner simulations. The tables in the four sections about the behavior on the reduced validation sets also report figures v.mae^0 and v.res^0 with respect to the latter base value: While the first row in these tables contains duplicates of the results from the appendix, the second row contains the new evaluations. Since base value $y^0_{16{,}000}$ is associated with a lower standard error, it is supposed to be the more reliable one. Therefore, it is worth noting that figure v.res^0 from the tables in the appendix can easily be transformed such that it is calculated with respect to $y^0_{16{,}000}$ as well by subtracting from v.res^0 the difference of $y^0_{16{,}000} - y^0_{1{,}000} = 14.4$ which the two base values incur. For the majority of the derived proxy functions such a transformation will in fact reveal a higher approximation quality than suggested by the tables.

We will not explicitly state the base residual (42) as it is just (39) minus (41). Unlike figures (38) and (39), figures (40) and (41) do not forgive a bad fit of the base value if the validation values are well approximated by a proxy function. Contrariwise, if a proxy function shows the same systematic deviation from the validation values and base value, (40) and (41) will be close to zero whereas (38) and (39) will be not.

The residual figures (39) and (41) have to be interpreted relative to the MAE figures (38) and (40). This means, we only speak of a bias if a residual figure is large compared to the MAE figures. Besides that, for residual figures ns.res, ns.res^0, cr.res and cr.res^0, the homogeneity resulting from the ordering of the underlying validation data sets (highest 5% losses) has to be taken into account. While a large residual implies a bias, a small residual does not necessarily indicate no bias since a small residual might as well be caused by a sign change of a bias along the ordering.

14.1.3 Economic Variables

We derive the OLS proxy functions for two economic variables, namely for the best estimate liability (BEL) and available capital (AC) over a one-year risk horizon, i.e., $Y(X) \in \{\mathrm{BEL}(X),\ \mathrm{AC}(X)\}$. Their approximation quality is assessed by validation figures (38) with respect to the relative and asset metric and (39). Essentially, AC is obtained as the market value of assets minus BEL. Since the market value of assets only depends on financial risk factors, AC reflects the negative behavior of BEL with respect to the actuarial risk factors. Due to the high similarity of BEL and AC, we will only derive BEL proxy functions with the other regression methods. The profit resulting from a certain risk constellation captured by an outer scenario X can be computed as $\mathrm{AC}(X)$ minus the base AC. Validation figures (40) and (41) address the approximation quality of this difference. Taking the negative of the profit yields the loss and evaluating the loss at all real-world scenarios the real-world loss distribution from which the SCR is derived as the 99.5% value-at-risk. The out-of-sample performances of two different OLS proxy functions of BEL on the Sobol, nested simulations and capital region sets serve as the benchmark for the other regression methods.

14.1.4 Numerical Stability

Let us discuss the subject of numerical stability of QR decompositions in the OLS regression design under a monomial basis. If the weighting in the weighted least-squares problems associated with GLMs, heteroscedastic FGLS regression and kernel regression is good-natured, similar arguments apply as they can also be solved via QR decompositions according to Green (1984) where the weighting is just a scaling. However, the weighting itself raises additional numerical questions that need to be taken into consideration when making the regression design choices. In GLMs, these choices are the random component (50) and link function (52), in FGLS regression it is the functional form of the heteroscedastic variance model (80) and in kernel regression it is the kernel function (95). The following arguments do not apply to GAMs and MARS models as these are constructed out of spline functions, see (63) and (85), respectively. In GAMs, the penalty matrix increases numerical stability.

McLean (2014) justifies that from the perspective of numerical stability performing a QR decomposition on a monomial design matrix Z is asymptotically equivalent to using a Legendre design matrix Z' and transforming the resulting coefficient estimator into the monomial one. Under the assumption of an orthonormal basis, Weiß & Nikolić (2019) have derived an explicit upper bound for the condition number of the non-diagonal matrix $\frac{1}{N}(Z')^{\mathrm{T}}(Z')$ for $N < \infty$, where the factor $\frac{1}{N}$ is used for technical reasons. This upper bound increases in (1) the number of basis functions, (2) the Hardy-Krause variation of the basis, (3) the convergence constant of the low-discrepancy sequence, and (4) the outer scenario dimension. The type of restriction setting which we define in Section 14.2.1 controls aspect (1) through the specification of $K_{\max}$ and aspect (2) through the limitation of exponents $d_1 d_2 d_3$. Aspects (3) and (4) are beyond the scope of the calibration and validation steps of the LSMC framework and therefore left aside here.

14.1.5 Interpolation & Extrapolation

In the LSMC framework, let us refer by interpolation to prediction inside the fitting space and by extrapolation to prediction outside the fitting space. Runge (1901) found that high-degree polynomial interpolation at equidistant points can oscillate toward the ends of the interval with the approximation error getting worse the higher the degree is. In a least-squares problem, Runge's phenomenon was shown by Dahlquist & Björck (1974) not to apply to polynomials of degree d fitted based on N equidistant points if the inequality $d < 2\sqrt{N}$ holds. With $N = 25,000$ fitting points the inequality becomes $d < 316$ so that we clearly do not have to impose any further restrictions in OLS, FGLS and kernel regression as well as in GLMs to keep this phenomenon under control. Splines as they occur in GAMs and MARS models do not suffer from this oscillation issue by construction.

Since Runge's phenomenon concerns the ends of the interval and the real-world scenarios for the insurer's full loss distribution forecast in the fourth step of the LSMC framework partly go beyond the fitting space, its scope comprises the extrapolation area as well. High-degree polynomial extrapolation can worsen the approximation error and therefore play a crucial role if many real-world scenarios go far beyond the fitting space.

14.1.6 Principle of Parsimony

Another problem that can occur in an adaptive algorithm is overfitting. Burnham & Anderson (2002) state that overfitted models often have needlessly large sampling variances which means that their precision of the predictions is poorer than that of more parsimonious models which are also free of bias. In cases where AIC leads to overfitting, implementing restriction settings of the form $K_{\max}$ - $d_1 d_2 d_3$ becomes relevant for adhering to the principle of parsimony.

14.2 Ordinary Least-Squares (OLS) Regression

14.2.1 Settings

We build the OLS proxy functions (46) of $Y(X) \in \{\mathrm{BEL}(X),\ \mathrm{AC}(X)\}$ with respect to an outer scenario X out of monomial basis functions that can be written as $e_k(X) = \prod_{l=1}^{15} X_l^{r_k^l}$ with $r_k^l \in \mathbb{N}_0$ so that each basis function can be represented by a 15-tuple $\left(r_k^1, \ldots, r_k^{15}\right)$. The final proxy function depends on the restrictions applied in the adaptive algorithm. The purpose of setting restrictions is to guarantee numerical stability, to keep the extrapolation behavior under control and the proxy functions parsimonious. To illustrate the impact of restrictions, we run the adaptive algorithm for BEL under two different restriction settings with the second one being so relaxed that it will not take effect in our example. Additionally, we run the adaptive algorithm under the first restriction setting for AC to give an example of how the behavior of BEL can transfer to AC. As the first ingredient of our restriction setting acts the maximum allowed number of terms $K_{\max}$. Furthermore, we limit the exponents in the monomial basis. Firstly we apply a uniform threshold to all exponents, i.e., $r_k^l \leq d_1$. Secondly we restrict the degree, i.e., $\sum_{l=1}^{15} r_k^l \leq d_2$. Thirdly we restrict the exponents in the interaction basis functions, i.e., if there are some $l_1 \neq l_2$ with $r_k^{l_1},\ r_k^{l_2} > 0$, we require $r_k^{l_1},\ r_k^{l_2} \leq d_3$. Let us denote this type of restriction setting by $K_{\max}$ - $d_1 d_2 d_3$.

As the first and second restriction settings, we choose 150-443 and 300-886, respectively, motivated by Teuguia et al. (2014) who found in their LSMC example in Chapter 4 with four risk factors and $50,000$ fitting scenarios entailing two inner simulations that the validation error computed based on 14 validation scenarios started to stabilize at degree 4 when using monomial or Legendre basis functions in different adaptive basis function selection procedures. Furthermore, they pointed out that the LSMC approach becomes infeasible for degrees higher than 12.

We apply R function *lm(·)* implemented in R package *stats* of R Core Team (2018).

14.2.2 Results

Table A2 contains the final BEL proxy function derived under the first restriction setting 150-443 with the basis function representations and coefficients. Thereby reflect the rows the iterations of the adaptive algorithm and depict thus the sequence in which the basis functions are selected. Moreover, the iteration-wise AIC scores and out-of-sample

MAEs (38) with respect to the relative metric in % on the Sobol, nested simulations and capital region sets are reported, i.e., v.mae, ns.mae and cr.mae. Table A3 contains the AC counterpart of the BEL proxy function derived under 150-443 and Table A4 the final BEL proxy function derived under the more relaxed restriction setting 300-886. Tables A5 and A6 indicate respectively for the BEL and AC proxy functions derived under 150-443 the AIC scores and all five previously defined validation figures evaluated on the Sobol, nested simulations and capital region sets after each tenth iteration. Similarly, Table A7 reports these figures for the BEL proxy function derived under 300-886. Here the last row corresponds to the final iteration.

Thereafter, we manipulate the validation values on all three validation sets twice insofar as we subtract respectively add pointwise 1.96 times the standard errors from respectively to them (inspired by the 95% confidence interval of the normal distribution). We then evaluate the validation figures for the final BEL proxy functions under both restriction settings on these manipulated sets of validation value estimates and depict them in Table A8 in order to assess the impact of the Monte Carlo error associated with the validation values. In addition, Table 5 provides information on how the best OLS model performs in terms of extrapolation.

14.2.3 Improvement by Relaxation

Tables A2 and A3 state that the adaptive algorithm terminates under 150-443 for both BEL and AC when the maximum allowed number of terms is reached. This gives reason to relax the restriction setting to, for instance, 300-886 which eventually lets the algorithm terminate due to no further reduction in the AIC score without hitting restrictions 886, compare Table A4 for BEL. In fact, only restrictions 224-464 are hit. Except for the already very small figures cr.mae, cr.mae^a and cr.res all validation figures are further improved by the additional basis functions, see Tables A5 and A7. The largest improvement takes place between iterations 180 and 190. The result that at maximum degrees 464 are selected is consistent with the result of Teuguia et al. (2014) who conclude in their numerical examples of Chapter 4 that under a monomial, Legendre or Laguerre basis the optimum degree is probably 4 or 5. Besides that, Bauer & Ha (2015) derive a similar result in their one risk factor LSMC example of Chapter 6 when using $50,000$ fitting scenarios and Legendre, Hermite, Chebychev basis functions or eigenfunctions.

According to our Monte Carlo error impact assessment in Table A8, the slight deterioration at the end of the algorithm is not sufficient to indicate an overfitting tendency of AIC. Under the standard choices of the five major components, compare Section 12.2, the adaptive algorithm manages thus to provide a numerically stable and parsimonious proxy function even without a restriction setting. Here, permitting a priori unlimited degrees of freedom is beneficial to capturing the complex interactions in the CFP model.

However, the standardized residual plot in Figure 22 shows for BEL that OLS regression is not able to fully model the variance structure. The standardized residuals $\frac{\widehat{\epsilon}_i}{s_{\mathrm{OLS}}}$ indicate a decrease of the variance with respect to risk factor X_1. Therefore, the assumption of homoscedastic errors is slightly violated in this numerical example.

14.2.4 Reduction of Bias

Overall, the systematic deviations indicated by the mean errors (39) and (41) are reduced significantly on the three validation sets by the relaxation but not completely eliminated.

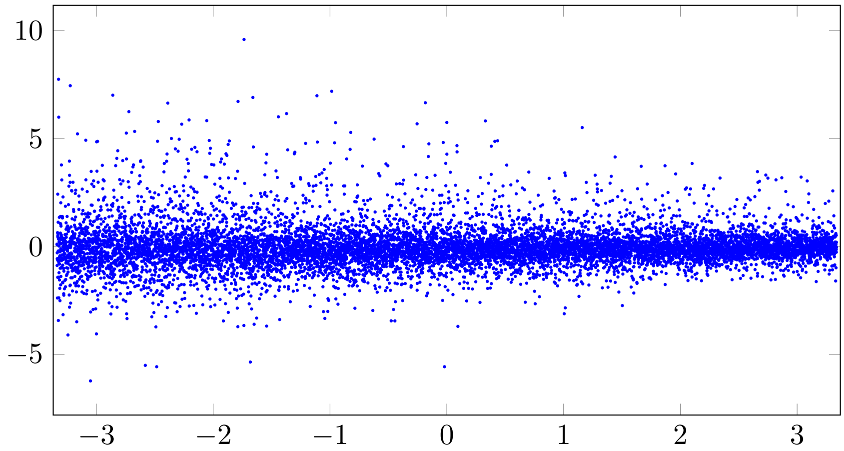

Figure 22: Standardized residual plot of the best OLS model with respect to X_1 indicating a slight violation of homoscedasticity.

For the 300-886 OLS residuals on the three sets, see the diamond-shaped residuals in Figures 23–25. While the reduction of the bias comes along with the general improvement stated above, the remainder of the bias indicates that sample size is not sufficiently large or that the functional form is not flexible enough to replicate the complex interactions in the CFP model. The polynomials might not be able to capture the sudden changes in steepness of BEL and AC which are a consequence of regulation and management actions.

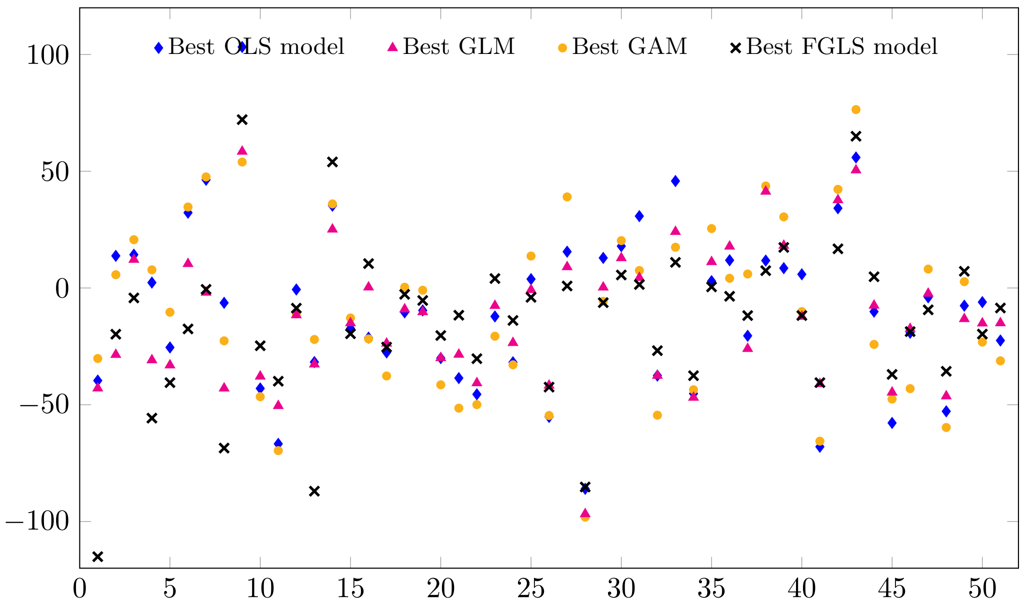

Figure 23: Residual plots on Sobol set.

The comparisons $|\text{v.res}| < |\text{v.res}^0|$, $|\text{cr.res}| < |\text{cr.res}^0|$ but $|\text{ns.res}| > |\text{ns.res}^0|$, holding under both restriction settings, indicate that on the Sobol and capital region sets primarily the base value is not approximated well whereas on the nested simulations set not only the base value but also the validation values are missed. The MAEs capture this result, too, i.e., $\text{v.mae}, \text{cr.mae} < \text{ns.mae}$ but $\text{ns.mae}^0 < \text{v.mae}^0, \text{cr.mae}^0$.

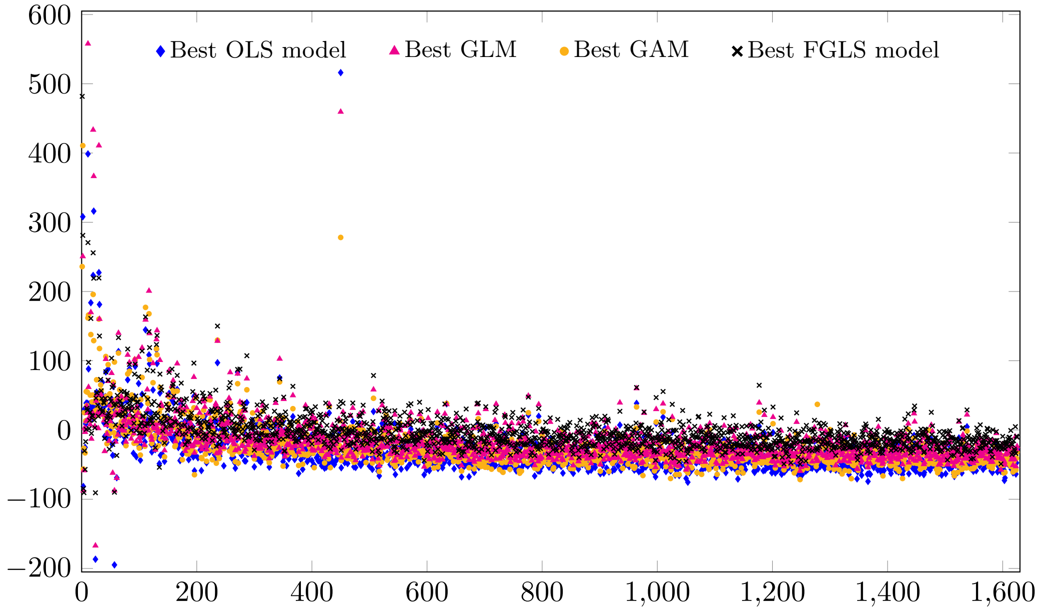

Figure 24: Residual plots on nested simulations set.

14.2.5 Relationship between BEL & AC

The MAEs with respect to the relative metric are much smaller for BEL than for AC since the two economic variables are subject to similar absolute fluctuations with, for example, in the base case BEL being approximately 20 times the size of AC. The similar absolute fluctuations are reflected by the iteration-wise very similar MAEs with respect to the asset metric of BEL and AC, compare v.mae^a, ns.mae^a and cr.mae^a given in % in Tables A5 and A6. Furthermore, they manifest themselves in the iteration-wise opposing mean errors v.res, v.res^0, ns.res and cr.res as well as in the similarly-sized MAEs v.mae^0, ns.mae^0 and cr.mae^0.

14.2.6 Reduced Validation Sets

Table 5 displays the out-of-sample validation figures of the best derived OLS proxy function of BEL evaluated based on the Sobol and the full and reduced nested simulations and capital region sets after the final iteration. Thereby, this table reports figures v.mae^0 and v.res^0 in the first row with respect to base value $y^0_{1,000}$ and in the second row with respect to more reliable base value $y^0_{16,000}$.

While the MAEs are consistently improved when excluding the points with scenarios lying outside of the fitting space, the mean errors seem to be impacted more randomly. The reason behind this seemingly random behavior is related to the sign change of the bias of the proxy function along the ordering of the nested simulations and capital region sets. Take, for example, figure cr.res which is equal to 0.8 when referring to the full capital region set, and equal to -10.3 when referring to the reduced capital region set. Now consider the diamond-shaped residuals in Figure 25 which yield the almost ideal mean error of 0.8 but at the same time show a systematic pattern. The residuals in the left half

k	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
Best OLS model evaluated based on full validation sets															
224	0.194	0.186	−8.7	6.659	34.3	0.268	0.259	−30.2	4.200	−1.6	0.168	0.165	0.8	5.007	29.4
Best OLS model evaluated based on reduced validation sets															
224	0.194	0.186	−8.7	5.329	19.9	0.253	0.244	−34.6	3.608	−6.0	0.124	0.121	−10.3	4.159	18.4

Table 5: Out-of-sample validation figures of the best derived OLS proxy function of BEL. MAEs in %.

of the diagram indicate an underestimation of BEL by the proxy function whereas the residuals in the right half indicate an overestimation. Removing the points with scenarios lying outside of the fitting space affects more residuals in the left than in the right half of the diagram. Therefore, the residuals in the right half dominate the mean error after the removal of the points. As a result, the mean error on the reduced capital region set becomes more negative and ends up at −10.3.

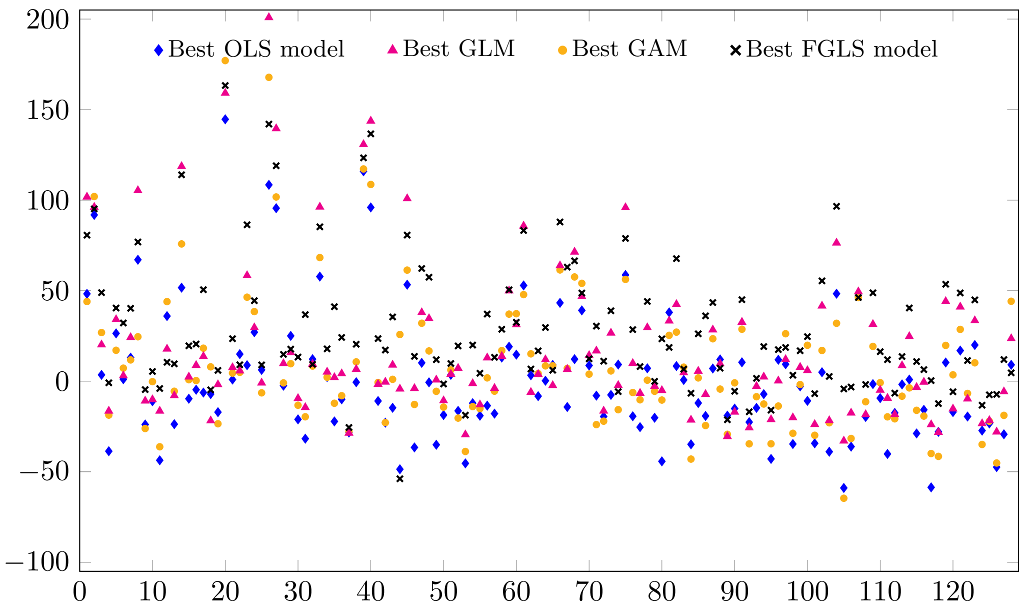

Figure 25: Residual plots on capital region set.

The second row of the table reveals for the best derived OLS proxy function an actually higher approximation quality on the Sobol set than suggested by the first row and thus all corresponding tables in the appendix. Moreover, it follows that the proxy function loses a part of its accuracy when extrapolating.

14.2.7 Summary

We applied the OLS regression algorithm in Section 14.2 under suitable restriction settings and found that relaxing the setting from 150-443 to 300-886 (i.e., no actual restriction) improved out-of-sample performance considerably. Thereby the bias indicated by the mean errors on the three validation sets was reduced, see Tables A5 and A7, but not eliminated so that we stated that the functional form of the proxy function still had some flaws, see Figure 24. We concluded that overall the adaptive algorithm managed to provide a numerically stable and parsimonious proxy function even without imposing a restriction setting and that the a priori unlimited degrees of freedom served to capture the complex CFP model better. Furthermore we pointed out that BEL and AC were subject to similar absolute fluctuations. Lastly, we showed that the best OLS model lost a part of its accuracy when extrapolating.

14.3 Generalized Linear Models (GLMs)

14.3.1 Settings

We derive the GLMs (52) of BEL under restriction settings 150-443 and 300-886 which we also employed for the derivations of the OLS proxy functions. Thereby, we run each restriction setting with the canonical choices of random components for continuous (non-negative) response variables, that is, the normal, gamma and inverse gaussian distributions, compare McCullagh & Nelder (1989). In cases where the economic variable can also attain negative values (e.g., AC), a suitable shift of the response values in a preceding step would be required. We combine each of the three random component choices with the commonly used identity, inverse and log link functions, i.e., $g(\mu) \in \left\{\mathrm{id}(\mu), \frac{1}{\mu}, \log(\mu)\right\}$, compare Chambers & Hastie (1992). In combination with the inverse gaussian random component, we consider additionally link function $\frac{1}{\mu^2}$. Further choices are conceivable but go beyond this first shot.

We take R function *glm(·)* implemented in R package *stats* of R Core Team (2018).

14.3.2 Results

While Tables A9–A11 display the AIC scores and five previously defined validation figures after each tenth iteration for the just mentioned combinations under 150-443, Tables A12–A14 do so under 300-886 and include the final iterations. Table A15 gives an overview of the AIC scores and validation figures corresponding to all considered final GLMs and highlights in green and red respectively the best and worst values observed per figure. Lastly, Table 6 comments on how the best GLM performs on extrapolated areas.

14.3.3 Improvement by Relaxation

The OLS regression is the special case of a GLM with normal random component and identity link function. This is why the first sections of Tables A9 and A12 coincide respectively with Tables A5 and A7. The adaptive algorithm terminates under 150-443 not only for this combination but also for all other ones when the maximum allowed number of terms is reached. Under 300-886 termination occurs due to no further reduction in the AIC score without hitting the restrictions - the different GLMs stop between the values 208-454 and 250-574 are attained.

For all GLMs except for the one with gamma random component and identity link, the AIC scores and eight most significant validation figures for measuring the approximation quality, namely leftmost figure v.mae to rightmost figure ns.res in the tables, are improved through the relaxation as can be seen in Table A15. For gamma random component with identity link, the deteriorations are negligible. Overall, figures ns.mae^0 and cr.mae^0 are deteriorated by at maximum 0.5% points and figures ns.res^0 and cr.res^0 by at maximum 4 units. Figures cr.mae and cr.mae^a are especially small under 150-443 so that slight deteriorations by at maximum 0.05% points under 300-886 towards the levels of v.mae and v.mae^a or ns.mae and ns.mae^a are not surprising. Similar arguments apply to the acceptability of the maximum deterioration of cr.res by 13 to 17 units for inverse gaussian with $\frac{1}{\mu^2}$ link. We conclude that the more relaxed restriction setting 300-886 performs better than 150-443 for all GLMs in our numerical example. This result appears plausible in comparison with the OLS result from the previous section and is hence also consistent with the OLS results of Teuguia et al. (2014) and Bauer & Ha (2015).

AIC cannot be said to show an overfitting tendency according to Tables A12–A14 and also Table A8 since the validation figures do not deteriorate in the late iterations more than they underlie Monte Carlo fluctuations, compare the OLS interpretation. Using GLMs instead of OLS regression in the standard adaptive algorithm, compare Section 12.2, lets the algorithm thus maintain its property to yield numerically stable and parsimonious proxy functions even without restriction settings.

14.3.4 Reduction of Bias

According to Table A15, inverse gaussian with $\frac{1}{\mu^2}$ link shows the most significant decrease in v.mae by -0.088% points when moving from 150-443 to 300-886. Under 300-886 this combination even outperforms all other ones (highlighted in green) whereas under 150-443 it is vice versa (highlighted in red). Hence, the performance of a random component link combination under 150-443 does not generalize to 300-886. On the Sobol and nested simulations sets, the MAEs (38) are not only considerably lower for inverse gaussian with $\frac{1}{\mu^2}$ link than for all others but also the closest together even when the capital region set is included. This speaks for a great deal of consistency.

In fact, the systematic overestimation of 81% of the points on the nested simulations set by inverse gaussian with $\frac{1}{\mu^2}$ link is certainly smaller than, for example, that of 89% by normal with identity link but still very pronounced. On the capital region set, the overestimation rates for these two combinations are 41% and 56%, respectively, meaning that here the bias is negligible. Surprisingly, for most GLMs the bias is here smaller than for inverse gaussian with $\frac{1}{\mu^2}$ link but since this result does not generalize to the nested simulations set, we regard the rather mediocre performance of inverse gaussian with $\frac{1}{\mu^2}$ link as a chance event. Interpreting the mean errors (39) provides similar insights.

In particular, for inverse gaussian $\frac{1}{\mu^2}$ link GLM the reduction of the bias comes along with the general improvement by the relaxation. The small remainder of the bias indicates not only that this GLM is a promising choice here but also that identifying well-suited regression methods and functional forms is crucial to further improving the accuracy of the proxy function. For the residuals on the three sets, see the triangle-shaped residuals in Figures 23–25.

14.3.5 Major & Minor Role of Link Function & Random Component

Apart from the just considered case, for all three random components, the relaxation to 300-886 yields the largest out-of-sample performance gains in terms of v.mae with identity link (between -0.047% and -0.058% points), closely followed by log link (between -0.033% and -0.047% points), and the least gains with inverse link (between -0.017% and -0.020% points). While with identity link the largest improvements before finalization take place for normal, gamma and inverse gaussian random components between iterations 180 to 190, 170 to 180, and 150 to 160, respectively, with log link they occur much sooner between iterations 120 to 130, 110 to 120, and 110 to 120, respectively, see Tables A12–A14. As a result of this behavior, under 150-443 log link performs better than identity link for normal and inverse gaussian whereas under 300-886 it is vice versa. Inverse link always performs worse than identity and log links, in particular under 300-886.

Applying the same link with different random components does not bring much variation under 300-886 with gamma and inverse gaussian being slightly better than normal for all considered links though. This observation is in line with the slight skewness of the distribution of BEL resulting from the asymmetric profit sharing mechanism in the CFP model: While the policyholders are entitled to participate at the surpluses of an insurance company, see e.g. Mourik (2003), the company has to bear its losses fully by itself. Since the normal random component performs only slightly worse than the skewed distributions, it should still be considered for practical reasons because it has a closed-form solution and a great deal of statistical theory has been developed for it, compare e.g. Dobson (2002). By conclusion, the choice of the link is more important than that of the random component so that trying alternative link functions might be beneficial.

14.3.6 Reduced Validation Sets

Table 6 indicates the out-of-sample validation figures of the best derived GLM of BEL evaluated based on the Sobol and the full and reduced nested simulations and capital region sets after the final iteration. Again, figures v.mae^0 and v.res^0 are calculated with respect to base value $y^0_{1,000}$ in the first row and base value $y^0_{16,000}$ in the second row.

k	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
Best GLM evaluated based on full validation sets															
250	0.174	0.166	−12.4	5.058	25.5	0.193	0.186	−14.6	3.833	8.8	0.188	0.184	17.3	6.266	40.8
Best GLM evaluated based on reduced validation sets															
250	0.174	0.166	−12.4	4.067	11.1	0.169	0.163	−19.6	3.046	3.8	0.107	0.105	2.5	4.942	26.0

Table 6: Out-of-sample validation figures of the best derived GLM of BEL. MAEs in %.

For the best GLM, we observe a similar pattern as for the best OLS model in Section 14.2.6 when excluding the extrapolation effects. While all MAEs are reduced when moving from the first to the second row, the mean errors behave a little more randomly. But the overall improvement for the best GLM is clearly more distinct than that of the best OLS model. The best GLM benefits particularly from removing the extrapolated points from the capital region set with even both mean errors cr.res and cr.res^0 improving significantly. See Figure 25 for an illustration of this result: There are comparatively many triangle-shaped residuals indicating a strong underestimation of BEL by the best GLM in the left two thirds of the diagram which disappear through the removal. Furthermore, the overall bias of the best GLM is much less severe than that of the best OLS model, compare also Figure 24, which makes the performance gains even higher.

Figures v.mae^0 and v.res^0 from the second row expose that the approximation quality of the best derived GLM is also actually higher than suggested by the first row and thus all corresponding tables in the appendix. Additionally, it can be concluded that the originally rather mediocre performance of the best GLM on the capital region set is partially driven by extrapolation.

14.3.7 Summary

Like in the OLS regression algorithm, we observed in all applied GLM algorithms in Section 14.3 that relaxing the setting from 150-443 to 300-886 (i.e., no actual restriction) helped to improve out-of-sample performance and reduce the bias. From the small remainder of the bias we deduced that identifying suitable regression methods and functional forms is crucial to further improving the accuracy of the proxy functions. We concluded that the adaptive algorithm maintained its property to yield numerically stable and parsimonious proxy functions without requiring restriction settings in the GLM context. The performance of a random component link combination under 150-443 did not generalize to 300-886. Moreover, we saw in the variation of the results that the choice of the link was more important than that of the random component so that regarding additional link functions might be beneficial. While continuous skewed random components led to slightly advantageous out-of-sample performances, the use of the normal random component had practical advantages. Compared to the OLS regression routine, there were GLM routine designs with better out-of-sample performances. While performing best on both the Sobol and nested simulations set, 300-886 inverse gaussian $\frac{1}{\mu^2}$ link GLM showed only a mediocre performance on the capital region set which was partially shown to be driven by extrapolation. For an overview of the results, see Table A15.

14.4 Generalized Additive Models (GAMs)

14.4.1 Settings

For the derivation of the GAMs (64) of BEL, we apply only restriction settings $K_{\max}$-443 with $K_{\max} \leq 150$ in the adaptive algorithm since we use smooth functions (63) constructed out of splines that may already have exponents greater than 1 to which the monomial first-order basis functions are raised. As the model selection criterion we take GCV (70) used by our chosen implementation by default. We vary different ingredients of GAMs while holding others fixed to carve out possible effects of these ingredients on the approximation quality of GAMs in adaptive algorithms and our application.

We rely on R function *gam(·)* implemented in R package *mgcv* of Wood (2018).

14.4.2 Results

Table A16 contains the validation figures for GAMs with varying number of spline functions per smooth function, i.e., $J \in \{4, 5, 8, 10\}$, after each tenth and the finally selected smooth function. In the case of adaptive forward stepwise selection the iteration numbers coincide with the numbers of selected smooth functions. In contrast, table sections with adaptive forward stagewise selection results do not display the iteration numbers in the smooth function column k. In Table A17, we display the effective degrees of freedom, p-values and significance codes of each smooth function of the $J = 4$ and $J = 10$ GAMs from the previous table at stages $k \in \{50, 100, 150\}$. The p-values and significance codes are based on a test statistic by Marra & Wood (2012) having its foundations in the frequentist properties of Bayesian confidence intervals analyzed in Nychka (1988). Tables A18 and A19 report the validation figures respectively for GAMs with numbers $J = 5$ and $J = 10$, where the type of the spline functions is varied. Thin plate regression splines, penalized cubic regression splines, duchon splines and Eilers and Marx style P-splines are considered. Thereafter, Tables A20 and A21 display the validation figures respectively for

GAMs with numbers $J = 4$ and $J = 8$ and different random component link function combinations. As in GLMs, we apply the normal, gamma and inverse gaussian distributions with identity, log, inverse and $\frac{1}{\mu^2}$ (only inverse gaussian) link functions.

Table A22 compares by means of two exemplary GAMs the effects of adaptive forward stagewise selection of length $L = 5$ and adaptive forward stepwise selection. Next, Table A23 contains a mixture of GAMs challenging the results which we will have deduced from the other GAM tables. Table A24 gives an overview of the validation figures corresponding to all derived final GAMs and highlights in green and red respectively the best and worst values observed per figure. In addition, Table 7 shows how the best GAM behaves in terms of extrapolation.

14.4.3 Improvement by Tailoring the Spline Function Number

Table A16 indicates that the MAEs (38) and (40) of the exemplary GAMs built up of thin plate regression splines with normal random component and identity link tend to increase with the number J of spline functions per dimension until $k = 100$. Running more iterations reverses this behavior until $k = 150$. Hence, as long as comparatively few smooth functions have been selected in the adaptive algorithm fewer spline functions tend to yield better out-of-sample performances of the GAMs whereas many smooth functions tend to perform better with more spline functions. A possible explanation of this observation is that an omitted-variable bias due to too few smooth functions is aggravated here by an overfitting due to too many spline functions. For more details on an omitted-variable bias, see e.g. Pindyck & Rubinfeld (1998), and for the needlessly large sampling variances and thus low estimation precision of overfitted models, see e.g. Burnham & Anderson (2002). Differently, the absolute values of the mean errors (39) and (41) tend to become smaller with increasing J regardless of k.

According to Table A17, the components of the effective degrees of freedom (69) associated with each smooth function tend to decrease slightly in k for $J = 4$ and $J = 10$. This is plausible as the explanatory power of each additionally selected smooth term is expected to decline by trend in the adaptive algorithm. Conditional on df > 1, that is for proportions of at least 40% of all smooth terms, the averages of the effective degrees of freedom belonging to $k \in \{50, 100, 150\}$ amount for $J = 4$ and $J = 10$ to $\{2.494, 2.399, 2.254\}$ and $\{5.366, 4.530, 4.424\}$, respectively. The values are by construction smaller than or equal to $J - 1$ since one degree of freedom per smooth function is lost to the identifiability constraints. Hence, for at least 40% of the smooth functions, on average $J = 6$ is a reasonable choice to capture the CFP model properly while maintaining computational efficiency, compare Wood (2017). The other side of the coin here is that up to 60% of the smooth functions are supposed to be replaceable by simple linear terms without losing accuracy so that here tremendous efficiency gains can be realized by making the GAMs more parsimonious. Furthermore, setting J individually for each smooth function can help to improve computational efficiency (if J should be set below average) and out-of-sample performance (if J should be set above average). However, such a tailored approach entails the challenge that the optimal J per smooth function is not stable across all k, compare row-wise the degrees of freedom in the table for $J = 4$ and $J = 10$.

14.4.4 Dependence of Best Spline Function Type

According to Tables A18 and A19, the adaptive algorithm terminates only due to no further decrease in GCV when the GAMs are composed of duchon splines, which are discussed in Duchon (1977). Whether GCV has an overfitting tendency here cannot be deduced from this example since only restriction settings with $K_{\max} \leq 150$ are tested. The thin plate regression splines by Wood (2003) and penalized cubic regression splines by Wood (2017) perform similarly to each other and significantly better than the duchon splines for both $J = 5$ and $J = 10$. For $J = 5$, the Eilers and Marx style P-splines proposed by Eilers & Marx (1996) perform by far best when $K_{\max} = 100$ smooth functions are allowed. However, for $J = 10$ they are outperformed by both the thin plate regression splines and penalized cubic regression splines when between $K_{\max} = 125$ and 150 smooth functions are allowed. This result illustrates well that the best choice of the spline function type varies with J and $K_{\max}$, meaning that it should be selected together with these parameters.

14.4.5 Minor Role of Link Function & Random Component

For GLMs, we have seen that varying the random component barely alters the validation results whereas varying the link function can make a noticeable impact. While this result mostly applies to the earlier compositions of GAMs in the adaptive algorithm as well, it certainly does not to the later ones. See, for instance, early composition $k = 40$ in Table A20. Here, identity link GAMs with gamma and inverse gaussian random components perform more similarly to each other than identity and log link GAMs with gamma random component or identity and log link GAMs with inverse gaussian random component do. Log link GAMs with gamma and inverse gaussian random components show such a behavior as well. However identity link GAM with the less flexible normal random component (no skewness) does not show at all a behavior similar to that of identity link GAMs with gamma or inverse gaussian random components. Now see later compositions $k \in \{70, 80\}$ to verify that all available GAMs in the table produce very similar validation results.

For another example see Table A21. For early composition $k = 50$, identity link GAMs with normal and gamma random components behave very similarly to each other just like log link GAMs with normal and gamma random components do. For later compositions $k \in \{100, 110\}$, again all available GAMs produce very similar validation results. A possible explanation of this result is that the impact of the link function and random component decreases with the number of smooth functions as the latter take over the modeling. By conclusion, the choices of the random component and link function do not play a major role when the GAMs are built up of many smooth functions.

14.4.6 Consistency of Results

Table A22 shows based on two exemplary GAMs constructed out of $J = 8$ thin plate regression splines per dimension varying in the random component and link function that the adaptive forward stagewise selection of length $L = 5$ and adaptive forward stepwise selection lead to very similar GAMs and validation results. As a result, stagewise selection should be preferred due to its considerable run time advantage. As we will see below, the run time can be further reduced without any drawbacks by dynamically selecting even more than 5 smooth functions per iteration.

The purpose of Table A23 is to challenge the hypotheses deduced above. Like Table A16, this table contains the results of GAMs with varying spline function number $J \in \{5, 8, 10\}$ and fixed spline function type. Instead of thin plate regression splines, now Eilers and Marx style P-splines are considered. Since adaptive forward stepwise and stagewise selection do not yield significant differences in the examples of Table A22, we do not expect that permutations thereof affect the results much here as well. This permits us to randomly assign three different adaptive forward selection approaches to the three exemplary proxy function derivation procedures. As one of these approaches, we choose a dynamic stagewise selection approach in which L is determined in each iteration as the proportion 0.25 of the size of the candidate term set. Again we see that as long as only $k \in \{90, 100\}$ smooth functions have been selected, $J = 5$ performs better than $J = 8$ and $J = 8$ better than $J = 10$. However, $k = 150$ smooth functions are not sufficient this time for $J = 10$ to catch up with the performance of $J = 5$. The observed performance order is consistent with the hypotheses of a high robustness of the GAMs with respect to the adaptive selection procedure and random component link function combination.

14.4.7 Potential of Improved Interaction Modeling

Table A24 presents as the most suitable GAM the one with highest allowed maximum number of smooth functions $K_{\max} = 150$ and highest number of spline functions $J = 10$ per dimension. The slight deterioration after $k = 130$ reported by Table A16 indicates that at least one of the parameters is already comparatively high. According to Table A17, there are a few smooth terms which might benefit from being composed of more than ten spline functions and increasing $K_{\max}$ might be helpful to capturing the interactions in the CFP model more appropriately, particularly in the light of the fact that the best GLM, having 250 basis functions, outperforms the best GAM on both the Sobol and nested simulations set, compare Table A15. The best GAM shows a comparatively low bias across the three validation sets though, see the dot-shaped residuals in Figures 23–25. Variations in the random component link function combination and adaptive selection procedure are not expected to change the performance much. By conclusion, we recommend the fast normal identity link GAMs (several expressions in the PIRLS algorithm simplify) with tailored spline function numbers per smooth function and simple linear terms under stagewise selection approaches of suitable lengths $L \geq 5$ and more relaxed restriction settings where $K_{\max} > 150$.

14.4.8 Reduced Validation Sets

Table 7 reports the out-of-sample validation figures of the best derived GAM of BEL evaluated based on the Sobol and the full and reduced nested simulations and capital region sets after the final iteration. As above, figures v.mae^0 and v.res^0 are computed with respect to base value $y^0_{1{,}000}$ in the first row and base value $y^0_{16{,}000}$ in the second row.

The best derived GAM behaves very similarly to the best GLM from Section 14.3.6 regarding the extrapolation effects. Comparing the MAEs row-wise shows that the MAEs in the second row are again consistently smaller. But the decrease of the MAEs is a little less pronounced here than for the best GLM. The mean errors tend to get smaller as well when excluding the extrapolated points. The improvement of the approximation quality on the capital region set strikes again the most. While, for example, for the best GLM figure cr.mae goes from 0.188% to 0.107%, for the best GAM it goes only from 0.173% to

k	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
Best GAM evaluated based on full validation sets															
150	0.212	0.203	−9.8	7.070	36.8	0.230	0.223	−24.3	3.575	7.9	0.173	0.170	8.3	6.337	40.4
Best GAM evaluated based on reduced validation sets															
150	0.212	0.203	−9.8	5.837	22.4	0.220	0.212	−28.3	3.031	3.8	0.120	0.117	−3.6	5.609	28.6

Table 7: Out-of-sample validation figures of the best derived GAM of BEL. MAEs in %.

0.120%.

The second row of the table reveals this time an actually higher approximation quality of the best derived GAM on the Sobol set than suggested by the first row and thus all corresponding tables in the appendix. Moreover, it follows that the best GAM performs similarly to the best OLS model from Section 14.2.6 when it comes to extrapolation.

14.4.9 Summary

We ran the different GAM algorithms in Section 14.4 only under restriction settings $K_{\max}$-443 with $K_{\max} \leq 150$. Whether GCV had an overfitting tendency in the adaptive algorithm could therefore not be assessed. We saw that as long as comparatively few smooth functions had been selected fewer spline functions performed better whereas many smooth functions did better with more spline functions, compare Table A16. We gave a possible explanation of these effects by arguing that an omitted-variable bias due to too few smooth functions might have been aggravated here by an overfitting due to too many spline functions. In order to realize the efficiency and performance gains incentivized by Table A17 by making the GAMs more parsimonious, we proposed to set the spline function numbers individually for each smooth function and to use linear terms where sufficient. Another result was that the spline function type should be selected conditional on the spline function number(s) and number of smooth functions, see Tables A18 and A19. As soon as the GAM had been composed of many smooth functions, the choices of both the link and random component turned out to be less crucial which made us recommended the fast normal identity GAMs in the exemplary application, compare Tables A20 and A21. Since adaptive forward stagewise selection of length $L = 5$ and adaptive forward stepwise selection led to very similar GAMs according to Table A22, we suggested to use the former selection approach due to its run time advantage. From the fact that the best found GLM had 250 terms and outperformed the best found GAM reported in Table A24, we deduced that using more than 150 smooth functions might improve the results. Lastly, it followed that the best GAM performed similarly good at extrapolation as the best OLS model.

14.5 Feasible Generalized Least-Squares (FGLS) Regression

14.5.1 Settings

Like the OLS proxy functions and GLMs, we derive the FGLS proxy functions (76) under restriction settings 150-443 and 300-886. For the performance assessment of FGLS regression, we apply type I and II algorithms with variance models of different complexity, where the type I results are obtained as a by-product of the type II algorithm since the latter algorithm builds upon the former one. We control the complexity through the maximum allowed numbers of variance model terms $M_{\max} \in \{2, 6, 10, 14, 18, 22\}$.

We combine R functions *nlminb(·)* and *lm(·)* implemented in R package *stats* of R Core Team (2018).

14.5.2 Results

Tables A25 and A26 display respectively the adaptively selected FGLS variance models of BEL corresponding to maximum allowed numbers of terms $M_{\max}$ based on final 150-443 and 300-886 OLS proxy functions given in Tables A2 and A4. For reasons of numerical stability and simplicity, only basis functions with exponents summing up to at max two are considered as candidates. Additionally, the AIC scores and MAEs with respect to the relative metric are reported in the tables. By construction, these results are simultaneously the type I algorithm outcomes. Tables A27 and A28 summarize respectively under 150-443 and 300-886 all iteration-wise out-of-sample test results of the type I FGLS proxy functions. The results of the type II algorithm after each tenth and the final iteration of adaptive FGLS proxy function selection are respectively displayed by Tables A29 and A30 for the two restriction settings. Table A31 gives an overview of the AIC scores and validation figures corresponding to all final FGLS proxy functions and highlights as in the previous overview tables in green and red respectively the best and worst values observed per figure. Lastly, Table 8 reveals how the best FGLS model performs on extrapolated areas.

14.5.3 Consistency Gains by Variance Modeling

By looking at Tables A25 and A26 we see similar out-of-sample performance patterns during adaptive variance model selection based on the basis function sets of 150-443 and 300-886 OLS proxy functions. In both cases, the p-values of Breusch-Pagan test indicate that heteroscedasticity is not eliminated but reduced when the variance models are extended, i.e., when $M_{\max}$ is increased. For instance, the standardized residual plot in Figure 26 confirms for the type II $M_{\max} = 14$ proxy function derived under 300-886 that

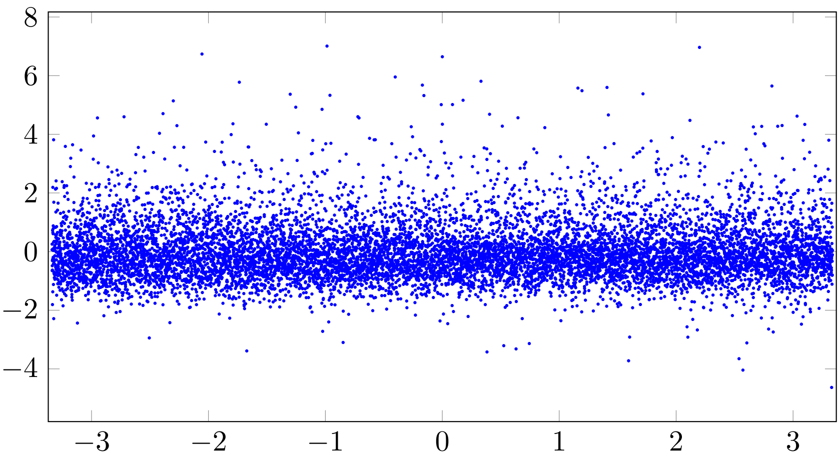

Figure 26: Standardized residual plot of the best FGLS model with respect to X_1 indicating heteroscedasticity modeling.

our FGLS regression approach captures a fair amount of variance structure. Accordingly, as opposed to the standardized OLS residuals in Figure 22, the standardized FGLS residuals $\frac{\widehat{\epsilon}_i}{\widehat{\sigma}_i}$, where $\widehat{\sigma}_i = \sqrt{\exp\left(\mathbf{v}^{i,\mathrm{T}}\widehat{\boldsymbol{\alpha}}_{\mathrm{ML}}\right)}$, and which are exactly the residuals of transformed regression problem (71) here, suggest independence of the variance from risk factor X_1. In fact, in a more good-natured LSMC example Hartmann (2015) shows that a type I alike algorithm manages to fully eliminate heteroscedasticity. While the MAEs (38) barely change on the Sobol set, they decrease significantly on the nested simulations set and increase noticeably on the capital region set. Under 300-886 the effects are considerably smaller than under 150-443 since the capital region performance of 300-886 OLS proxy function is less extraordinarily good than that of 150-443 OLS proxy function. The MAEs on the three sets approach each other under both restriction settings. Hence the reductions in heteroscedasticity lead to consistency gains across the three validation sets.

Tables A27 and A28 complete the just discussed picture. The remaining validation figures on the Sobol set improve through type I FGLS regression slightly compared to OLS regression. Like ns.mae, figure ns.res and the base residual improve a lot with increasing $M_{\max}$ under 150-443 and a little less under 300-886 but ns.mae^0 and ns.res^0 do not alter much as the aforementioned two figures cancel each other out here. On the capital region set, the figures deteriorate or remain comparatively high in absolute values. The type I FGLS figures converge fast so that increasing $M_{\max}$ successively from 10 to 22 barely affects the out-of-sample performance anymore. As a result of heteroscedasticity modeling, the proxy functions are shifted such that overall approximation quality increases. Unfortunately, this does not guarantee an improvement in the relevant region for SCR estimation as our example illustrates well.

14.5.4 Monotonicity in Complexity

Let us address the type II FGLS results under 150-443 in Table A29 now. For $M_{\max} = 2$, figures (40) and (41) are improved on all three validation sets significantly compared to OLS regression with the type I figures lying in-between. The other validation figures are similar for OLS, type I and II FGLS regression, which traces the performance gains in (40) and (41) back to a better fit of the base value. For $M_{\max} = 6$ to 22, the type II figures show the same effects as the type I ones but more pronouncedly. These effects are by trend the more distinct the more complex the variance models become. The type II figures stabilize less than the type I ones because of the additional variability coming along with adaptive FGLS proxy function selection. Hartmann (2015) shows in terms of Sobol figures in her LSMC example that increasing the complexity while omitting only one regressor from the variance model can deteriorate the out-of-sample performance dramatically. Intuitively, it is plausible that the FGLS validation figures are the farther from the OLS figures away the more elaborately heteroscedasticity is modeled.

Now let us relate the type II FGLS results under 300-886 in Table A30 to the other FGLS results. Under 300-886 for $M_{\max} = 2$, figures (40) and (41) are already at a comparatively good level with both OLS and type I FGLS regression so that they do not alter much or even deteriorate with type II FGLS regression. Like under 150-443 for $M_{\max} = 6$ to 22, the type II figures show the effects of the type I ones more pronouncedly. Under both restriction settings, ns.mae and ns.res decrease thereby significantly. While this barely causes ns.res^0 to change under 150-443, it lets ns.res^0 increase in absolute values under 300-886. The slight improvements on the Sobol set and the deteriorations on the capital region set carry over to 300-886. When $M_{\max}$ is increased up to 22, the type II FGLS validation figures under 300-886 do not stop fluctuating. The variability entailed by adaptive FGLS proxy function selection intensifies thus through the relaxation of the restriction setting in this numerical example. According to Breusch-Pagan test, heteroscedasticity is neither eliminated by the type II algorithm here nor by a type II alike approach of Hartmann (2015) in her more good-natured example.

14.5.5 Improvement by Relaxation

Among all FGLS proxy functions listed in Table A31, we consider type II with $M_{\max} = 14$ in variance model selection under 300-886 as the best performing one. Apart from nested simulations validation under the type I algorithm, 300-886 performs better than 150-443. Since on the other hand the type II algorithm performs better than the type I algorithm under the respective restriction settings, 300-886 and the type II algorithm are the most promising choices here. Differently $M_{\max} = 14$ does not constitute a stable choice due to the high variability coming along with 300-886 and the type II algorithm.

While all type I FGLS proxy functions are by definition composed of the same basis functions as the OLS proxy function, the compositions of the type II FGLS proxy functions vary with $M_{\max}$ because of their renewed adaptive selections. Consequently, under 300-886 all type I FGLS proxy functions hit the same restrictions 224-464 as the OLS proxy function, whereas the restrictions hit by the type II FGLS proxy functions vary between 224-454 and 258-564. This variation is consistent with the OLS and GLM results from the previous sections and the OLS results of Teuguia et al. (2014) and Bauer & Ha (2015).

AIC does not have an overfitting tendency according to Tables A27–A30 as the validation figures do not deteriorate in the late iterations more than they underly Monte Carlo fluctuations, compare the OLS and GLM interpretations. Using FGLS instead of OLS regression in the standard adaptive algorithm, compare Section 12.2, lets the algorithm thus yield numerically stable and parsimonious proxy functions without restriction settings as well.

14.5.6 Reduction of Bias

The type II $M_{\max} = 14$ FGLS proxy function under 300-886 reaches with 258 terms the highest observed number across all numerical experiments and not only outperforms all derived GLMs and GAMs in terms of combined Sobol and nested simulations validation, it also shows by far the smallest bias on these two validation sets and approximates the base value comparatively well. This observation speaks for a high interaction complexity of the CFP model. The reduction of the bias comes again along with a general improvement by the relaxation. Given the fact that the capital region set presents the most extreme and challenging validation set in our analysis, the still mediocre performance here can be regarded as acceptable for now. Nevertheless, especially the bias on this set motivates the search for even more suitable regression methods and functional forms. For the residuals of the best FGLS proxy function on the three sets, see the x-shaped residuals in Figures 23–25.

14.5.7 Reduced Validation Sets

Table 8 contains the out-of-sample validation figures of the best derived FGLS proxy function of BEL evaluated based on the Sobol and the full and reduced nested simulations and capital region sets after the final iteration. Again, figures v.mae^0 and v.res^0 are calculated with respect to base value $y^0_{1,000}$ in the first row and base value $y^0_{16,000}$ in the second row.

k	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
Best FGLS model evaluated based on full validation sets															
258	0.172	0.165	−14.4	4.371	10.4	0.134	0.129	−2.1	3.504	8.2	0.214	0.210	28.2	6.063	38.6
Best FGLS model evaluated based on reduced validation sets															
258	0.172	0.165	−14.4	3.868	−4.0	0.114	0.110	−5.8	2.961	4.6	0.144	0.141	17.6	5.341	27.9

Table 8: Out-of-sample validation figures of the best derived FGLS proxy function of BEL. MAEs in %.

In this table, we see again a decrease in all MAEs when transitioning from the first to the second row. Like for the best OLS model, GLM and GAM, this effect is the most pronounced on the capital region set. For the best FGLS model, the mean errors also behave a little more randomly but are overall reduced. The pattern following from the exclusion of the extrapolated points is similar to the one of the best GLM because the approximation quality on the capital region set improves considerably both in terms of MAEs and mean errors. However, like for the best GLM, the latter can be viewed as a comparatively likely outcome given the rather mediocre performance beforehand.

Figures v.mae^0 and v.res^0 from the second row show this time that the approximation quality of the best derived FGLS model is actually higher than suggested by the first row and thus all corresponding tables in the appendix. Furthermore, the table suggests that the best FGLS model performs similarly good at extrapolation as the best GLM.

14.5.8 Summary

Among the applied FGLS algorithms in Section 14.5, the type I algorithms led to consistency gains across the three validation sets. According to Breusch-Pagan test, they induced at least a reduction in heteroscedasticity in the generalized least-squares problem, which tended to be the more pronounced the more complex the variance models became but which converged fast, compare Tables A25 and A26. Despite the overall improvement in out-of-sample performance and the base approximation, they led to a deterioration in the relevant region for SCR estimation. The type II algorithms showed the effects of the type I algorithms in an amplified and more volatile way. The type II routines under 300-886 (i.e., no actual restriction) constituted systematically the best choices except for on the most extreme and challenging capital region set where their performances were still acceptable. But there was no systematically best choice of variance model complexity due to the high variability accompanied by the type II routines under 300-886. The best found FGLS routine reached with 258 terms the highest observed number across all numerical experiments and outperformed all considered GLM and GAM routines in terms of combined Sobol and nested simulations validation. Furthermore, it had by far the lowest bias on these two validation sets. This result spoke again for a high interaction complexity of the CFP model. We concluded that the adaptive algorithm maintained its property to yield numerically stable and parsimonious proxy functions without requiring restriction settings in the FGLS context. Nonetheless, the bias of the best FGLS routine on the capital region set motivated the search for even more suitable regression methods and functional forms, see Figure 25. Like for the best GLM, the bias was partially shown to be driven by extrapolation. For an overview of the results, see Table A31.

14.6 Multivariate Adaptive Regression Splines (MARS)

14.6.1 Settings

We undertake two-step approaches to identify well-suited generalized MARS models out of numerous possibilities conditional on smoothing parameter $c \in \{0, 2, 3\}$. It turns out that, regardless of c, comparatively few basis functions are selected in the forward passes of the numerical experiments. As higher values of c penalize the selection of additional basis functions even more severely, we forgo testing values $c > 3$. Furthermore, we see that the three applied choices of c lead to MARS models with very similar out-of-sample performances. To keep the analysis simple, we therefore decide to limit the subsequent reporting to the case $c = 0$.

In the first step, we vary several MARS ingredients over a wide range and obtain in this way a large number of different MARS models. To be more specific, we vary the maximum allowed number of terms $K_{\max} \in \{50, 113, 175, 237, 300\}$ and the minimum threshold for the decrease in the residual sum of squares $t_{\min} \in \{0, 1.25, 2.5, 3.75, 5\} \cdot 10^{-5}$ in the forward pass, the order of interaction $o \in \{3, 4, 5, 6\}$ and pruning method $p \in \{\text{'n'}, \text{'b'}, \text{'f'}, \text{'s'}\}$ with 'n' = 'none', 'b' = 'backward', 'f' = 'forward' and 's' = 'seqrep' in the backward pass, and the random component link function combination of the GLM extension. In addition to the 10 random component link function combinations applied in the numerical experiments of the GLMs, compare, for instance, Table A15, we use poisson random component with identity, log and squareroot link functions. We work with the default fast MARS parameter fast.k = 20 of our chosen implementation.

We use R function *earth(·)* implemented in R package *earth* of Milborrow (2018).

14.6.2 Results

In total, these settings yield $5 \cdot 5 \cdot 4 \cdot 4 \cdot 13 = 5{,}200$ MARS models with a lot of duplicates in our first step. We validate the $5{,}200$ MARS models on the Sobol, nested simulations and capital region sets through evaluation of the five validation figures. Then we collect the five best performing MARS models in terms of each validation figure per set which gives us in total $5 \cdot 5 = 25$ best performing models per first step validation set. Since the MAEs (38) with respect to the relative and asset metric entail the same best performing models, only $5 \cdot 4 = 20$ of the collected models per first step set are potentially different. Based on the ingredients of each of these 20 MARS models per first step set, we define $5 \cdot 5 = 25$ new sets of ingredients varying only with respect to $K_{\max}$ and $t_{\min}$ and derive the corresponding new but similar MARS models in the second step. As a result, we obtain in total $20 \cdot 25 = 500$ new MARS models per first step set. Again, we assess their out-of-sample performances through evaluation of the five validation figures on the three validation sets. Out of the 500 new MARS models per first step set, we collect then the best performing ones in terms of each validation figure per second step set. Now this gives us in total $5 \cdot 3 = 15$ best MARS models per first step set, or taking into account that the MAEs (38) with respect to the relative and asset metric entail once more the same best performing models, $4 \cdot 3 = 12$ potentially different best models per first step set. In total, this makes $12 \cdot 3 = 4 \cdot 9 = 36$ best MARS models, which can be found in Table A32 sorted by first and second step validation sets.

14.6.3 Poor Interaction Modeling & Extrapolation

In Table A32, the out-of-sample performances of all MARS models derived in our two-step approach are sorted using the first step validation set as the primary and the second step validation set as the secondary sort key. Let us address the first step second step validation set combinations by the headlines in Table A32. By construction, the combinations Sobol set^2, Nested simulations set^2 and Capital region set^2 yield respectively the MARS models with the best validation figures (38), (39), (40) and (41) on the Sobol, nested simulations and capital region sets. See that in the table all corresponding diagonal elements are highlighted in green. But the best MAEs (38) and (40) are not even close to what OLS regression, GLMs, GAMs and FGLS regression achieve. Finding small mean errors (39) and (41) regardless of the other validation figures is not sufficient. The performances on the nested simulations and capital region sets, comprising several scenarios beyond the fitting space, are especially poor. All these results indicate that MARS models do not seem very suitable for our application. Despite the possibility to select up to 300 basis functions, the MARS algorithm selects only at maximum 148 basis functions, which suggests that without any alterations, the algorithm is not able to capture the behavior of the CFP model properly, in particular extrapolation behavior is comparatively poor.

The MARS model with the set of ingredients $K_{\max} = 50$, $t_{\min} = 0$, $o = 4$, $p =$ 'b', inverse gaussian random component and identity link function is selected as the best one six times out of 36, or once for each Sobol and nested simulations first step validation set combination. Furthermore, this model performs best in terms of v.res^0, ns.mae^0 and ns.mae^a. Since there is no other MARS model with a similarly high occurrence and performance, we consider it the best performing and most stable one found in our two-step approach. For illustration of a MARS model, see this one in Table A33. The fact that this best MARS model performs worse than other ones in terms of several validation figures stresses the infeasibility of MARS models in this application.

14.6.4 Limitations

Table A32 suggests that, up to a certain upper limit, the higher the maximum allowed number of terms $K_{\max}$ the higher tends the performance on the Sobol set to be. However, this result does not generalize to the nested simulations and capital region sets. Since at maximum 148 basis functions are selected here even if up to 300 basis functions are allowed, extending the range of $K_{\max}$ in the first step of this numerical experiment would not affect the output in this regard. The threshold $t_{\min}$ is an instrument controlling the number of basis functions selected in the forward pass up to $K_{\max}$ which cannot be extended below zero, meaning that its variability has already been exhausted here as well. For the interaction order o similar considerations as for $K_{\max}$ apply. The pruning method p used in the backward pass does not play a large role compared to the other ingredients as it only helps to reduce the set of selected basis functions. In terms of Sobol validation, inverse gaussian random component with identity link performs best, whereas in terms of nested simulations and capital region validation, inverse gaussian random component with any link or log link with normal or poisson random component perform best. We conclude that if there was a suitable MARS model for our application, our two-step approach would have found it.

14.6.5 Summary

By applying a great variety of MARS algorithms in Section 14.6 in a two-step approach, we ensured that no comparatively well suited MARS model would have been missed in our analysis. All tested MARS algorithms selected at maximum 148 basis functions and showed rather poor out-of-sample performances as well as a weak extrapolation behavior compared to the previously discussed regression approaches, see Table A32. The conclusion was that MARS routines were not able to model the complex interactions in the CFP model appropriately.

14.7 Kernel Regression

14.7.1 Settings

We make a series of adjustments affecting either the structure or the derivation process of the multidimensional LC and LL proxy functions (96) and (98) to get as broad a picture of the potential of kernel regression in our application as possible. Our adjustments concern the kernel function and its order, the bandwidth selection criterion, the proportion of fitting points used for bandwidth selection, and the sets of basis functions of which the local proxy functions are composed of. Thereby we combine in various ways the gaussian, Epanechnikov and uniform kernels, orders $o \in \{2, 4, 6, 8\}$, bandwidth selection criteria LOO-CV and AIC, and between $2,500$ (proportion bw $= 0.1$) and $25,000$ (proportion bw $= 1$) fitting points for bandwidth selection.

We work with R functions *npregbw(·)* and *npreg(·)* implemented in R package *np* of Racine & Hayfield (2018).

14.7.2 Results

Additionally, we alternate the four basis function sets contained in Tables A34 and A35. The first two basis function sets with $K_{\max} \in \{16, 27\}$ are derived by adaptive forward stepwise selection based on OLS regression, the third one with $K_{\max} = 15$ by risk factor wise linear selection and the last one with $K_{\max} = 22$ by a combination thereof. All combinations including their out-of-sample performances can be found in Table A36. Again, the best and worst values observed per validation figure are highlighted in green and red, respectively.

14.7.3 Poor Interaction Modeling & Extrapolation

We draw the following conclusions based on the validation results in Table A36. The comparisons of LC and LL regression applied with gaussian kernel and 16 basis functions or Epanechnikov kernel and 15 basis functions suggest that LL regression performs better than LC regression. However, even the best Sobol, nested simulations and capital region results of LL regression are still outperformed by OLS regression, GLMs, GAMs and FGLS regression. Possible explanations for this observation are that kernel regression is not able to model the interactions of the risk factors equally well with its few basis functions and that local regression approaches perform rather poorly close to and especially beyond the boundary of the fitting space because of the thinned out to missing data basis in this region. While the first explanation applies to all three validation sets, the latter one applies only to the nested simulations and capital region sets on which the validation figures are indeed worse than on the Sobol set. While LC regression produces interpretable results with the sets of 22 and 27 basis functions, the more complex LL regression does not in most cases.

14.7.4 Limitations

On the Sobol and capital region sets, both LC and LL regression show similar behaviors when relying on gaussian kernel and 16 basis functions compared to Epanechnikov kernel and 15 basis functions. But on the nested simulations set, gaussian kernel and 16 basis functions are the superior choices. Using a uniform kernel with LC regression deteriorates the out-of-sample performance. The results of LC regression indicate furthermore that an extension of the basis function sets from 15 to 27 only slightly affects the validation performance. With gaussian kernel switching from 16 to 27 basis functions barely has an impact and with Epanechnikov kernel only the nested simulations and capital region validation performance improve when using 27 as opposed to 15, 16 or 22 basis functions. While increasing the order of the gaussian or Epanechnikov kernel deteriorates the validation figures dramatically, for the uniform kernel the effects can go in both directions. AIC performs worse than LOO-CV when used for bandwidth selection of the gaussian kernel in LC regression. For LC regression, increasing the proportion of fitting points entering bandwidth selection improves all validation figures until a specific threshold is reached. But thereafter the nested simulations and capital region figures are deteriorated. For LL regression no such deterioration is observed.

Overall we do not see much potential in kernel regression for our practical example compared to most of the previously analyzed regression methods. Nonetheless in order to achieve comparatively good kernel regression results, we consider LL regression more promising than LC regression due to the superior but still poor modeling close to and beyond the boundary of the fitting space. We would apply it with gaussian, Epanechnikov or other similar kernel functions. A high proportion of fitting points for bandwidth selection is recommended and it might be worth trying alternative comparatively small basis function sets reflecting the risk factor interactions better than in our examples.

14.7.5 Summary

We applied numerous variants of kernel regression algorithms in Section 14.7. We found that the LL regression algorithms performed better than the LC ones but still worse than the previously discussed routines, see Table A36. We traced the rather poor out-of-sample performances back to an insufficient interaction modeling by too few basis functions and a poor behavior of local regression approaches close to and beyond the boundary of the fitting space.

15 Conclusion

Summary

For high-dimensional variable selection applications such as the calibration step in the LSMC framework, we have presented various deterministic machine learning regression approaches ranging from ordinary and generalized least-squares regression variants over GLM and GAM approaches to multivariate adaptive regression splines and kernel regression approaches. At first we have justified the combinability of the ingredients of the regression routines such as the estimators and proposed model selection criteria in a theoretical discourse. Afterwards we have applied numerous configurations of these machine learning routines to the same slightly disguised real-world example in the LSMC framework. With the aid of different validation figures, we have analyzed the results, compared the out-of-sample performances and advised to use certain routine designs.

Outlook

In our slightly disguised real-world example and LSMC setting, the adaptive OLS regression, GLM, GAM and FGLS regression algorithms turned out to be well-suited machine learning methods for proxy modeling of life insurance companies with potential for both performance and computational efficiency gains by fine-tuning model hyperparameters and implementation designs. For recommendations of specific hyperparameter settings and designs, see the aforementioned suggestions. Differently, the MARS and kernel regression algorithms were not found to be convincing in our application. To study the robustness of our results, the approaches can be repeated in multiple other LSMC examples.

After all, none of our tested approaches was able to completely eliminate the bias observed in the validation figures and to yield consistent results across the three validation sets though. Investigations on whether these observations are systematic for the approaches, a result of the Monte Carlo error or a combination thereof help to further narrow down the circle of recommended regression techniques. To assess the variance and bias of the proxy functions, seed stability analyses in which the sets of fitting points are varied and convergence analyses in which sample size is increased need to be carried out. While such analyses would be computationally very costly, they would provide valuable insights into how to further improve approximation quality, that is, whether additional fitting points are necessary to reflect the underlying CFP model more accurately, whether more suitable functional forms and estimation assumptions are required for a more appropriate proxy modeling, or whether both aspects are relevant. Furthermore, one could deduce from such an analysis the sample sizes needed by the different regression algorithms to meet certain validation criteria. Since the generation of large sample sizes is currently computationally expensive for the industry, algorithms getting along with comparatively few fitting points should be striven for.

Picking a suitable calibration algorithm is most important from the viewpoint of capturing the CFP model and hence the SCR appropriately. Therefore, if the bias observed in the validation figures indicates indeed issues with the functional forms of our approaches, doing further research on techniques not entailing such a bias or at least a smaller one is vital. On the one hand, one can fine-tune the approaches of this thesis by trying different configurations thereof and/or bringing in randomness such as by the method of bootstrap aggregating, see e.g. Breiman (1994). On the other hand, one can analyze further machine learning alternatives such as the stochastic ones mentioned in the introduction, see e.g. Krah et al. (2020*b*). Ideally, various approaches like adaptive OLS regression, GLM, GAM and FGLS regression algorithms, artificial neural networks, tree-based methods and support vector machines would be fine-tuned and compared based on the same realistic and comprehensive data basis. Since a major challenge of machine learning calibration algorithms is hyperparameter selection, future research should be dedicated to efficient hyperparameter search algorithms and, as a means of mitigation thereof, stabilization methods such as ensemble methods. As a starting point for these kinds of investigations, going beyond the scope of this thesis, serves the aforementioned source.

Taking the definition of hyperparameters one step further, the regression approach itself (OLS, GLM, GAM, FGLS, artificial neural networks, etc.) could be identified with an additional hyperparameter which the hyperparameter search algorithm should select from. Thereby, constraints conditional on, for example, run time, approximation quality, complexity or certain hyperparameters could be imposed. For further reductions in run time, amongst others, the nature of the adaptive algorithm could be taken advantage of.

Part III

A Least-Squares Monte Carlo Approach in Valuing Life Insurance Contracts with Early Exercise Features

Résumé

Life insurance contracts with early exercise features can be priced by an algorithm using the least-squares Monte Carlo method. We consider equity-linked contracts with American/Bermudan-style surrender options and minimum interest rate guarantees payable upon contract termination. In the simulation framework, randomness and jumps in the movements of the interest rate, stochastic volatility, stock market and mortality are permitted. For the simultaneous valuation of numerous insurance contracts of which the initial values of the underlying stochastic processes vary, a hybrid probability measure and an additional regression function are introduced. An efficient seed-related simulation procedure accounting for the forward discretization bias is presented. Furthermore, a concept for the selection of consistent basis functions serving also as a validation concept is proposed. We apply our suggested procedures and concepts in an extensive numerical example.

16 Introduction

LSMC for Pricing Contracts

In this third and last part of the thesis, we describe a different application field of the least-squares Monte Carlo (LSMC) method in life insurance business. We take up the setting by Bacinello et al. (2009) to price life insurance contracts with early exercise features and extend it by a hybrid probability measure such as introduced by Bauer & Ha (2015)and Natolski & Werner (2016) and used in the foregoing parts of this thesis for the differentiation between the outer risk scenarios and inner Monte Carlo simulations. This new setting combines the LSMC idea of the original application in finance to price American or Bermudan options, see e.g. Longstaff & Schwartz (2001), with that under the Solvency II directive to switch between pricing and projection to obtain full loss distribution forecasts, see e.g. the first part of this thesis. Furthermore, we account for the discretization bias according to a comparatively efficient seed-related modification of the procedure proposed by Desmettre & Korn (2015) and present a suitable validation concept in the new setting.

More precisely, our objective is to extend the theoretical model of equity-linked endowment insurance contracts with surrender options from Bacinello et al. (2009) by a hybrid probability measure to allow the simultaneous valuation of numerous insurance contracts varying in the initial values of the stochastic processes of the modeled risk factors. These initial values take over the role of the outer scenarios in the hybrid probability framework (compare the other two parts of this thesis). Without increasing the computational effort, we achieve the simultaneous valuation by diversifying the simulation budget across Monte Carlo simulations with different initial values and introducing a new regression function with respect to these initial values at contract inception. We consider randomness and jumps in both the reference fund value to which the contract is linked and mortality. In addition, we allow randomness in the evolutions of the interest rates and stochastic volatility. Moreover, we choose the order of the discretization bias in coincidence with the order of the Monte Carlo error to avoid a misbalance between the number of Monte Carlo simulations and forward discretization step size. Last but not least, we present a validation concept for the selection of the basis functions under the hybrid probability measure and select the basis functions accordingly. We give detailed implementation instructions and illustrate our findings by numerical examples.

Outline

At first, we introduce the theoretical model based on Bacinello et al. (2009) and modify it where necessary to include the hybrid probability measure proposed by Bauer & Ha (2015) or Natolski & Werner (2016). For this purpose, we formalize the insurance contract in Section 17.1, describe the valuation framework in Section 17.2, model the financial and demographic risk factors in Section 17.3, and deal with the valuation of the contract in Section 17.4. We move on to the next section for the implementation of this model. Here, we start with some general remarks in Section 18.1 and discretize the underlying continuous time processes in Section 18.2. Then we define in Section 18.3 the required regression problems and suggest a concrete LSMC algorithm for valuing the various contracts simultaneously. Afterwards, we discuss in Section 18.4 the implications of the discretization bias and present in Section 18.5 a validation concept for the selection of the basis functions. We complete the last part of this thesis with a numerical example. In Section 19.1, we explain additional model specifications and address numerical obstacles. Thereafter, we describe in Section 19.2 how we achieve basis function consistency and balance out the Monte Carlo error and discretization bias in a preliminary phase. In Section 19.3, we discuss the results of our final runs. In Sections 17.5, 18.6 and 19.4, we give short summaries.

17 Theoretical Model

17.1 Contract

In the model setup of Bacinello et al. (2009), the equity-linked endowment contract with maturity $T > 0$ embeds a surrender option and minimum interest rate guarantees. The LSMC approach will tackle here the early exercise feature coming with the surrender option. The contract can be entered by an individual of age x at time 0, and the life insurance policy pays a lump sum benefit F_T^s at time T upon survival or a benefit F_τ^d at time $\tau \in (0, T]$ in case the individual passes away at τ. The American-style surrender option provides the policyholder with the option to withdraw from the contract at each time $\theta \in (0, T)$ and to obtain in return a surrender value F_θ^w. The equity-linkage is reflected in the contract through its dependence on a reference market fund value $S = (S_t)_{t \geq 0}$. The minimum guaranteed interest rate κ_e on the benefit payments varies with the manner in which the contract is terminated, that is, with survival, death and surrender, i.e., $e \in \{s, d, w\}$. Additionally, we assume that the policyholder finances the contract with a single initial premium F_0 equal to the initial reference fund value S_0. Contracts featuring all of the above properties can be represented by *terminal guarantees*. Their values F_t^e follow the expression

$$F_t^e = F_0 \max \left(\frac{S_t}{S_0}, \exp(\kappa_e t) \right), \ e \in \{s, d, w\}. \tag{103}$$

Contingent on the evolutions of the reference fund value and mortality, there might be times at which it is more attractive for a policyholder to surrender a contract against provision of the surrender value than to stay in the contract. In this respect, the time θ at which an individual decides to terminate a contract before maturity can be considered as an exercise *policy*. With the aid of the original LSMC approach developed for option pricing (see also the references given in Part I of this thesis), we will determine an optimal surrender policy θ^* for a rational policyholder. For this purpose, we define the cumulated benefits paid up to a fixed time t given a policy θ by

$$G_t(\theta) = F_T^s 1_{\tau > T, T \leq t \wedge \theta} + F_\tau^d 1_{\tau \leq t \wedge T \wedge \theta} + F_\theta^w 1_{\theta \leq t, \theta < \tau \wedge T}. \tag{104}$$

Here, the first term of the sum yields the payment at maturity if neither early withdrawal nor death happen until T. The second term provides the death benefit if the individual does not surrender the contract before the time of death and if death occurs until T. The third term yields the payment upon early withdrawal if the individual surrenders the contract before maturity and the time of death.

17.2 Valuation Framework

In contrast to Bacinello et al. (2009) who define a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{Q})$ with only a risk-neutral probability measure $\mathbb{Q}$ to capture the financial and demographic randomness, we specify a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with a physical probability measure $\mathbb{P}$ as suggested by Bauer & Ha (2015). A frictionless securities market is assumed so that the existence of a risk-neutral measure $\mathbb{Q}$ as the one defined by Bacinello et al. (2009) and equivalent to $\mathbb{P}$ can be guaranteed in suitable model settings in the absence of arbitrage. Under $\mathbb{Q}$ the price of any security is given by the expected value of its cumulated dividends discounted at the risk-free rate.

The filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ captures the information flow to the insurer and the policyholder. Thus, the σ-algebra $\mathcal{F}_t$ contains all information about the market that have been revealed up to time t. In particular, the time of death τ and the exercise policy θ are $\mathbb{F}$-stopping times, meaning that at any time t it is known whether death or surrender have occurred by t.

We model the time of death by

$$\tau = \inf \{t : \Gamma_t > \xi\}, \tag{105}$$

where $\Gamma = (\Gamma_t)_{t\geq 0}$ is a non-decreasing process with $\Gamma_t = \int_0^t \mu_s \mathrm{d}s$ for any time t and a nonnegative process $\mu = (\mu_t)_{t\geq 0}$ for the intensity of mortality, and where ξ is exponentially distributed with parameter one.

The hybrid probability measure consisting of the two equivalent measures $\mathbb{P}$ and $\mathbb{Q}$ is introduced to enable the valuation of numerous insurance contracts at comparatively low computational costs. Removal of the measure $\mathbb{P}$ would result in the setting of Bacinello et al. (2009) who value in fact a single contract at almost the same costs. The measure $\mathbb{P}$ effective at contract inception randomizes the initial values of the stochastic processes reflecting the financial and demographic randomness. Without $\mathbb{P}$, there is thus no diversification across these initial values so that no relationship between them and the contract value is deducible and no fast valuation of numerous insurance contracts possible. The measure $\mathbb{Q}$ living on $[0, T]$ represents the possible Monte Carlo paths. As our market is incomplete (due to the jumps in the various risk processes), the measure $\mathbb{Q}$ is not unique. However, we assume that one such $\mathbb{Q}$ is chosen for valuation purposes and do not consider the arguments that led to this particular risk-neutral measure. As opposed to the risk management applications presented in the foregoing two parts of this thesis, $\mathbb{P}$ is only effective at contract inception here and not upon a time interval. We denote the cube on which $\mathbb{P}$ is defined in analogy to the foregoing parts by S_{fit}. However, we use $\mathbb{P}$ instead of $\mathbb{P}'$ in this part as we do not need to distinguish between them here.

17.3 Risk Factors

Surrender decisions are driven in particular by financial and demographic risk factors. Accordingly, Bacinello et al. (2009) focus on interest rate risk, stock market performance and mortality risk. The financial risks are modeled by an extended Bates model. Bakshi et al. (1997) demonstrate that this model is well-suited for simulating the price behavior of equity derivatives.

For the term structure of interest rates, a standard Cox-Ingersoll-Ross (CIR) model is used, i.e., the short rate r_t is assumed to follow the dynamics given by

$$\mathrm{d}r_t = \zeta_r (\delta_r - r_t) \mathrm{d}t + \sigma_r \sqrt{r_t} \mathrm{d}Z_t^r, \tag{106}$$

where $\zeta_r, \delta_r, \sigma_r > 0$ and Z^r is a standard Brownian motion. This square-root process ensures nonnegative interest rates and is mean-reverting towards the long-run value δ_r with speed of adjustment ζ_r and volatility σ_r.

Similarly, the squared non-jump stochastic volatility of the stock value to which the insurance contract is linked is modeled by a standard CIR process, i.e.,

$$\mathrm{d}K_t = \zeta_K \left(\delta_K - K_t\right) \mathrm{d}t + \sigma_K \sqrt{K_t} \mathrm{d}Z_t^K, \tag{107}$$

with $\zeta_K, \delta_K, \sigma_K > 0$ and Z^K a standard Brownian motion. Nonnegativity and mean-reversion are guaranteed for this square-root process analogously to the interest rate dynamics in (106).

The evolution of the reference fund value $S = \exp(Y)$ follows the process

$$\mathrm{d}Y_t = \left(r_t - \frac{1}{2}K_t - \lambda_Y \mu_Y\right) \mathrm{d}t + \sqrt{K_t} \left(\rho_{Yr} \mathrm{d}Z_t^r + \rho_{YK} \mathrm{d}Z_t^K + \sqrt{1 - \rho_{Yr}^2 - \rho_{YK}^2} \mathrm{d}Z_t^Y\right) + \mathrm{d}J_t^Y, \tag{108}$$

where r, K, Z^r, Z^K stem from the processes (106), (107), Z^Y is a Brownian motion, ρ_{Yr}, ρ_{YK} are correlation coefficients satisfying $\rho_{Yr}^2 + \rho_{YK}^2 \leq 1$, and J^Y is a compound Poisson process independent of Z^r, Z^K, Z^Y with jump arrival rate $\lambda_Y > 0$, mean μ_Y and lognormally distributed jump sizes. More precisely, the jump diffusion process J^Y defined by Bakshi et al. (1997) evolves according to

$$\mathrm{d}J_t^Y = j_t^Y \mathrm{d}q_t^Y, \tag{109}$$

where q^Y is Poisson distributed with parameter λ_Y such that $\Pr\left\{\mathrm{d}q_t^Y = 1\right\} = \lambda_Y \mathrm{d}t$ and $\Pr\left\{\mathrm{d}q_t^Y = 0\right\} = 1 - \lambda_Y \mathrm{d}t$, interpretable as a jump counter, and j_t^Y is the percentage jump size in case $\mathrm{d}q_t^Y = 1$ where $1 + j_t^Y$ is lognormally distributed with mean $\log(1 + \mu_Y) - \frac{1}{2}\sigma_Y^2$ and variance σ_Y^2.

The intensity of mortality is modeled through the left continuous version of the process

$$\mathrm{d}\mu_t = \zeta_\mu \left(m(t) - \mu_t\right) \mathrm{d}t + \sigma_\mu \sqrt{\mu_t} \mathrm{d}Z_t^\mu + \mathrm{d}J_t^\mu, \tag{110}$$

with $\zeta_\mu, m(\cdot), \sigma_\mu > 0$, Z^μ a Brownian motion, and J^μ a compound Poisson process independent of Z^μ with jump arrival rate $\lambda_\mu \geq 0$ and exponentially distributed jump sizes. Similarly to the dynamics in (109), the jump diffusion process J^μ can be written as

$$\mathrm{d}J_t^\mu = j_t^\mu \mathrm{d}q_t^\mu, \tag{111}$$

with q^μ Poisson distributed with parameter λ_μ so that we have here $\Pr\{\mathrm{d}q_t^\mu = 1\} = \lambda_\mu \mathrm{d}t$ and $\Pr\{\mathrm{d}q_t^\mu = 0\} = 1 - \lambda_Y \mathrm{d}t$, and j_t^μ the jump size in case $\mathrm{d}q_t^\mu = 1$ where j_t^μ is exponentially distributed with mean γ_μ.

The state variable process $X = (X_t)_{t\geq 0} = ((r_t, K_t, S_t, \mu_t))_{t\geq 0} = (r, K, S, \mu)$ generates the filtration $\mathbb{F}$ introduced above. The measure $\mathbb{P}$ randomizes the initial values of the risk factors $X_0 = (r_0, K_0, S_0, \mu_0)$ and captures thus possible stresses. Conditional on X_0, the risk factors evolve with respect to $\mathbb{Q}$. We refer to the $\mathbb{P}$-randomized initial values as the *outer scenarios* and to the $\mathbb{Q}$-randomized risk factor evolutions as the *inner scenarios* in analogy to the foregoing two parts of this thesis. If numerous complex insurance contracts with embedded options varying in the outer scenarios were valued by a *nested simulations approach* as described in Bauer et al. (2012), conditional on each outer scenario X_0 a large number of Monte Carlo simulations based on various inner scenarios $(X_t)_{t\geq 0}$ would be required. By budgeting only one or few inner simulations to each outer scenario, our extended LSMC approach is able to produce considerably less computational costs. Here, a $\mathbb{Q}$-based fair valuation of a given insurance contract is finally obtained by plugging the corresponding outer scenario in the last regression function from the backward induction procedure, that is, the one at contract inception. For this central idea, compare also the second section of the first part of this thesis in which the *nested valuation problem* and LSMC solution are discussed.

17.4 Valuation

The financial market consists of the investment stock S with evolution (108) and a money market account yielding the instantaneous risk-free rate $(r_t)_{t\geq 0}$. Based on the latter, the accumulation rate can be formalized as $B_t = \exp\left(\int_0^t r_s \mathrm{d}s\right)$ for $t \geq 0$. In the absence of arbitrage, under the given choice of the risk-neutral measure $\mathbb{Q}$ and the assumption that the surrender decision is based on the entire information available over time, the time-t value of the insurance contract for a fixed exercise policy θ is given by the risk-neutral formula

$$V_t(\theta) = B_t E^{\mathbb{Q}}\left[\int_t^{\infty} B_u^{-1} \mathrm{d}G_u(\theta) \mid \mathcal{F}_t\right], \tag{112}$$

where $G_u(\theta)$ is defined in (104). This time-t value reflects the cumulated benefits paid by the contract at all future times $u \geq t$.

The price of the insurance contract specified in (103) is given by inserting the solution θ^* to the optimal stopping problem

$$V_0^* = V_0(\theta^*) = \sup_{\theta\in\mathcal{T}_{\mathbb{F}}} V_0(\theta) = \sup_{\theta\in\mathcal{T}_{\mathbb{F}},\theta\leq\tau} V_0(\theta), \tag{113}$$

with τ denoting the time of death characterized in (105) and $\mathcal{T}_{\mathbb{F}}$ denoting the set of finite valued $\mathbb{F}$-stopping times. Since the solution θ^* maximizes the initial arbitrage-free insurance contract value, it is called *rational* exercise policy. The expectation with respect to $\mathbb{Q}$ in (112) depends on the outer scenario associated with the insurance contract to be valued. The outer scenario enters the σ-algebra $\mathcal{F}_t$ and defines as the only information available at contract inception the initial σ-algebra $\mathcal{F}_0$.

17.5 Summary

The model includes an equity-linked endowment insurance contract with a surrender option and minimum interest rate guarantees depending upon contract termination, see Section 17.1. By defining a filtered probability space with a hybrid probability measure the valuation framework has been established in Section 17.2. Thereafter, the state variable process has been composed of the movements of the interest rate, stochastic volatility, stock market and mortality in Section 17.3. The interest rate and stochastic volatility risk factors are modeled by standard CIR processes and the reference fund value and mortality risk factors are allowed to make jumps. Moreover, inner and outer scenarios have been introduced and the nested valuation problem addressed. In Section 17.4, the time-t value of the insurance contract has been written with respect to the exercise policy and the insurance contract value at inception defined as the time-0 value evaluated at the solution to the optimal stopping problem.

18 Least-Squares Monte Carlo Approach

18.1 General Remarks

The implementation of the theoretical model described above requires the following steps. The continuous time processes have to be discretized in the time dimension: Besides the forward discretization, a backward discretization is needed to replace the continuous optimal stopping problem by a discrete one. Typically the backward discretization is carried out on a coarser time grid. Furthermore, a regression method with a set of suitable basis functions has to be chosen for the LSMC algorithm.

As an enhancement, we propose to select the order of the forward discretization bias in coincidence with the order of the Monte Carlo error and to rely on consistent basis functions. While the former helps to determine a suitable forward discretization step size, the latter ensures consistent results with Bacinello et al. (2009) and can be considered a validation procedure. However, an analysis of the discretization bias and a consistent basis function selection require several complete runs of the LSMC algorithm. The final regression function at contract inception is found when the discretization bias is accounted for and the results show a sufficiently high basis function consistency.

The implementation algorithm which we describe in the following works in particular for the parameter choices made in the numerical experiment by Bacinello et al. (2009).

18.2 Simulation Setting

18.2.1 Forward Discretization

For simulation, the state variable process $X = (r, K, S, \mu)$ consisting of the financial and demographic risk factors defined by (106), (107), (108) and (110) has to be discretized in time. As this discretization takes place from contract inception until maturity, we call it *forward discretization*. We use an adapted version of the natural Euler-Maruyama method to simulate process (108) according to Korn et al. (2010)[p. 320]. Given the initial value $Y_0 = \exp(S_0)$, we obtain for each $t = t_l = \frac{l}{L}T,\ l = 1, \ldots, L$, with L standing for the number of forward discretization steps so that one time step is of length $\delta t = \frac{T}{L}$,

$$
\begin{aligned}
Y_{t+1} = Y_t \quad &+ \quad \left(r_t - \frac{1}{2}K_t - \lambda_Y \mu_Y\right)\delta t \\
&+ \quad \sqrt{K_t}\left(\rho_{Yr}\delta Z_t^r + \rho_{YK}\delta Z_t^K + \sqrt{1-\rho_{Yr}^2-\rho_{YK}^2}\,\delta Z_t^Y\right) + \delta J_t^Y. \quad (114)
\end{aligned}
$$

Exemplary Monte Carlo paths of stochastic process (114) are depicted in Figure 27.

To simulate the remaining processes (106), (107) and (110), we follow the proposal of Alfonsi (2005) and use the so-called explicit scheme $E(0)$ ensuring nonnegative values as long as $\sigma_r^2 \le 4\zeta_r\delta_r$ resp. $\sigma_K^2 \le 4\zeta_K\delta_K$ resp. $\sigma_\mu^2 \le 4\zeta_\mu m(t)$ for each $t \ge 0$. These nonnegativity conditions can be immediately seen in the definitions below since the quadratic and compound Poisson process parts are always nonnegative. Given the initial values

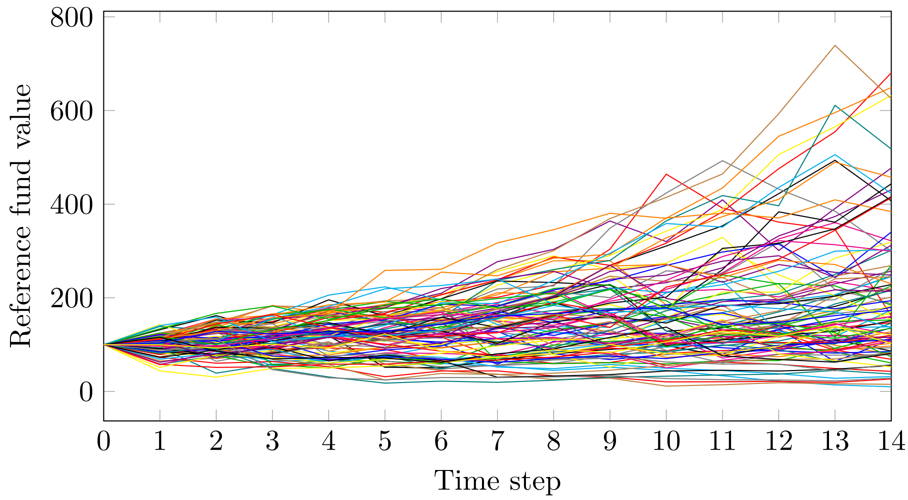

Figure 27: Set of stochastically simulated reference fund values on the backward discretization time grid conditional on $S_0 = 100$ and forward discretization step size $\delta t = 0.01$ used for regression.

$r_0, K_0, \mu_0 = m(0)$, we have for each $t = t_l = \frac{l}{L}T,\ l = 1, \ldots, L$,

$$r_{t+1} = \left(\left(1 - \frac{\zeta_r}{2}\delta t\right)\sqrt{r_t} + \frac{\sigma_r}{2\left(1 - \frac{\zeta_r}{2}\delta t\right)}\delta Z_t^r\right)^2 + \left(\zeta_r \delta_r - \frac{\sigma_r^2}{4}\right)\delta t, \tag{115}$$

$$K_{t+1} = \left(\left(1 - \frac{\zeta_K}{2}\delta t\right)\sqrt{K_t} + \frac{\sigma_K}{2\left(1 - \frac{\zeta_K}{2}\delta t\right)}\delta Z_t^K\right)^2 + \left(\zeta_K \delta_K - \frac{\sigma_K^2}{4}\right)\delta t, \tag{116}$$

$$\mu_{t+1} = \left(\left(1 - \frac{\zeta_\mu}{2}\delta t\right)\sqrt{\mu_t} + \frac{\sigma_\mu}{2\left(1 - \frac{\zeta_\mu \delta t}{2}\right)}\delta Z_t^\mu\right)^2 + \left(\zeta_\mu m(t) - \frac{\sigma_\mu^2}{4}\right)\delta t + \delta J_t^\mu \tag{117}$$

Exemplary Monte Carlo paths of stochastic process (115) are depicted in Figure 28.

Thereby, we draw $\delta Z_t^r, \delta Z_t^K, \delta Z_t^Y$ and δZ_t^μ from the normal distribution with zero mean and variance δt for each time step $t_l = \frac{l}{L}T,\ l = 1, \ldots, L$. To obtain δJ_t^Y and δJ_t^μ, we simulate the compound Poisson process by the alternative jump time simulation in Korn et al. (2010)[p. 312]. While we use the jump intensity λ_Y and lognormal height distribution specified for $1 + j_t^Y$ below Equation (109) to simulate J_t^Y, we use the jump intensity λ_μ and exponential height distribution specified for j_t^μ below Equation (111) to simulate J_t^μ.

The alternative jump time simulation works for $b \in \{Y, \mu\}$ as follows:

1. Draw the number of jumps $N(T)$ from the Poisson distribution with parameter $\lambda_b T$.
2. Draw $N(T)$ independent random variables u_j from the uniform distribution $U[0, T]$.
3. Draw $N(T)$ independent random variables h_j from the specified height distribution.
4. Assign the $u_j, j = 1, \ldots, N(T)$, to times $t_l = \frac{l}{L}T,\ l = 1, \ldots, L$, on the forward time grid. We do the assignment by setting $l_j = \left\lfloor u_j \frac{L}{T} \right\rfloor$ and thus $t_{l_j} = \left\lfloor u_j \frac{L}{T} \right\rfloor \frac{T}{L}$ where $\lfloor \cdot \rfloor$ stands for the floor function.

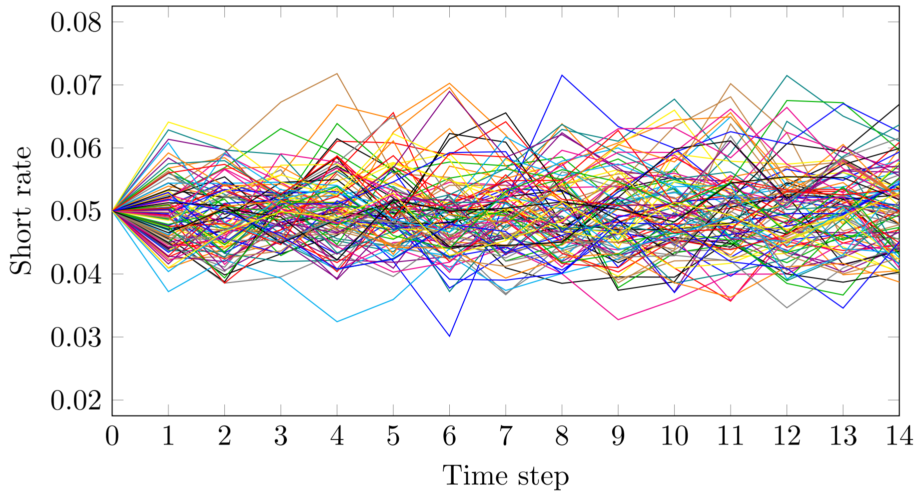

Figure 28: Set of stochastically simulated short rates on the backward discretization time grid conditional on $r_0 = 0.05$ and forward discretization step size $\delta t = 0.01$ used for regression.

5. Set $\delta J^b_{t_{l_j}} = h_j,\ j = 1, \ldots, N(T)$.
6. Set $\delta J^b_{t_l} = 0$ if $l \neq l_j,\ j = 1, \ldots, N(T),\ l = 1, \ldots, L$.

Since the event $t_{l_j} = t_{l_i},\ j \neq i$, is very unlikely with our parameter choices, we simply neglect it and overwrite $\delta J^b_{t_{l_j}}$ by h_i or $\delta J^b_{t_{l_i}}$ by h_j where applicable. The compound Poisson process is then given by $J^b_{t_k} = \sum_{l=1}^{k} \delta J^b_{t_l},\ k = 1, \ldots, L$.

In our extended setting, the simulated processes X do not only differ because of the randomness in the Brownian motions and compound Poisson processes but also because of their different initial conditions reflected by the outer scenarios $X_0 = (r_0, K_0, S_0, \mu_0)$.

18.2.2 Backward Discretization

After the Monte Carlo simulations of (114), (115), (116) and (117) have been carried out according to the forward discretization described above, the *backward discretization* in time is required to solve the optimal stopping problem (113) by backward induction. From an economic perspective, the American surrender option of the insurance contract is replaced by a Bermudan surrender option by this discretization. To achieve consistent results, the backward discretization time grid is selected as a subgrid of the forward discretization time grid. Usually, the backward discretization grid is chosen to be coarser as it determines the number of computationally expensive LSMC regression steps. Let M, a divisor of L, be the number of backward discretization steps. Then we obtain the backward time grid $\mathbb{T} = \left\{ t_m = \frac{m}{M} T \mid m = M, \ldots, 0 \right\}$, and the discretized version of the continuous optimal stopping problem becomes

$$\sup_{\theta \in \mathcal{T}_{\mathbb{F},\mathbb{T}}, \theta \leq \tau} V_0(\theta), \tag{118}$$

where $\mathcal{T}_{\mathbb{F},\mathbb{T}}$ denotes the family of $\mathbb{T}$-valued $\mathbb{F}$-stopping times. Since $B_0 = 1$ and $G_t(\theta) = \text{const}$ for $t \geq \theta$, by using (112) expression (118) reduces to

$$\sup_{\theta \in \mathcal{T}_{\mathbb{F},\mathbb{T}}, \theta \leq \tau} V_0(\theta) = \sup_{\theta \in \mathcal{T}_{\mathbb{F},\mathbb{T}}, \theta \leq \tau} E^{\mathbb{Q}} \left[\int_0^\theta B_u^{-1} \mathrm{d}G_u(\theta) \mid \mathcal{F}_0 \right]. \tag{119}$$

Defining the function which is evaluated at $t = \theta$ and of which the expectation is taken in (119) by

$$g_t = \int_0^t B_u^{-1} \mathrm{d}G_u(\theta) \tag{120}$$

provides the process $(g_t)_{t \geq 0}$. Introducing the Snell envelope (see e.g. Föllmer & Schied (2004)[p. 280–282]) of this process allows the realization of a dynamic programming principle involving at each time step $t_m = \frac{m}{M}T,\ m = M-1, \ldots, 1$, a comparison between the payoff from surrendering the insurance contract and the *continuation value*. By the continuation value we refer to the expected payoff from not exercising the surrender option. If the policyholder survives maturity, an approximate solution $\theta^* = \theta_0^*$ to the optimal stopping problem (118) is computed in the LSMC algorithm through the backward procedure

$$\begin{cases} \theta_M^* = t_M, \\ \theta_m^* = t_m 1_{g_{t_m} > U_m} + \theta_{m+1}^* 1_{g_{t_m} \leq U_m} \text{ for } m = M-1, \ldots, 1, \\ \theta_0^* = \theta_1^*, \end{cases} \tag{121}$$

with the immediate payoff g_{t_m} and continuation value $U_m = E^{\mathbb{Q}}\left[g_{\theta_{m+1}^*} \mid \mathcal{F}_{t_m}\right]$ evaluated at time t_m.

At contract inception, the policyholder certainly does not surrender the contract as she would otherwise lose $V_0^* - F_0 > 0$ with $V_0^* > F_0$. From the insurer's perspective, such a contract offer is expected to be profitable if its asset management is able to earn more than $V_0^* - F_0$ per premium (costs neglected). Since we make a distinction between surrender and survival benefits, the policyholder is formally not permitted to withdraw from the contract at maturity.

18.3 Least-Squares Regression

18.3.1 Basis Functions

In this section, we estimate the continuation values from above for $m = M-1, \ldots, 1$ and value the contract at inception conditional on any outer scenario.

Under the assumption of a Markovian environment, we can write the continuation values U_m from above as $U_m = E^{\mathbb{Q}}\left[g_{\theta_{m+1}^*} \mid X_{t_m}\right] = u\left(t_m \mid X_{t_m}\right)$ for some Borel functions $u\left(t_m \mid \cdot\right)$, $m = M-1, \ldots, 1$. The first approximation is made when replacing each $u\left(t_m \mid X_{t_m}\right)$ with a projection from $L^2(\Omega)$ onto the H-dimensional vector space generated by a suitable set of basis functions $\{e_1, \ldots, e_H, \ldots\}$.

For fixed H and each m, let $\boldsymbol{e}(\cdot) = (e_1(\cdot), \ldots, e_H(\cdot))^{\mathrm{T}}$ be the basis function vector and let $\boldsymbol{\beta}_{\boldsymbol{m}}^* = \left(\beta_{m,1}^*, \ldots, \beta_{m,H}^*\right)^{\mathrm{T}}$ be the ordinary least-squares estimator given as the solution to regression problem

$$\boldsymbol{\beta}_{\boldsymbol{m}}^* = \arg \min_{\boldsymbol{\beta}_{\boldsymbol{m}} \in \mathbb{R}^H} \sum_{i=1}^{N_m} \left(w^i(t_m) - \boldsymbol{\beta}_{\boldsymbol{m}}^{\mathrm{T}} \boldsymbol{e}\left(X_{t_m}^i\right)\right)^2, \tag{122}$$

where N_m is the number of simulations in which the insured is alive at time t_m, where $X_{t_m}^i$ is the simulated value of X_{t_m} and $w^i(t_m) = B_{t_m}^i \left(B_{\theta_{m+1}^*}^i\right)^{-1} P^i\left(\theta_{m+1}^*\right)$ is the *simulated* time-t_m continuation value in the i-th simulation. These time-t_m continuation values are the results of the regressions and comparisons from the previous backward induction steps $M-1,\ldots,m+1$. At time t_{M-1}, the simulated continuation value is simply the discounted survival benefit given the policyholder survives maturity, otherwise the simulation is excluded from the time-t_{M-1} regression.

The time-t_m continuation value of the i-th simulation *estimated* based on (122) is given by the second approximation

$$\widetilde{u}^i\left(t_m \mid X^i_{t_m}\right) = \boldsymbol{\beta}^{*,\mathrm{T}}_{\boldsymbol{m}} \boldsymbol{e}\left(X^i_{t_m}\right). \tag{123}$$

Since the policyholder does not know the future development, she has to estimate the continuation value based on what she knows at time t_m. If the surrender benefit at t_m is greater than the estimated time-t_m continuation value $\widetilde{u}^i\left(t_m \mid X^i_{t_m}\right)$, the *simulated* continuation value $w^i\left(t_{m-1}\right)$ in the next backward induction step will be this surrender benefit after discounting.

Besides the continuation values, we estimate the contract value conditional on the outer scenarios by regression. Technically, the contract value is a continuation value, too, and can thus be written as $V^*_0 = E^{\mathbb{Q}}\left[g_{\theta^*_0} \mid \mathcal{F}_0\right] = u\left(X_0\right)$ with $u\left(X_0\right)$ being a Borel function and X_0 representing an outer scenario. Now we replace $u\left(X_0\right)$ with an approximation from $L^2\left(\Omega\right)$ onto the H_0-dimensional vector space generated by a potentially different set of basis functions $\left\{e^0_1,\ldots,e^0_{H_0},\ldots\right\}$.

Let H_0 be fixed, let $\boldsymbol{e^0}\left(\cdot\right) = \left(e^0_1\left(\cdot\right),\ldots,e^0_{H_0}\left(\cdot\right)\right)^{\mathrm{T}}$ be the basis function vector and let $\boldsymbol{\beta^*_0} = \left(\beta^*_{0,1},\ldots,\beta^*_{0,H_0}\right)^{\mathrm{T}}$ be the ordinary least-squares estimator solving regression problem

$$\boldsymbol{\beta^*_0} = \arg\min_{\boldsymbol{\beta_0}\in\mathbb{R}^{H_0}} \sum_{i=1}^{N}\left(w^i\left(0\right) - \boldsymbol{\beta^{\mathrm{T}}_0}\boldsymbol{e^0}\left(X^i_0\right)\right)^2, \tag{124}$$

where N is the sample size, X^i_0 is the outer scenario and $w^i\left(0\right) = \left(B^i_{\theta^*_0}\right)^{-1} P^i\left(\theta^*_0\right)$ is the *simulated* continuation value in the i-th simulation. Similarly to the continuation values above, the insurance contract value conditional on outer scenario X_0 can be estimated by

$$\widetilde{V}^*_0\left(X_0\right) = \boldsymbol{\beta^{*,\mathrm{T}}_0}\boldsymbol{e^0}\left(X_0\right). \tag{125}$$

18.3.2 LSMC Algorithm

Suppose that N paths of the state variable process $X = (r, K, S, \mu)$ have been simulated according to (114), (115), (116) and (117) and that N exponential random variables ξ with parameter one have been simulated.

Let $\left(\mu^i_{t_l}\right)_{0\leq l\leq L}$ be the i-th evolution of the intensity of mortality (110) and let ξ^i be the i-th simulated exponential random variable. Then the i-th simulated time of death τ^i is obtained by definition (105) as

$$\tau^i = \min\left\{t_l : \Gamma^i_{t_l} > \xi^i\right\}, \tag{126}$$

where $\Gamma^i_{t_l} = \sum_{s=1}^{l}\mu^i_{t_s}\delta t$ is the simulated value of $\Gamma_{t_l} = \int_0^t \mu_s \mathrm{d}s$. By convention, we set $\tau^i = \infty$ if the policyholder survives maturity T.

Moreover, let $\left(r^i_{t_l}\right)_{0\leq l\leq L}$, $\left(K^i_{t_l}\right)_{0\leq l\leq L}$ and $\left(S^i_{t_l}\right)_{0\leq l\leq L}$ be the i-th evolutions of the term structure of interest rates (106), stochastic volatility (107) and reference fund value as the exponential of (108), respectively. We compute the simulated discount factors by $v^i_{t_l,t_k} = B^i_{t_l}\left(B^i_{t_k}\right)^{-1} = \exp\left(-\sum_{s=l+1}^{k} r^i_{t_s}\delta t\right)$ with $t_l < t_k$.

By following Equation (103), we calculate the simulated benefits paid upon death $F^{d,i}_{\tau^i} = F_0 \max\left(\frac{S^i_{\tau^i}}{S_0}, \exp\left(\kappa_d \tau^i\right)\right)$ if $\tau^i \leq T$, upon survival $F^{s,i}_T = F_0 \max\left(\frac{S^i_T}{S_0}, \exp\left(\kappa_s T\right)\right)$ otherwise, and upon surrender $F^{w,i}_{t_m} = F_0 \max\left(\frac{S^i_{t_m}}{S_0}, \exp\left(\kappa_w t_m\right)\right)$ for $m = M-1, \ldots, 1$ in case $t_m < \tau^i$.

The insurance contract value $\widetilde{V}^*_0$ can now be derived by the LSMC algorithm depicted in Figure 29.

18.4 Forward Discretization Bias

18.4.1 Decomposition of MSE

Technically, we are now able to run the LSMC algorithm. However, some questions like how to choose the forward discretization step size or which basis functions to use are still unanswered. Answering these questions goes beyond the scope of Bacinello et al. (2009) and Bauer & Ha (2015) and will be the focus of the following. To find a reasonable forward discretization step size conditional on the number of Monte Carlo simulations, we apply the procedure proposed by Desmettre & Korn (2015).

We aim to select the order of the forward discretization bias in coincidence with the one of the Monte Carlo error to avoid a misbalance between the number of Monte Carlo simulations N and the forward discretization step size δt. If these two parameters are not balanced out, one of the two error sources dominates the other one. Both kinds of dominations have undesirable effects: If the Monte Carlo error dominates the discretization bias, the forward discretization step size can be increased without losing accuracy, and if it is the other way around, the results are biased due to the discretization of the stochastic processes. Hence we would end up either with a suboptimal computational power usage or biased estimation results.

The decomposition of the mean squared error (MSE) into the Monte Carlo error and discretization bias will now be presented. For reasons of simplification, we assume for this analysis that there is no diversification across the outer scenarios, i.e., $X^i_0 = X_0,\ i = 1, \ldots, N$, so that the physical probability measure $\mathbb{P}$ vanishes and the setting of Bacinello et al. (2009) is established. As a result, the regression (124) at contract inception reduces to simply averaging over the continuation values $w^i(0),\ i = 1, \ldots, N$. Moreover, let a backward discretization time grid $\mathbb{T}$ based on which the backward induction steps are performed be fixed. In addition, let the *true* insurance contract value $\widetilde{V}^*_0(N, X)$ be generated by the algorithm in Figure 29 conditional on N simulations, all together indicated by X, with an infinitesimally small forward discretization step size $\delta t \to 0$. Similarly, let $\widetilde{V}^*_0\left(N, X^{\delta t}\right)$ be generated by the algorithm in Figure 29 based on N simulations, all together indicated by $X^{\delta t}$, with a forward discretization step size of $\delta t > 0$. Then the expected mean squared error (MSE) made when calculating $\widetilde{V}^*_0\left(N, X^{\delta t}\right)$ to approximate $\widetilde{V}^*_0(N, X)$ is

$$\begin{aligned}
\text{MSE} &= E\left[\widetilde{V}^*_0\left(N, X^{\delta t}\right) - E\left[\widetilde{V}^*_0(N, X)\right]\right]^2 \\
&= E\left[\widetilde{V}^*_0\left(N, X^{\delta t}\right) - E\left[\widetilde{V}^*_0\left(N, X^{\delta t}\right)\right] + E\left[\widetilde{V}^*_0\left(N, X^{\delta t}\right)\right] - E\left[\widetilde{V}^*_0(N, X)\right]\right]^2
\end{aligned}$$

Step 1: *(Initialization)*

Set $I = \{1 \leq i \leq N\}$.

for all $i \in I$ **do**

  if $\tau^i \leq T$ **then**

    $\theta^{*,i} = \tau^i$ and $P^i_{\theta^{*,i}} = F^{d,i}_{\theta^{*,i}}$

  else

    $\theta^{*,i} = T$ and $P^i_{\theta^{*,i}} = F^{s,i}_{\theta^{*,i}}$

Step 2: *(Backward induction)*

for $m = M - 1$ to 0 **do**

  Set $I_m = \left\{1 \leq i \leq N : \tau^i > t_m\right\}$.

  (1) *(Simulated continuation values)*

  for all $i \in I_m$ **do**

    $w^i_{t_m} = v^i_{t_m,\theta^{*,i}} P^i_{\theta^{*,i}}$

  (2) *(Estimated continuation values)*

  if $m \neq 0$ **then**

    Regress $\left(w^i_{t_m}\right)_{i \in I_m}$ against $\left(\boldsymbol{e}\left(X^i_{t_m}\right)\right)_{i \in I_m}$.

    for all $i \in I_m$ **do**

      $\widetilde{u}^i_{t_m} = \boldsymbol{\beta}^{*,\mathrm{T}}_{\boldsymbol{m}} \boldsymbol{e}\left(X^i_{t_m}\right)$

      if $F^{w,i}_{t_m} > \widetilde{u}^i_{t_m}$ **then**

        $\theta^{*,i} = t_m$ and $P^i_{\theta^{*,i}} = F^{w,i}_{\theta^{*,i}}$

  else

    Regress $\left(w^i_0\right)_{i \in I_0}$ against $\left(\boldsymbol{e}^{\mathbf{0}}\left(X^i_0\right)\right)_{i \in I_0}$.

Step 3: *(Insurance contract value)*

Compute the contract value for X_0:

$\widetilde{V}^*_0 = \boldsymbol{\beta}^{*,\mathrm{T}}_{\mathbf{0}} \boldsymbol{e}^{\mathbf{0}}\left(X_0\right)$

Figure 29: LSMC approach for valuing life insurance contracts.

$$\begin{aligned}
&= E\left[\widetilde{V}_0^*\left(N, X^{\delta t}\right) - E\left[\widetilde{V}_0^*\left(N, X^{\delta t}\right)\right]\right]^2 + \left(E\left[\widetilde{V}_0^*\left(N, X^{\delta t}\right)\right] - E\left[\widetilde{V}_0^*\left(N, X\right)\right]\right)^2 \\
&\quad + 2\,\underbrace{E\left[\widetilde{V}_0^*\left(N, X^{\delta t}\right) - E\left[\widetilde{V}_0^*\left(N, X^{\delta t}\right)\right]\right]}_{=\,0}\left(E\left[\widetilde{V}_0^*\left(N, X^{\delta t}\right)\right] - E\left[\widetilde{V}_0^*\left(N, X\right)\right]\right) \\
&= \underbrace{\mathrm{Var}\left[\widetilde{V}_0^*\left(N, X^{\delta t}\right)\right]}_{\text{Monte Carlo error}} + \underbrace{\left(E\left[\widetilde{V}_0^*\left(N, X^{\delta t}\right)\right] - E\left[\widetilde{V}_0^*\left(N, X\right)\right]\right)^2}_{\text{Discretization bias}}. \qquad (127)
\end{aligned}$$

18.4.2 Harmonization Algorithm

A first algorithm close to that by Desmettre & Korn (2015) with the purpose of selecting the order of the discretization bias in coincidence with the one of the Monte Carlo error consists of the following three steps:

1. Start with a rather large forward discretization step size δt and then increase the number of Monte Carlo simulations N until you reach the desired accuracy of the insurance contract value at confidence level $1-\alpha$, i.e.,
$$\left[\widetilde{V}_0^*\left(N, X^{\delta t}\right) - z_{1-\frac{\alpha}{2}}\frac{\widehat{\sigma}_N}{\sqrt{N}}, \widetilde{V}_0^*\left(N, X^{\delta t}\right) + z_{1-\frac{\alpha}{2}}\frac{\widehat{\sigma}_N}{\sqrt{N}}\right], \qquad (128)$$
where $z_{1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution and $\widehat{\sigma}_N$ in dependence of $w^i\left(0\right),\ i = 1, \ldots, N$, is given by
$$\widehat{\sigma}_N^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(w^i\left(0\right) - \widetilde{V}_0^*\left(N, X^{\delta t}\right)\right)^2. \qquad (129)$$
This variance is an estimator of the Monte Carlo error $\mathrm{Var}\left[\widetilde{V}_0^*\left(N, X^{\delta t}\right)\right]$ given in Equation (127). The order of the Monte Carlo error is generally equal to $O\left(\frac{1}{N}\right)$. Our objective is to find a forward discretization step size δt at which the order of the discretization bias $O\left(\epsilon^2\right)$ is equal to $O\left(\frac{1}{N}\right)$.

2. Decrease the forward discretization step size δt while keeping the final number of Monte Carlo simulations N from the first step fixed. To compensate random fluctuations, repeat the calculation process with the same δt several times and take the average over the resulting insurance contract values as the estimate. Given the forward discretization step size δt is large so that the discretization bias dominates the Monte Carlo error at the start of this step, the estimated insurance contract value should change as δt decreases. Repeat this step until the estimate of the insurance contract value stabilizes.

3. The stabilization indicates the completion of the search for δt at which the order of the discretization bias $O\left(\epsilon^2\right)$ is equal to $O\left(\frac{1}{N}\right)$. Since the discretization bias ϵ^2 can be written in terms of δt, that is $\epsilon^2 = \epsilon^2\left(\delta t\right)$, we can derive a relationship between the order of the discretization bias and the order of the Monte Carlo error, namely, $\epsilon^2\left(\delta t\right) = \frac{1}{N}$. After rearranging the terms, the forward discretization step size δt can be expressed conditional on the number of Monte Carlo simulations N, i.e.,
$$\delta t = \delta t\left(N\right). \qquad (130)$$
Hence, δt and N should be selected such that they satisfy this relationship whenever the Monte Carlo error and discretization bias are supposed to be harmonized.

18.4.3 Same Seed

Unlike Desmettre & Korn (2015), we do not run the LSMC algorithm several times with the same forward discretization step sizes in the second step to compensate random fluctuations by averaging. Instead we generate the Brownian motions Z^r, Z^K, Z^Y, Z^μ and compound Poisson processes J^Y, J^μ in (114), (115), (116) and (117) only once with our smallest forward discretization step size δt_{min} and derive the processes for larger forward discretization step sizes δt based on the realizations for δt_{min}. In this way, we use the same seed throughout the entire harmonization procedure. Figure 30 contains exemplary Monte Carlo paths of the intensity of mortality for different forward discretization step sizes that have all been derived based on the realizations of the same simulation with $\delta_{\text{min}} = 0.001$.

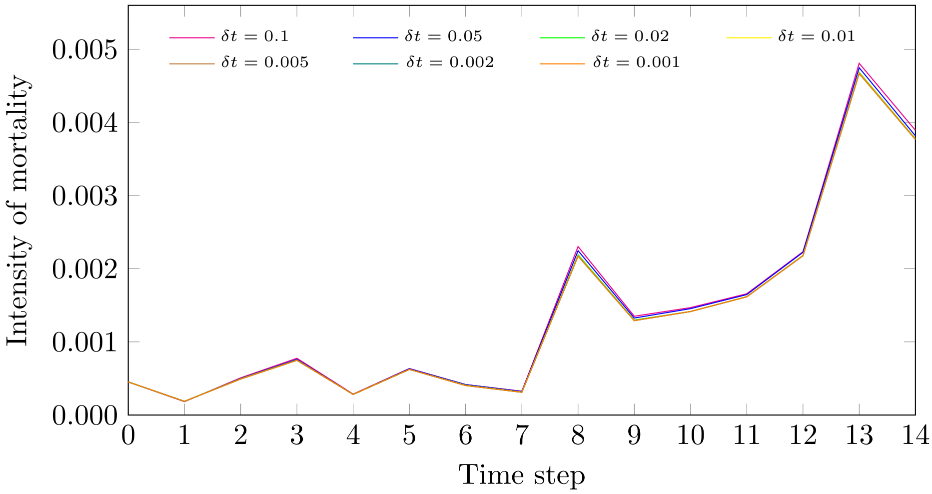

Figure 30: Stochastically simulated intensity of mortality on the backward discretization time grid conditional on $\mu_0 = m(0)$ for varying forward discretization step size δt.

The justification why such an approach is mathematically valid can be divided into two parts and goes as follows. In the first part, we define constructions for all suitable larger forward discretization step sizes based on the Brownian motions and compound Poisson processes that we have generated based on δt_{min}. In the second part, we show that these constructions have indeed the properties of Brownian motions and compound Poisson processes. We start with the proof for the Brownian motions and complete with the proof for the compound Poisson processes.

18.4.4 Brownian Motions

Proof. Let $Z^{b,\delta t_{\text{min}}}$, $b \in \{r, K, Y, \mu\}$, be standard Brownian motions generated with the smallest forward discretization step size δt_{min} and let $L_{\text{min}} = \frac{T}{\delta t_{\text{min}}}$. Then all components $\delta Z^{b,\delta t_{\text{min}}}_{t_{l_{\text{min}}}}$, $l_{\text{min}} = 1, \ldots, L_{\text{min}}$, are by definition independent and normally distributed with zero mean and variance δt_{min}, i.e.,

$$\delta Z^{b,\delta t_{\text{min}}}_{t_{l_{\text{min}}}} \sim N\left(0, \delta t_{\text{min}}\right). \tag{131}$$

Let $\delta t = a\delta t_{\min},\ a \in \mathbb{N}$, be a larger forward discretization step size where in addition a is a divisor of $L_{\min}$. We construct the Brownian motions for δt based on those of $\delta t_{\min}$ by

$$\delta Z_{t_{al}}^{b,\delta t} = \sum_{l_{\min}=a(l-1)+1}^{al} \delta Z_{t_{l_{\min}}}^{b,\delta t_{\min}},\ l = 1,\ldots,L, \tag{132}$$

with $L = \frac{L_{\min}}{a}$. Due to (131), all $al - (a\,(l-1)+1)+1 = a$ summands on the right-hand side of Equation (132) are independent and identically distributed. By applying the results of Eisenberg & Sullivan (2008), who prove once more that the sum of two independent and normally distributed random variables is normally distributed, and by induction, it follows that

$$\delta Z_{t_{al}}^{b,\delta t} \sim N\left(0, a\delta t_{\min}\right). \tag{133}$$

By using $\delta t = a\delta t_{\min}$, we obtain

$$\delta Z_{t_{al}}^{b,\delta t} \sim N\left(0, \delta t\right) \tag{134}$$

which we wanted to show to prove that the constructions in (132) are indeed standard Brownian motions for the larger forward discretization step size δt. □

18.4.5 Compound Poisson Processes

Proof. Similarly, let $J^{b,\delta t_{\min}},\ b \in \{Y,\mu\}$, be compound Poisson processes generated with the smallest forward discretization step size $\delta t_{\min}$ and let $L_{\min} = \frac{T}{\delta t_{\min}}$. According to the description of the alternative jump time simulation in Section 18.2.1, all components $\delta J_{t_{l_{\min}}}^{b,\delta t_{\min}},\ l_{\min} = 1,\ldots,L_{\min}$, can be written as

$$\delta J_{t_{l_{\min}}}^{b,\delta t_{\min}} = \begin{cases} h_j & \text{if } l_{\min} = l_j,\ j = 1,\ldots,N(T), \\ 0 & \text{otherwise,} \end{cases} \tag{135}$$

where $l_j = \left\lfloor u_j \frac{L_{\min}}{T} \right\rfloor$ and thus $t_{l_j} = \left\lfloor u_j \frac{L_{\min}}{T} \right\rfloor \frac{T}{L_{\min}}$ with u_j uniformly distributed on $[0,T]$. The compound Poisson process is then given by

$$J_{t_k}^{b,\delta t_{\min}} = \sum_{l_{\min}=1}^{k} \delta J_{t_{l_{\min}}}^{b,\delta t_{\min}},\ k = 1,\ldots,L_{\min}. \tag{136}$$

Again, let $\delta t = a\delta t_{\min},\ a \in \mathbb{N}$, be a larger forward discretization step size where a is a divisor of $L_{\min}$. We construct the compound Poisson processes for δt based on those of $\delta t_{\min}$ by

$$\delta J_{t_{al}}^{b,\delta t} = \sum_{l_{\min}=a(l-1)+1}^{al} \delta J_{t_{l_{\min}}}^{b,\delta t_{\min}},\ l = 1,\ldots,L, \tag{137}$$

with $L = \frac{L_{\min}}{a}$ so that we obtain

$$J_{t_{ak}}^{b,\delta t} = \sum_{l=1}^{k} \delta J_{t_{al}}{}^{b,\delta t} = \sum_{l=1}^{k} \left(\sum_{l_{\min}=a(l-1)+1}^{al} \delta J_{t_{l_{\min}}}^{b,\delta t_{\min}} \right),\ k = 1,\ldots,L. \tag{138}$$

Univariate (#12)			Multivariate (#22)			
r	r^2	r^3	rK	r^2K	rK^2	rKS
K	K^2	K^3	rS	r^2S	rS^2	$rK\mu$
S	S^2	S^3	$r\mu$	$r^2\mu$	$r\mu^2$	$rS\mu$
μ	μ^2	μ^3	KS	K^2S	KS^2	$KS\mu$
			$K\mu$	$K^2\mu$	$K\mu^2$	
			$S\mu$	$S^2\mu$	$S\mu^2$	

Table 9: Starting set of basis functions.

By employing an index transformation and using (136), this expression becomes

$$J_{t_{ak}}^{b,\delta t} = \sum_{l_{\min}=1}^{ak} \delta J_{t_{l_{\min}}}^{b,\delta t_{\min}} = J_{t_{ak}}^{b,\delta t_{\min}}, \ k = 1, \ldots, L, \tag{139}$$

which is by definition a compound Poisson process, see (136), differing only in the forward discretization step size. Hence the constructions in (137) are indeed compound Poisson processes for the larger forward discretization step size δt. $\square$

18.5 Basis Functions Consistency

18.5.1 Principle of Parsimony

We can subsume the process according to which we select the basis functions under two principles. The objective is to choose as few basis functions as possible and at the same time as many basis functions as necessary to explain the continuation values adequately. By doing so, we follow again the principle of parsimony by Burnham & Anderson (2002) which we have already explained in the previous part of this thesis. While our first principle ensures that no basis functions with very little or no explanatory power enter the regressions, our second principle ensures that all basis functions with significant explanatory power are considered. To account for the first principle we do not use basis functions which show very high correlations with simpler basis functions. We follow the second principle by selecting the basis functions at times $t > 0$ and at time $t = 0$ *consistently.* As we will see, the second principle also serves as a validation concept. Below, we describe the two principles in detail.

18.5.2 Exclusion by Correlation

We start the regressions with a set of polynomial basis functions up to order three since this is the choice of Bacinello et al. (2009). In contrast to their choice of orthogonal basis functions, we decide for ordinary monomial basis functions and apply QR decompositions to the design matrices. Since our model comprises the four risk factors defined in (106), (107), (108) and (110), the starting set makes in total $H = 35$ basis functions including a constant function to allow for an intercept. It is reasonable to reflect the process of the reference fund value (108) either by polynomial basis functions with respect to S or Y. As an implication of the second principle, we choose S. For illustration purposes, we list our starting set of basis functions (without the constant function) in Table 9.

Throughout all $t > 0$ regressions $m = M - 1, \ldots, 1$, let $\boldsymbol{e}(\cdot) = (e_1(\cdot), \ldots, e_H(\cdot))^{\mathrm{T}}$ be the basis function vector as defined in Section 18.3.1. Then $\boldsymbol{e}\left(X_{t_m}^i\right)$, $i = 1, \ldots, N$, with

$X^i_{t_m} = \left(r^i_{t_m}, K^i_{t_m}, S^i_{t_m}, \mu^i_{t_m}\right)$ is the basis function vector evaluated at the time-t_m values of the state variable process in the i-th simulation. For all pairs $\left(e_{h_1}(\cdot), e_{h_2}(\cdot)\right)$ of basis functions, $h_1, h_2 = 1, \ldots, H$ with $h_1 \neq h_2$, we calculate Pearson's correlation coefficients $r^{t_m}_{h_1,h_2}$, i.e.,

$$r^{t_m}_{h_1,h_2} = \frac{\sum_{i=1}^N \left(e_{h_1}\left(X^i_{t_m}\right) - \bar{e}_{h_1}\left(X_{t_m}\right)\right)\left(e_{h_2}\left(X^i_{t_m}\right) - \bar{e}_{h_2}\left(X_{t_m}\right)\right)}{\sqrt{\sum_{i=1}^N \left(e_{h_1}\left(X^i_{t_m}\right) - \bar{e}_{h_1}\left(X_{t_m}\right)\right)^2}\sqrt{\sum_{i=1}^N \left(e_{h_2}\left(X^i_{t_m}\right) - \bar{e}_{h_2}\left(X_{t_m}\right)\right)^2}}, \tag{140}$$

where $\bar{e}_{h_1}\left(X_{t_m}\right) = \frac{1}{N}\sum_{i=1}^N e_{h_1}\left(X^i_{t_m}\right)$ and $\bar{e}_{h_2}\left(X_{t_m}\right) = \frac{1}{N}\sum_{i=1}^N e_{h_2}\left(X^i_{t_m}\right)$. If for a fixed pair (h_1, h_2), all absolute values of the correlation coefficients $|r^{t_m}_{h_1,h_2}|$, $m = M-1, \ldots, 1$, exceed a given threshold, we exclude the more complex basis function of $\left(e_{h_1}(\cdot), e_{h_2}(\cdot)\right)$ from the $t > 0$ regressions. In this way, we simplify the regressions, increase their numerical stability, reduce the issue of off-setting coefficients, and make the economic interpretation of the remaining basis functions easier.

For now, suppose that the $t = 0$ regression is conducted just like the $t > 0$ regressions so that we have $\boldsymbol{e}^{\mathbf{0}}(\cdot) = \boldsymbol{e}(\cdot)$. We cannot tell whether the obtained basis functions from applying the first principle explain the continuation values adequately because there might still be basis functions with significant explanatory power missing. To cope with this question, we introduce the second principle, a validation concept, according to which the basis functions at times $t > 0$ and at time $t = 0$ are selected *consistently*.

18.5.3 Validation Concept

To apply the second principle, we define a set of validation scenarios $\left\{X^j_0 \mid j = 1, \ldots, V\right\}$ consisting of hand-picked outer scenarios. The validation scenarios should be selected such that they represent the range of relevant outer scenarios well. Then we run the LSMC algorithm depicted in Figure 29 exclusively for each validation scenario X^j_0, $j = 1, \ldots, V$, and therefore set $X^{j,i}_0 = X^j_0,\ i = 1, \ldots, N,\ j = 1, \ldots, V$, as in the previous section. By doing so, we establish again the setting of Bacinello et al. (2009) in which the regression at contract inception reduces to averaging. As a result, we directly obtain all corresponding validation insurance contract values $\widetilde{V}^{*,j}_0$, $j = 1, \ldots, V$. We call the pairs $\left(X^j_0, \widetilde{V}^{*,j}_0\right)$, $j = 1, \ldots, V$, *validation points* in analogy to the notion in the foregoing two parts of this thesis. To generate V validation points, we perform the algorithm in Figure 29 V times.

In addition to the V *exclusive* LSMC runs, we run the *extended* LSMC algorithm once. Then we validate the results from the extended LSMC run and the V exclusive runs by comparison. We call the basis functions used in the $t > 0$ regressions and in the $t = 0$ regression *consistent* and the validation successful if, for the extended LSMC run and the V exclusive LSMC runs,

- the pointwise deviations of the contract value estimates do not exhibit a massive systematic pattern over all validation scenarios;
- the contract value estimates do not deviate by more than a given threshold from each other;
- the same basis functions per regression are used unless they need to be dropped or replaced properly for degenerating reasons in the $t = 0$ regression.

Degenerating reasons prevail if model characteristics (e.g., a return over time at inception) or the absence of outer scenarios (e.g., averaging at inception in the *exclusive* LSMC runs) cause different basis functions to take on the same values. We repeat the validation procedure with different constellations of economically promising basis functions until consistency is achieved and the validation is successful. In other words, the basis functions are calibrated such that they provide consistent results in the exclusive LSMC runs and extended LSMC run.

18.6 Summary

To apply the theoretical model, a suitable implementation algorithm has been set up after some general remarks in Section 18.1. At first, forward discretization methods working for the parameter values in the numerical example have been presented in Section 18.2. While for the reference fund value an adapted version of the Euler-Maruyama method has been used, for the remaining risk factors the so-called explicit scheme $E(0)$ has been used to avoid negative simulation values. Furthermore, a convenient procedure for the simulation of compound Poisson processes has been described. Thereafter, the backward induction procedure has been introduced as an approximation technique to find a solution to the optimal stopping problem. Since we have only been able to simulate discrete processes, practically, we have switched over to pricing an insurance contract with a Bermudan- instead of an American-style surrender option. To estimate the continuation values in the backward induction procedure at each time step, in Section 18.3, a Markovian environment has been assumed and two approximations conditional on finite sets of basis functions and ordinary least-squares estimators have been derived. At contract inception, the insurance contract values with respect to infinitely many outer scenarios have been obtained by an additional regression function. The resulting LSMC algorithm written with the aid of pseudo-code has completed the regression theme.

In Section 18.4, we have shown that a misbalance between the number of simulations and forward discretization step size causes either a suboptimal computational power usage or biased estimation results. To eliminate these undesirable effects, a harmonization algorithm has been presented and a seed-related modification thereof for run time reductions proposed. Afterwards, the concept of consistent basis functions has been introduced in Section 18.5. Basis functions with very little explanatory power have been excluded based on Pearson's correlation coefficients in accordance with the principle of parsimony. Furthermore, only consistent basis functions have been allowed. The concept of consistent basis functions has also been said to serve as a validation concept.

19 Numerical Example

19.1 General Remarks

19.1.1 Model Specifications

We take up the numerical experiment of Bacinello et al. (2009) so that we can use their results as a benchmark. Let the physical probability measure $\mathbb{P}$, characterizing the outer scenarios, be the multivariate uniform distribution on a 4-dimensional cube S_{fit} around a given central outer scenario $X_0 = (r_0, K_0, S_0, \mu_0)$, which we call the *base scenario.* The components of the base scenario are referred to as the *base values.* While the definitions of the outer scenario values r_0, K_0 and μ_0 are unambiguous, the one of the initial reference fund value S_0 is not. It is possible to let the insured invest in the reference fund either before it has been stressed by $\mathbb{P}$ or afterwards. We decide for the more interesting former definition as it lets the insured participate at the shock and leads to a wider range of insurance contract values. In addition, we let the single initial premium F_0 be equal to the initial reference fund value S_0^{before} so that the insured invests in exactly one unit. After the investment, the reference fund is stressed and its value becomes $S_0 := S_0^{\text{after}}$.

To derive the function $m(\cdot)$ in discretized process (117) representing the time-dependent long-run values of the intensity of mortality we follow the suggestion in Bacinello et al. (2009) by fitting a Weibull intensity $m(t) = c_1^{-c_2} c_2 (x+t)^{c_2-1}$ with $c_1 > 0, c_2 > 1$ to the survival probabilities implied by Table SIM2001 (male) employed in the Italian endowment insurance market.

19.1.2 Forward Discretization Schemes

The parameter choices of Bacinello et al. (2009) partly demand schemes going beyond the natural Euler-Maruyama method. For instance, the simulated values of (107) turn out to become negative in some simulations when using the Euler-Maruyama scheme since the Gaussian increment in this scheme is not bounded from below, see Alfonsi (2005). This is the reason why we only use the Euler-Maruyama method to simulate process (108), see (114), and why we use the explicit scheme $E(0)$ to simulate the other processes (106), (107) and (110), see (115), (116) and (117), respectively. The forward discretization step size used to determine the final insurance contract value will depend on the analysis undertaken to balance the Monte Carlo error and discretization bias out.

Process (110) might even become negative when simulated by the explicit scheme $E(0)$ as the inequality $\sigma_\mu^2 \leq 4\zeta_\mu m(t)$ from Section 18.2.1 does not hold for each $t \geq 0$ in our numerical experiment. Since the inequality is violated only marginally though, the simulated values (117) turn negative only occasionally. For this reason, we do not search for another scheme but deal with this issue by reversing negativity. If we have, for example, $\mu_{t+1} < 0$ for $t = t_l,\ l = 1, \ldots, L$, we subtract the term $\left(\zeta_\mu m(t) - \frac{\sigma_\mu^2}{4}\right)\delta t + \delta J_t^\mu$ which we just added.

To derive the results for several forward discretization step sizes with the same seed according to the procedure in Section 18.4, we simulate the Brownian motions and compound Poisson processes only once with the smallest forward discretization step size $\delta t_{\min}$, save the paths and reaccess them for the larger forward discretization step sizes $\delta t = a\delta t_{\min},\ a \in \mathbb{N}$.

19.1.3 Memory Size & Parallelization

Another numerical difficulty with the parameter choices of Bacinello et al. (2009) concerns the memory capacities of our hardware. Running the software 32-bit R on a 64-bit Windows version allows to store memory sizes of at maximum 4 GB, which is exceeded when all Brownian motions Z^r, Z^K, Z^Y, Z^μ, compound Poisson processes J^Y, J^μ and further required values of the stochastic processes $r, K, S, \mu, \tau, F^s, F^d, F^w$ are supposed to be stored $N = 10,000$ times simultaneously with forward discretization step size $\delta t = 0.001$. By applying software 64-bit R under 64-bit Windows, where maximum memory sizes of 8 TB can be achieved at the time we do the calculations, see R Documentation (2017), we overcome this computational challenge. In settings in which the maximum values of obtainable memory are exceeded, one can, for instance, store $N_1 < N$ Brownian motions and compound Poisson processes at the same time, derive the further required values of the stochastic processes $r, K, S, \mu, \tau, F^s, F^d, F^w$, and finally delete the N_1 Brownian motions and compound Poisson processes. Then the available memory can be allocated to the Brownian motions and compound Poisson processes required for the derivation of the next $N_2 \leq N - N_1$ stochastic processes, and so on. Thereby only the values of the processes on the backward time grid need to be stored as only these are processed in the regressions.

Run time reductions can be achieved by operating on multiple cores. The stochastic processes corresponding to different simulations can be generated separately on multiple cores. For such a parallelization, the software R provides package *doParallel* which executes loops from package *foreach* in parallel.

19.2 Preliminary Considerations

19.2.1 Initialization of Settings

Our initial set of polynomial basis functions up to order three comprises $H = 35$ basis functions of which all except for the constant function are shown in Table 9. Besides the constant function, the initial set includes risk factor wise the univariate polynomials of degrees one to three, and 22 additional basis functions reflecting the interactions between the risk factors. The latter 22 basis functions are products of the univariate polynomials with degrees summing up to at maximum three. In all regressions, we work with QR decompositions of the resulting design matrices to ensure numerical stability, compare Section 14.1.4.

For the derivation of consistent basis functions, our calibration procedure takes the following course. We make first guesses for a reasonable number of simulations N and forward discretization step size δt. Our initializations of $N = 50,000$ and $\delta t = 0.1$ are inspired by Bacinello et al. (2009) who set $N = 19,000$ and $\delta t = 0.01$. We thereby decide for a larger sample size to increase the accuracy and for a larger forward discretization step size to decrease the run time. After we will have found a consistent set of basis functions according to the procedure described in Section 19.2.3, we will adjust our first guesses as shown in Section 19.2.4 to harmonize the Monte Carlo error and discretization bias. The initializations should be made carefully to ensure that the basis functions remain consistent until after the harmonization.

Since we set the maturity of the insurance contract equal to $T = 15$ years, a one-year backward discretization step size is appropriate from both a computational and economic perspective. With these choices, the numerical setting comprises $L = 150$ forward and $M = 15$ backward discretization steps.

19.2.2 Validation Scenarios

We hand-pick a set of $V = 29$ validation scenarios on the 4-dimensional cube S_{fit}. Let the set $\left\{X_0^i \mid i = 1, \ldots, N'\right\}$ be a realization of $\mathbb{P}$ with $N' \geq N$ on S_{fit}. For each risk factor, we take the empirical 10%- and 90%-quantiles on $\left\{X_0^i \mid i = 1, \ldots, N'\right\}$ and denote them by $b-, b+,\ b \in \{r, K, S, \mu\}$. We decide for these particular quantiles as they are neither too close to the base values nor too close to the boundaries. Furthermore, we combine selected quantiles with selected base values r_0, K_0, S_0, μ_0. Preferably, the 29 validation scenarios are well-distributed over S_{fit}. We assume that if we achieve consistent results in

Base/Univariate (#9)	**Bi-/Trivariate (#10)**	**Multivariate (#10)**
base	$S+, \mu-$	$r-, K-, S-, \mu-$
$r-$	$K-, \mu+$	$r-, K-, S-, \mu+$
$r+$	$r-, S+$	$r-, K-, S+, \mu-$
$K-$	$r-, \mu-$	$r-, K+, S-, \mu-$
$K+$	$r+, \mu-$	$r+, K-, S-, \mu-$
$S-$	$K-, \mu+$	$r+, K+, S+, \mu+$
$S+$	$r-K+$	$r+, K+, S+, \mu-$
$\mu-$	$K+, S-$	$r+, K+, S-, \mu+$
$\mu+$	$K+, S+, \mu-$	$r+, K-, S+, \mu+$
	$r-, S+, \mu+$	$r-, K+, S+, \mu+$

Table 10: Set of validation scenarios.

rather extreme validation scenarios like the ones containing the 10%- and 90%-quantiles, we do so for less extreme ones as well. In Table 10, we report the hand-picked validation scenario selection.

As the first validation scenario, we select the base scenario $X_0 = (r_0,\ K_0,\ S_0,\ \mu_0)$, which has a central position on S_{fit}. The next eight validation scenarios comprise single risk factor deviations from the base scenario such that for each risk factor an upwards and downwards stress is included. These validation scenarios capture the effects of single risk factors on the insurance contract value. Each of the subsequent ten validation scenarios contains two or three arbitrary risk factor shocks. The last ten validation scenarios are deviated in each risk factor dimension and are thus the most challenging scenarios of our selection.

19.2.3 Basis Function Calibration

We run the extended LSMC algorithm under realization $\left\{X_0^i \mid i = 1, \ldots, N\right\}$ of measure $\mathbb{P}$ and use in each regression the basis functions from Table 9 plus a constant function to allow for an intercept. Afterwards, we calculate all correlation coefficients (140) and exclude the basis functions showing an absolute correlation of at least 0.97 with a simpler basis function in each regression. The only remaining multivariate basis functions are rK and rS, which means that they capture all essential interactions between r, K and S in (108) and (114). Finally we run both the extended LSMC algorithm and the exclusive LSMC algorithm for each of the 29 validation scenarios from Table 10 with the reduced set of basis functions. By having in addition the exclusive LSMC runs with the basis functions from Table 9 as reference runs available, we observe that the exclusion of highly correlated basis functions improves indeed the validation results, compare Section 18.5.1.

However, the validation is not successful yet as the estimated insurance contract values show a systematic pattern. The insurance contract values derived in the extended and exclusive LSMC runs deviate from each other whenever the corresponding validation scenarios include the 10%- and 90%-quantiles of the reference fund value, i.e., $S-$ and $S+$. The contract values derived by the extended LSMC algorithm are smaller than the ones derived by the exclusive LSMC algorithm if the validation scenario contains $S-$ whereas it is vice versa if it contains $S+$.

To eliminate this systematic pattern, we reconsider the economic setting and search for economically more appropriate basis functions. By analyzing the contract value (103),

Univariate (#7)		Multivariate (#1)	Excl. & Ext. $t = 0$ LSMC Regressions (#1)	Extended $t > 0$ LSMC Regressions (#3)		
r						
K	K^2	rK				
S	S^2		rS	$\frac{S}{S_0}$	$\left(\frac{S}{S_0}\right)^2$	$r\frac{S}{S_0}$
μ	μ^2					

Table 11: Final set of consistent basis functions.

we recognize that not only the reference fund value S_t but also the return $\frac{S_t}{S_0}$ drives the policyholder's surrender behavior. As long as all outer scenarios are equal, we have $S_0^i = S_0^j,\ i \neq j$, for all $i, j = 1, \ldots, N$ meaning that the basis functions in terms of the reference fund value and its return are equivalent. While the return is thus not required for the modeling in the exclusive LSMC runs, it certainly is in the extended LSMC run. Besides that, we find that using basis functions with respect to Y_t instead of S_t does not improve the results. For this test, we draw new initial values Y_0^i, $i = 1, \ldots, N$, from a uniform distribution around the new base value $Y_0 = \log(S_0)$ and assign them to the realization of $\mathbb{P}$ by substituting them for S_0^i, $i = 1, \ldots, N$.

After having run the extended LSMC algorithm and the exclusive LSMC algorithm with the economically more appropriate basis functions, our validation is successful and consistency in the basis functions is achieved. The final set of basis functions (without the constant function) for the parameterizations $\kappa_s = \kappa_d = 0.04,\ \kappa_w = 0$ and $\kappa_s = \kappa_d = 0,\ \kappa_w = 0.06$ is listed in Table 11. The basis functions in the first two main columns "univariate" and "multivariate" enter all regressions of both the exclusive and extended LSMC algorithm. The last two main columns "exclusive & extended $t = 0$ LSMC regressions" and "extended $t > 0$ LSMC regressions" contain the basis functions exclusively used in the specified regressions. In the exclusive LSMC algorithm, the $t = 0$ regression reduces to averaging so that we only need to provide basis functions for the $t > 0$ regressions here. In the extended LSMC algorithm, the $t = 0$ regression plays a special role as well because the basis functions in the last main column degenerate either into a constant function or the univariate basis function r. The basis functions degenerating into a constant function are simply dropped and the one degenerating into r is replaced by rS.

19.2.4 Forward Discretization Bias

Due to limited computational capacities, we use only $N = 10,000$ simulations to balance out the Monte Carlo error and discretization bias according to the three steps in Section 18.4. A transition from, for example, $50,000$ to $10,000$ simulations is appropriate as long as the results remain sufficiently precise. The choice of $N = 10,000$ simulations is motivated by the procedure described in Desmettre & Korn (2015). To determine a suitable forward discretization step size δt_0 conditional on N, we conduct the exclusive LSMC algorithm for the base scenario $X_0 = (r_0,\ K_0,\ S_0,\ \mu_0)$ with several different forward discretization step sizes. In each LSMC run, we use the basis functions shown in the left three main columns of Table 11 plus a constant function.

For the parameterizations $\kappa_s = \kappa_d = 0.04,\ \kappa_w = 0$ and $\kappa_s = \kappa_d = 0,\ \kappa_w = 0.06$, the

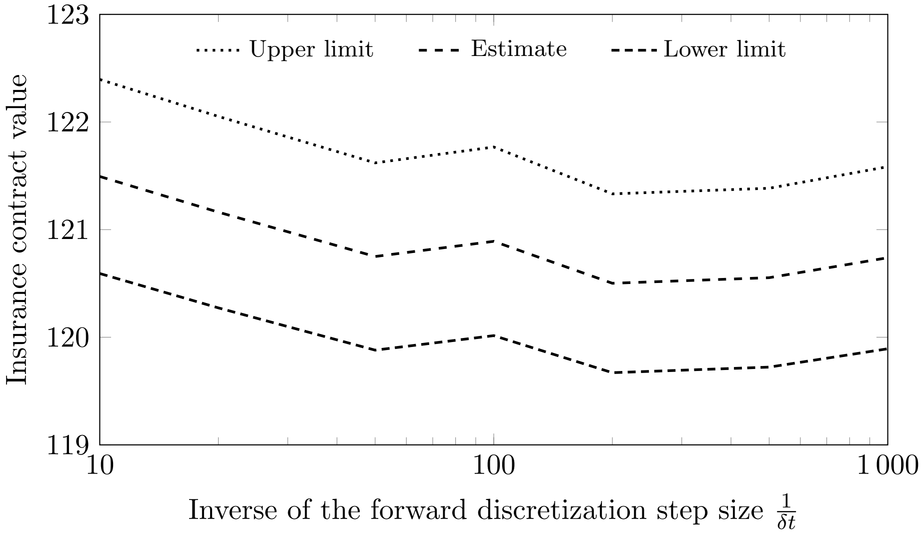

Figure 31: Estimated insurance contract values with 95%-confidence intervals for varying forward discretization step sizes δt, parameterization $\kappa_s = \kappa_d = 0.04, \kappa_w = 0$.

estimation results from the repetitions of the second step, in which the forward discretization step size is reduced gradually from $\delta t = 0.1$ to $\delta t = 0.001$, are depicted in Figures 31 and 32, respectively. The confidence intervals in both figures have levels of $1 - \alpha = 95\%$. As the forward discretization step size decreases, the estimated insurance contract values decrease at first and then stabilize. Since a stabilization indicates that the Monte Carlo error and discretization bias are harmonized, one can stop the refinement at this point. According to Figures 31 and 32, the estimated insurance contract values begin to stabilize at $\delta t_0 = 0.01$. Therefore, $\delta t_0 = 0.01$ will be the forward discretization step size of our choice in the final calculations in Section 19.3.

In deviation from Desmettre & Korn (2015), we do not perform the exclusive LSMC algorithm multiple times with the same forward discretization step sizes to obtain average insurance contract values. Instead we generate the Brownian motions Z^r, Z^K, Z^Y, Z^μ and compound Poisson processes J^Y, J^μ in (114), (115), (116) and (117) only once with our smallest forward discretization step size δt_{min} and construct the realizations for larger forward discretization step sizes based on those for δt_{min}. By this approach, we ensure that we use the same seed throughout the entire harmonization procedure and eliminate further undesirable random fluctuations.

19.3 Final Results

19.3.1 Settings

For a male individual of age $x = 40$ entering an equity-linked endowment insurance contract with the option to surrender the contract before its maturity of $T = 15$ years as defined by (103), we run $N = 10,000$ simulations for two different constellations of minimum interest rate guarantees (κ, κ_w) where we set $\kappa := \kappa_s = \kappa_d$. Even though in practice only constellations with $\kappa_w \leq \kappa$ are reasonable, we also test a constellation with $\kappa_w > \kappa$ for numerical purposes. We decide for the two constellations $\kappa = 0.04,\ \kappa_w = 0$ and $\kappa = 0,\ \kappa_w = 0.06$ and let the individual invest in exactly one unit of the reference fund

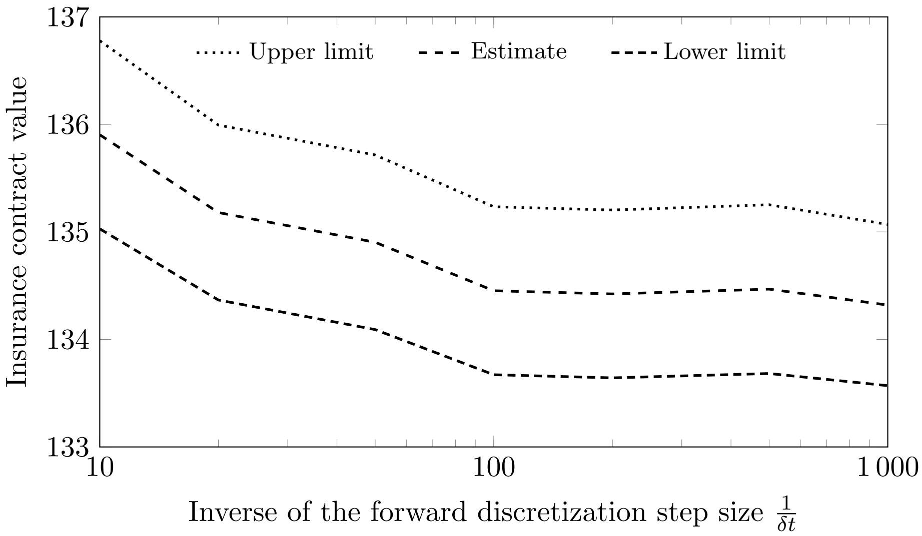

Figure 32: Estimated insurance contract values with 95%-confidence intervals for varying forward discretization step sizes δt, parameterization $\kappa_s = \kappa_d = 0, \kappa_w = 0.06$.

and set thus $F_0 := S_0$.

On the one hand, we perform $V = 29$ exclusive LSMC runs where each run corresponds to another validation scenario from Table 10 so that the different insurance contracts associated with these validation scenarios are valued separately. This setting in which no regression takes place at contract inception is the one introduced by Bacinello et al. (2009). On the other hand, we carry out the extended LSMC run with the additional regression function at contract inception once and thereby achieve the simultaneous valuation of infinitely many insurance contracts. For the latter task, we have drawn outer scenarios from the the physical measure $\mathbb{P}$ on the 4-dimensional cube S_{fit} around the base scenario $X_0 = (r_0, K_0, S_0, \mu_0) = (0.05, 0.04, 100, m(0))$ with $m(0) = c_1^{-c_2} c_2 (x)^{c_2 - 1}$. We let r_0^i, K_0^i, S_0^i and μ_0^i, $i = 1, \ldots, N$, be distributed uniformly on $[0, 2r_0], [0, 2K_0], [75, 125]$ and $\left[m(0) - 4 \cdot 10^{-4}, m(0) + 4 \cdot 10^{-4}\right]$, respectively. In the regressions, we use the final set of consistent basis functions of which all except for the constant function are reported in Table 11. They differ for the exclusive and extended LSMC runs only for degenerating reasons.

As the forward discretization step size we set $\delta t_0 = \frac{T}{L} = 0.01$ years which has been shown in the previous section to work well with our choice of $N = 10,000$ simulations. As the backward discretization step size we set $\frac{T}{M} = 1$ year which is suitable from a computational and economic perspective. The entire parameterization is summarized in Table 12.

19.3.2 LSMC Runs

The insurance contract values V_0^* estimated by the exclusive and extended versions of the algorithm depicted in Figure 29 can be viewed for the constellations $\kappa = 0.04$, $\kappa_w = 0$ and $\kappa = 0$, $\kappa_w = 0.06$ in Tables A37 and A38, respectively. The notation of the validation scenarios is explained in Section 19.2.2. Per constellation of guaranteed minimum interest rates, we have performed the $V = 29$ exclusive LSMC runs and the single extended LSMC run under three different Monte Carlo settings. In the first setting, we use different

General	$\boldsymbol{r}$	$\boldsymbol{K}$	$\boldsymbol{S}$	$\boldsymbol{\mu}$
$N = 10{,}000$	$r_0 = 0.05$	$K_0 = 0.04$	$S_0 = 100.00$	$\mu_0 = m(0)$
$T = 15$	$\zeta_r = 0.60$	$\zeta_K = 1.50$	$\rho_{YK} = -0.70$	$\zeta_\mu = 0.50$
$L = 1500$	$\delta_r = 0.05$	$\delta_K = 0.04$	$\rho_{Yr} = 0.00$	$\sigma_\mu = 0.03$
$M = 15$	$\sigma_r = 0.03$	$\sigma_K = 0.40$	$\lambda_Y = 0.50$	$\lambda_\mu = 0.10$
$H = 12/10$			$\mu_Y = 0.00$	$\gamma_\mu = 0.01$
$H_0 = 10/1$			$\sigma_Y = 0.07$	$c_1 = 83.70$
$V = 29$				$c_2 = 8.30$
$F_0 = S_0$				$x = 40$

Table 12: Parameterization in the extended/exclusive LSMC algorithm.

realizations of the Brownian motions and compound Poisson processes across all scenarios. We report the results in the first three connected columns "Am. - Random Seed" of Tables A37 and A38, respectively. Differently, in the second setting, the $29 + 1$ LSMC runs per constellation of guaranteed minimum interest rates are conducted based on the same Brownian motions and compound Poisson processes so that they have the same seed. Besides, the LSMC runs from Section 19.2.4 for the derivation of the forward discretization step size δt_0 were performed based on this seed as well. The results can be viewed in the second three connected columns "Am. - Same Seed" of Tables A37 and A38, respectively. In the last three columns "Eu. - Same Seed", we present the results of the third setting in which the American/Bermudan option is replaced by a European option.

When replacing the American/Bermudan option by a European option in the insurance contract, the parameter κ_w becomes irrelevant as the surrender option vanishes. The insurance contract value can then be computed by executing only Steps 1 and 3 of the algorithm in Figure 29. The difference between the American/Bermudan and the European insurance contract value yields the pure American/Bermudan option value.

19.3.3 Economic Interpretation

The results allow the following economic interpretation. Conditional on a specific stress, an insurance contract with $\kappa = 0$, $\kappa_w = 0.06$ is worth more than one with $\kappa = 0.04$, $\kappa_w = 0$ as the values associated with the former contract are greater than the ones associated with the latter contract over all validation scenarios. This means that a minimum interest rate guarantee of $\kappa_w = 0.06$ paid out upon surrender is worth more than a minimum interest rate guarantee of $\kappa = 0.04$ paid out upon survival or death. This result is plausible because, as long as the probability of death until maturity is very small, it is likely that there are several times at which it is profitable for the insured to surrender the contract whereas there is only one time at which the insured either survives maturity or dies prior to that. Moreover, in our comparison the minimum interest rate guarantee paid upon surrender is larger than the one paid upon survival or death.

As a plausibility check, we compare our base scenario results from the second and third Monte Carlo settings "Am. - Same Seed" and "Eu. - Same Seed" with the results obtained by Bacinello et al. (2009). For $\kappa = 0.04$, $\kappa_w = 0$, our algorithm yields for the American version an insurance contract value of 121.10 compared to 123.09 and for the European version a value of 121.40 compared to 122.90. The higher value obtained for the European product thereby lies within the error tolerance for the value of the American product and signals that early exercise is not optimal for the policyholder under the given parameter constellation. For $\kappa = 0,\ \kappa_w = 0.06$, our algorithm values the American version by 135.55 compared to 137.13 and the European version by 105.23 compared to 107.19. Here, early exercise clearly is optimal. We assume that the deviations arise from several different implementation choices because, according to the confidence intervals depicted in Figures 31 and 32 and the comparison with the results from the first Monte Carlo setting "Am. - Random Seed", randomness is not able to fully explain the deviations.

19.3.4 Sensitivity Analysis

Tables A37 and A38 reveal the following sensitivities of the insurance contract values to the risk factors. While a negative shock on the interest rate r_0 increases the insurance contract value, a positive shock on r_0 decreases it due to discounting effects. The higher the interest rate is, the larger is the discounting and thus the lower the contract value. A negative shock on the reference fund value S_0 leads to a lower value whereas a positive shock on S_0 leads to a higher value. This result is reasonable because the benefits paid upon contract termination depend on the reference fund value. If we have $F_0 = S_0$, the benefits in (103) become $F_t^e = \max\left(S_t, S_0 \exp\left(\kappa_e t\right)\right),\ e \in \{s, d, w\}$, so that they are proportional to the shocked value of S_0 whenever $S_0 \exp\left(\kappa_e t\right) \geq S_t$ holds. In the case of $S_0 \exp\left(\kappa_e t\right) < S_t$, the benefits are equal to S_t, which also increases with S_0 according to (114). Based on the results derived with sample size $N = 10,000$, we can only state that the effects of the stresses on the stochastic volatility K_0 and the intensity of mortality μ_0 are low compared to those associated with r_0 and S_0. In the exclusive LSMC run for the base scenario, the standard error $\frac{\widehat{\sigma}_N^2}{\sqrt{N}}$ with $\widehat{\sigma}_N^2$ from (129) turns out to be 0.48 for $\kappa = 0.04,\ \kappa_w = 0$ and 0.40 for $\kappa = 0,\ \kappa_w = 0.06$.

Given a confidence level of $1 - \alpha = 95\%$, the sample size of $N = 10,000$ provides a confidence interval length of 1.88 for $\kappa = 0.04,\ \kappa_w = 0$ and of 1.56 for $\kappa = 0,\ \kappa_w = 0.06$. To achieve a confidence interval length of 0.2, sample sizes of $N = 900,000$ and $N = 600,000$, respectively, would be necessary. We assume these sample sizes would show that the insurance contract value increases slightly in K_0. This is our expectation because with a higher stochastic volatility, the reference fund is more likely to attain comparatively large and small values where the large values affect the insurance contract value more than the small ones since the benefits paid out are bounded from below by the minimum interest rate guarantees. If $\kappa < \kappa_w$, we expect a slight fall of the insurance contract value in μ_0 as a higher intensity of mortality induces a higher probability for the insured to die before maturity and thus less surrender opportunities so that overall lower minimum interest rate guarantees are expected. However, if $\kappa \geq \kappa_w$, we expect in accordance with Bacinello et al. (2009) a slight rise of the insurance contract value in μ_0 since then the value of the standard endowment insurance contract is increased (can be shown) while there is no effect in the opposite direction.

19.4 Summary

In Section 19.1, we have defined the shocks to the risk factors, derived a function for the intensity of mortality based on Italian survival probabilities, and addressed numerical obstacles incurred with our parameter choices. For instance, the explicit scheme $E(0)$ has been motivated and the issue of limited memory capacities discussed. Then, the basis function selection procedure has been initialized by first guesses of a suitable number of simulations and forward discretization step size in Section 19.2. Validation scenarios have been hand-picked to serve the selection of consistent basis functions in the exclusive and extended LSMC runs. Moreover, the principle of parsimony has been ensured by excluding basis functions showing correlations above 0.97 with simpler ones. To achieve consistency, economically appropriate basis functions have been selected. For the harmonization of the forward discretization step size and number of Monte Carlo simulations, a comparatively efficient seed-related procedure has been applied. As expected, the estimated insurance contract values have stabilized after gradually reducing the forward discretization step size.

We have performed the final LSMC runs for two different constellations of guaranteed minimum interest rates in Section 19.3. Per constellation, we have calculated the insurance contract values for the three categories "American/Bermudan surrender option - different seeds", "American/Bermudan surrender option - same seed", "European surrender option - same seed" and reported them in Tables A37 and A38. We have concluded that the insurance contract was worth more if $\kappa_s = \kappa_d = 0,\ \kappa_w = 0.06$ than if $\kappa_s = \kappa_d = 0.04,\ \kappa_w = 0$. Additionally, we have found that the insurance contract value increased if the interest rate was stressed negatively or the reference fund positively. Moreover, we have made statements about confidence interval lengths and described how we expected the insurance contract value to change if the stochastic volatility or mortality were stressed.

20 Conclusion

Summary

In the third and last part of this thesis, we have described how life insurance contracts with early exercise features can be valued by an LSMC-based approach. We have extended the model setting in Bacinello et al. (2009) by a hybrid probability measure such as the one introduced in Bauer & Ha (2015) and Natolski & Werner (2016), formulated a validation concept and have taken the forward discretization bias into account. By applying the physical probability measure we have accomplished run time reductions and raised the question of how to select suitable basis functions. We have answered this question by formulating a validation concept according to which the basis functions are selected consistently. Additionally, we have broached the issue of forward discretization biases in LSMC settings based on Desmettre & Korn (2015), confirmed the prevalence of this issue in our model and proposed a comparatively efficient seed-related implementation technique to find the optimal forward discretization step size conditional on a given sample size. This seed-related technique gets along with only one Monte Carlo simulation per stochastic process for all forward discretization step sizes together.

Outlook

In the setting above, we introduced the physical probability measure to capture variations in the risk factors at contract inception. But the application is not limited to risk factors. We expect that an extension to variations in other product features such as minimum interest rate guarantees is possible. Similarly, this holds for the application in Parts I and II: Parameters of the CFP model such as the ones steering the management actions can be included in the LSMC model for a sensitivity analysis.

Furthermore, machine learning approaches such as from Part II or variations thereof can be applied to value life insurance contracts with early exercise features.

References

Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle, *in* B. N. Petrov & F. Csáki, eds, 'Proceedings of the 2nd International Symposium on Information Theory', Akadémiai Kiadó, Budapest, Hungary, pp. 267–281.

Alfonsi, A. (2005), 'On the discretization schemes for the CIR (and bessel squared) processes', *Monte Carlo Methods and Applications* **11**(4), 355–384.

Bacinello, A. R., Biffis, E. & Millossovich, P. (2009), 'Pricing life insurance contracts with early exercise features', *Journal of Computational and Applied Mathematics* **233**, 27–35.

Bakshi, G., Cao, C. & Chen, Z. (1997), 'Empirical performance of alternative option pricing models', *Journal of Finance* **52**(5), 2003–2049.

Barrie & Hibbert (2011), 'A least squares Monte Carlo approach to liability proxy modeling and capital calculation', www.barrhibb.com. By Adam Koursaris.

Bates, D. S. (1996), 'Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options', *The Review of Financial Studies* **9**(1), 69–107.

Bauer, D. & Ha, H. (2015), A least-squares Monte Carlo approach to the calculation of capital requirements, *in* 'The World Risk and Insurance Economics Congress', Munich, Germany.

Bauer, D., Reuss, A. & Singer, D. (2012), 'On the calculation of the solvency capital requirement based on nested simulations', *The Journal of the International Actuarial Association* **42**(2), 453–499.

Bettels, C., Fabrega, J. & Weiß, C. (2014), 'Anwendung von Least Squares Monte Carlo (LSMC) im Solvency-II-Kontext - Teil 1', *Der Aktuar* **2**, 85–91.

Beutner, E., Pelsser, A. & Schweizer, J. (2016), 'Theory and validation of replicating portfolios in insurance risk management', *SSRN Electronic Journal* .
URL: *https://ssrn.com/abstract=2557368*

Black, F. & Karasinski, P. (1991), 'Bond and option pricing when short rates are lognormal', *Financial Analysts Journal* **47**(4), 52–59.

Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**(3), 637–654.

Boisbunon, A., Canu, S., Fourdrinier, D., Strawderman, W. & Wells, M. T. (2014), 'AIC, C_P and estimators of loss for elliptically symmetric distributions', https://arxiv.org.

Born, R. (2018), Künstliche Neuronale Netze im Risikomanagement, Master's thesis, Universität zu Köln, Germany.

Brace, A., Gątarek, D. & Musiela, M. (1997), 'The market model of interest rate dynamics', *Mathematical Finance* **7**(2), 127–147.

Breiman, L. (1994), 'Bagging predictors', Technical Report 421. Department of Statistics at University of California at Berkeley.

Breusch, T. S. & Pagan, A. R. (1979), 'A simple test for heteroscedasticity and random coefficient variation', *Econometrica* **47**(5), 1287–1294.

Buja, A., Hastie, T. & Tibshirani, R. (1989), 'Linear smoothers and additive models', *The Annals of Statistics* **17**(2), 453–510.

Burnham, K. P. & Anderson, D. R. (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2 edn, Springer-Verlag, New York, USA.

Cairns, A. J. G., Blake, D. & Dowd, K. (2006), 'A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration', *Journal of Risk and Insurance* **73**(4), 687–718.

Carriere, J. F. (1996), 'Valuation of the early-exercise price for options using simulations and nonparametric regression', *Insurance: Mathematics and Economics* **19**, 19–30.

Carroll, R. J. & Ruppert, D. (1988), *Transformation and Weighting in Regression*, Chapman & Hall, New York, USA.

Castellani, G., Fiore, U., Marino, Z., Passalacqua, L., Perla, F., Scognamiglio, S. & Zanetti, P. (2018), 'An investigation of machine learning approaches in the Solvency II valuation framework', *SSRN Electronic Journal* .
URL: *https://ssrn.com/abstract=3303296*

Cathcart, M. J. (2012), Monte Carlo Simulation Approaches to the Valuation and Risk Management of Unit-Linked Insurance Products with Guarantees, PhD thesis, Heriot Watt University, Scotland.

Chambers, J. & Hastie, T., eds (1992), *Statistical Models in S*, Wadsworth & Brooks/Cole, Boca Raton, London, New York, Washington D.C., chapter 'Generalized Linear Models'.

Committee on Finance Research of Society of Actuaries (2016), 'Economic scenario generators - a practical guide', Research Paper.

Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**(2), 385–408.

Craven, P. & Wahba, G. (1979), 'Smoothing noisy data with spline functions', *Numerische Mathematik* **31**, 377–403.

Dahlquist, G. & Björck, A. (1974), *Numerical Methods*, Prentice-Hall, Englewood Cliffs, USA.

Desmettre, S. & Korn, R. (2015), 10 computational challenges in finance, *in* C. De Schryver, ed., 'FPGA Based Accelerators for Financial Applications', Springer International Publishing AG, Cham, Switzerland, pp. 9–15.

Dobson, A. J. (2002), *An Introduction to Statistical Modelling*, 2 edn, Chapman & Hall/CRC, Boca Raton, London, New York, Washington D.C.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A. & Vapnik, V. (1997), Support vector regression machines, *in* 'Advances in Neural Information Processing Systems 9', MIT Press, Denver, USA, pp. 155–161.

Drăgulescu, A. A. & Yakovenko, V. M. (2002), 'Probability distribution of returns in the Heston model with stochastic volatility', *Quantitative Finance* **2**(6), 443–453.

Duchon, J. (1977), Splines minimizing rotation-invariant semi-norms in Solobev spaces, *in* W. Schempp & K. Zeller, eds, 'Constructive Theory of Functions of Several Variables', Springer, Berlin, Germany, pp. 85–100.

Dutang, C. (2017), 'Some explanations about the IWLS algorithm to fit generalized linear models', hal-01577698, HAL, France.

Efron, B. (1983), 'Estimating the error rate of a prediction rule: Improvement on cross-validation', *Journal of the American Statistical Association* **78**(382), 316–331.

Eilers, P. H. & Marx, B. D. (1996), 'Flexible smoothing with B-splines and penalties', *Statistical Science* **11**(2), 89–121.

EIOPA (2014), 'Technical specifications for the preparatory phase', Technical Paper. Implementation guidelines for the Solvency II framework.

EIOPA (2019), 'Methodological principles of insurance stress testing', Discussion Paper.

Eisenberg, B. & Sullivan, R. (2008), 'Why is the sum of independent normal random variables normal?', *Mathematics Magazine* **81**, 362–366.

European Central Bank (2020), 'Euro area yield curves'. Accessed on 15 June 2020.
URL: *https://www.ecb.europa.eu/stats/financial_markets_and_interest_rates*

European Commission (2014), 'Delegated Regulation (EU) supplementing Directive 2009/138/EC on the taking-up and pursuit of the business of insurance and reinsurance (Solvency II)', Regulation.

European Parliament & European Council (2009), 'Directive 2009/138/EC on the taking-up and pursuit of the business of insurance and reinsurance (Solvency II)', Directive.

Föllmer, H. & Schied, A. (2004), *Stochastic Finance: An Introduction in Discrete Time*, 2 edn, Walter de Gruyter, Berlin, New York.

Friedman, J. H. (1991), 'Multivariate adaptive regression splines (with discussion)', *The Annals of Statistics* **19**(1), 1–141.

Friedman, J. H. (1993), 'Fast MARS', Technical Report 110. Department of Statistics at Stanford University.

Friedman, J. H. & Silverman, B. W. (1989), 'Flexible parsimonious smoothing and additive modeling', *Technometrics* **31**(1), 3–21.

Friedman, J. H. & Stuetzle, W. (1981), 'Projection pursuit regression', *Journal of the American Statistical Association* **76**, 817–823.

Gay, D. M. (1990), 'Usage summary for selected optimization routines', Computing Science Technical Report 153. AT&T Bell Laboratories, Murray Hill.

Glasserman, P. (2004), *Monte Carlo Methods in Financial Engineering (Stochastic Modelling and Applied Probability)*, Springer Science+Business Media, New York, USA.

Glejser, H. (1969), 'A new test for heteroskedasticity', *Journal of the American Statistical Association* **64**(325), 316–323.

Gordy, M. B. & Juneja, S. (2010), 'Nested simulations in portfolio risk measurement', *Management Science* **56**, 1833–1848.

Green, P. J. (1984), 'Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives', *Journal of the Royal Statistical Society, Series B* **46**(2), 149–192.

Greene, W. H. (2002), *Econometric Analysis*, 5 edn, Prentice Hall, Upper Saddle River, USA.

Gürtler, N. (2011), Der MCEV in der Lebens- und Schadenversicherung geeignet für die Unternehmenssteuerung oder nicht?, *in* M. Heep-Altiner & M. Berg, eds, 'Proceedings zum 1. FaRis & DAV Symposium', FaRis & DAV Symposium, Cologne, Germany, pp. 7–20.

Hartmann, S. (2015), Verallgemeinerte lineare Modelle im Kontext des Least Squares Monte Carlo Verfahrens, Master's thesis, Katholische Universität Eichstätt-Ingolstadt, Germany.

Harvey, A. C. (1976), 'Estimating regression models with multiplicative heteroscedasticity', *Econometrica* **44**(3), 461–465.

Hastie, T. & Tibshirani, R. (1986), 'Generalized additive models', *Statistical Science* **1**(3), 297–318.

Hastie, T. & Tibshirani, R. (1990), *Generalized Additive Models*, Chapman & Hall, London, UK.

Hastie, T., Tibshirani, R. & Friedman, J. H. (2017), *The Elements of Statistical Learning*, 2 edn, Springer Series in Statistics, New York, USA.

Hayashi, F. (2000), *Econometrics*, Princeton University Press, Princeton, USA.

Hejazi, S. A. & Jackson, K. R. (2017), 'Efficient valuation of SCR via a neural network approach', *Journal of Computational and Applied Mathematics* **313**, 427–439.

Heston, S. L. (1993), 'A closed-form solution for options with stochastic volatility with applications to bond and currency options', *The Review of Financial Studies* **6**(2), 327–343.

Hocking, R. R. (1976), 'The analysis and selection of variables in linear regression', *Biometrics* **32**(1), 1–49.

Hoerl, A. E. & Kennard, R. W. (1970), 'Ridge regression: Biased estimation for nonorthogonal problems', *Technometrics* **12**(1), 55–67.

Huang, D. S. (1970), *Regression and Econometric Methods*, John Wiley & Sons, New York, USA.

Hudson, R. S. & Gregoriou, A. (2015), 'Calculating and comparing security returns is harder than you think: A comparison between logarithmic and simple returns', *International Review of Financial Analysis* **38**, 151–162.

Hurvich, C. M., Simonoff, J. S. & Tsai, C.-L. (1998), 'Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion', *Journal of the Royal Statistical Society, Series B* **60**(2), 271–293.

Investment Committee of DAV (2015), 'Proxy-Modelle für die Risikokapitalberechnung', Results Report. By working group Aggregation Techniques of Deutsche Aktuarvereinigung.

Investment Committee of DAV (2017), 'Anforderungen an einen ökonomischen Szenariogenerator', Fachgrundsatz. By working group Capital Market Modeling of Deutsche Aktuarvereinigung.

Jarrow, R. A., Lando, D. & Turnbull, S. M. (1997), 'A Markov model for the term structure of credit risk spreads', *The Review of Financial Studies* **10**(2), 481–523.

Kandasamy, K. & Yu, Y. (2016), Additive approximations in high dimensional nonparametric regression via the SALSA, *in* 'Proceedings of the 33rd International Conference on Machine Learning', Vol. 48, JMLR: W&CP, New York, USA, pp. 69–78.

Kazimov, N. (2018), Least Squares Monte Carlo modeling based on radial basis functions, Master's thesis, Universität Ulm, Germany.

Kochanski, M. (2010), 'Solvency capital requirement for German unit-linked insurance products', *German Risk and Insurance Review (GRIR)* **6**, 33–70.

Kopczyk, D. (2018), 'Proxy modeling in life insurance companies with the use of machine learning algorithms'. Working Paper.
URL: *https://ssrn.com/abstract=3396481*

Korn, R., Korn, E. & Kroisandt, G. (2010), *Monte Carlo Methods in Finance and Insurance*, CRC Press, Boca Raton, USA.

Krah, A.-S. (2015), Suitable information criteria and regression methods for the polynomial fitting process in the LSMC model, Master's thesis, Julius-Maximilians-Universität Würzburg, Germany.

Krah, A.-S., Nikolić, Z. & Korn, R. (2018), 'A least-squares Monte Carlo framework in proxy modeling of life insurance companies', *Risks* **6**(2), 62.

Krah, A.-S., Nikolić, Z. & Korn, R. (2020*a*), 'Machine learning in least-squares Monte Carlo proxy modeling of life insurance companies', *Risks* **8**(1), 21.

Krah, A.-S., Nikolić, Z. & Korn, R. (2020*b*), 'Least-squares Monte Carlo proxy modeling in life insurance: Neural networks', *Risks* **8**(4), 116.

Lee, R. D. & Carter, L. R. (1992), 'Modeling and forecasting U.S. mortality', *Journal of the American Statistical Association* **87**(419), 659–671.

Li, Q. & Racine, J. (2004), 'Cross-validated local linear nonparametric regression', *Statistica Sinica* **14**, 485–512.

Li, Q. & Racine, J. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, Princeton, USA.

Longstaff, F. A. & Schwartz, E. S. (2001), 'Valuing American options by simulation: A simple least-squares approach', *The Review of Financial Studies* **14**(1), 113–147.

Magnus, J. R. (1978), 'Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix', *Journal of Econometrics* **7**(3), 281–312.

Mai, J.-F. & Scherer, M. (2012), *Simulating Copulas*, World Scientific, Singapore.

Mallows, C. L. (1973), 'Some comments on C_P', *Technometrics* **15**(4), 661–675.

Marra, G. & Wood, S. N. (2012), 'Coverage properties of confidence intervals for generalized additive model components', *Scandinavian Journal of Statistics* **39**(1), 53–74.

Marx, B. D. & Eilers, P. H. (1998), 'Direct generalized additive modeling with penalized likelihood', *Computational Statistics & Data Analysis* **28**, 193–209.

McCullagh, P. & Nelder, J. A. (1989), *Generalized Linear Models*, 2 edn, Chapman & Hall, London, New York.

McLean, D. (2014), 'Orthogonality in proxy generator', Presentation, Moody's Analytics, Insurance-ERS. Legendre Polynomial / QR Decomposition Equivalence in Multiple Polynomial Regression.

Milborrow, S. (2018), *earth: Multivariate Adaptive Regression Splines.* Derived from mda:mars by Trevor Hastie and Rob Tibshirani. Uses Alan Miller's Fortran utilities with Thomas Lumley's leaps wrapper. R package version 4.6.3.
URL: *https://cran.r-project.org/web/packages/earth*

Milliman (2013), 'Least squares Monte Carlo for fast and robust capital projections', us.milliman.com. By Michael Leitschkis, Mario Hörig, Florian Ketterer, Christian Bettels.

Mourik, T. (2003), 'Market risk of insurance companies', Discussion Paper IAA Insurer Solvency Assessment Working Party.

Nadaraya, E. A. (1964), 'On estimating regression', *Theory of Probability and Its Applications* **9**(1), 141–142.

Natolski, J. & Werner, R. (2014), 'Mathematical analysis of different approaches for replicating portfolios', *European Actuarial Journal* **4**(2), 411–435.

Natolski, J. & Werner, R. (2016), 'Mathematical foundation of the replicating portfolio approach', *SSRN Electronic Journal* .
URL: *https://ssrn.com/abstract=2771254*

Nelder, J. A. & Wedderburn, R. W. M. (1972), 'Generalized linear models', *Journal of the Royal Statistical Society, Series A* **135**(3), 370–384.

Niederreiter, H. (1992), *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, USA.

Nikolić, Z., Jonen, C. & Zhu, C. (2017), 'Robust regression technique in LSMC proxy modeling', *Der Aktuar* **1**, 8–16.

Nychka, D. (1988), 'Bayesian confidence intervals for smoothing splines', *Journal of the American Statistical Association* **83**, 1134–1143.

Pindyck, R. S. & Rubinfeld, D. L. (1998), *Econometric Models and Economic Forecasts*, Irwin/McGraw-Hill, University of Michigan, USA.

R Core Team (2018), *stats: R statistical functions*, R Foundation for Statistical Computing, Vienna, Austria. R package version 3.2.0.
URL: *http://www.R-project.org*

R Documentation (2017), 'memory.size', https://www.rdocumentation.org.

Racine, J. S. & Hayfield, T. (2018), *np: Nonparametric Kernel Smoothing Methods for Mixed Data Types*. R package version 0.60-8.
URL: *https://cran.r-project.org/web/packages/np*

Reichenwallner, B. (2014), Modellwahlverfahren für proxy-modelle im kontext von solvency ii, Master's thesis, Paris-Lodron-Universität Salzburg, Austria.

Runge, C. (1901), 'Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten', *Zeitschrift für Mathematik und Physik* **46**, 224–243.

Schelthoff, T. (2019), Machine learning methods as alternatives to the least squares Monte Carlo model for calculating the solvency capital requirement of life and health insurance companies, Master's thesis, Universität zu Köln, Germany.

Schoenenwald, J. J. (2019), Modelli proxy per la determinazione dei requisiti di capitale secondo Solvency II, Master's thesis, Universitá degli Studi di Trieste, Italy.

Sell, R. (2019), Nicht-Parametrische Regression im Risikomanagement, Bachelor's thesis, Universität zu Köln, Germany.

Stone, M. (1974), 'Cross-validatory choice and assessment of statistical predictions', *Journal of the Royal Statistical Society, Series B* **36**(2), 111–147.

Suykens, J. A. & Vandewalle, J. (1999), 'Least squares support vector machine classifiers', *Neural Processing Letters* **9**(3), 293–300.

Takeuchi, K. (1976), 'Distributions of information statistics and criteria for adequacy of models', *Mathematical Science* **153**, 12–18. (in Japanese).

Teuguia, O. N., Ren, J. & Planchet, F. (2014), 'Internal model in life insurance: application of least squares monte carlo in risk assessment', Technical Report. Laboratoire de Sciences Actuarielle et Financière.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the Lasso', *Journal of the Royal Statistical Society, Series B* **58**(1), 267–288.

Tsitsiklis, J. N. & Van Roy, B. (2001), 'Regression methods for pricing complex american-style options', *IEEE Transactions on Neural Networks* **12**(4), 694–703.

Vasicek, O. (1977), 'An equilibrium characterization of the term structure', *Journal of Financial Economics* **5**, 177–188.

Watson, G. S. (1964), 'Smooth regression analysis', *Sankhyā: The Indian Journal of Statistics, Series A* **26**(4), 359–372.

Weiß, C. & Nikolić, Z. (2019), 'An aspect of optimal regression design for LSMC', *Monte Carlo Methods and Applications* **25**(4), 283–290.

Wood, S. N. (2000), 'Modelling and smoothing parameter estimation with multiple quadratic penalties', *Journal of the Royal Statistical Society, Series B* **62**(2), 413–428.

Wood, S. N. (2003), 'Thin plate regression splines', *Journal of the Royal Statistical Society, Series B* **65**(1), 95–114.

Wood, S. N. (2006), 'Generalized additive models', Lecture Notes, School of Mathematics, University of Bristol, U.K.

Wood, S. N. (2017), *Generalized Additive Models: An Introduction with R*, 2 edn, CRC Press, Boca Raton, USA.

Wood, S. N. (2018), *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation.* R package version 1.8-24.
URL: *https://cran.r-project.org/web/packages/mgcv*

Wood, S. N., Goude, Y. & Shaw, S. (2015), 'Generalized additive models for large data sets', *Journal of the Royal Statistical Society, Series C* **64**(1), 139–155.

Wood, S. N., Li, Z., Shaddick, G. & Augustin, N. H. (2017), 'Generalized additive models for gigadata: Modeling the U.K. black smoke network daily data', *Journal of the American Statistical Association* **112**(519), 1199–1210.

Wood, S. N., Pya, N. & Säfken, B. (2016), 'Smoothing parameter and model selection for general smooth models', *Journal of the American Statistical Association* **111**(516), 1548–1575.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. & Smith, G. M. (2009), *Mixed Effects Models and Extensions in Ecology with R*, Springer Science+Business Media, New York, USA, chapter 'GLM and GAM for Count Data', pp. 209–243.

Curriculum Vitae

Education

2018 – 2021	DOCTORAL STUDIES UNDER PROF. DR. RALF KORN Technical University of Kaiserslautern
2012 – 2015	MASTER OF SCIENCE IN ECONOMATHEMATICS Julius-Maximilians-University of Würzburg
2013 – 2014	TWO-SEMESTER EXCHANGE PROGRAM IN ECONOMICS University of Texas at Austin
2010 – 2012	BACHELOR OF SCIENCE IN ECONOMATHEMATICS Julius-Maximilians-University of Würzburg
2007 – 2010	ABITUR Freiherr-vom-Stein School Fulda

Work Experience

Since 2016	ACTUARY IN RISK MANAGEMENT Generali Deutschland AG, Cologne
Since 2019	LECTURER IN MATHEMATICS Hochschule Ruhr West, Mülheim an der Ruhr
2015	INTERN AND MASTERAND IN RISK MANAGEMENT Generali Deutschland AG, Cologne
2014	INTERN IN REVENUE MANAGEMENT Deutsche Lufthansa AG, Frankfurt am Main
2012 – 2013	STUDENT RESEARCH AND TEACHING ASSISTANT Julius-Maximilians-University of Würzburg

Lebenslauf

Studium

2018 – 2021	PROMOTION BEI PROF. DR. RALF KORN Technische Universität Kaiserslautern
2012 – 2015	MASTER OF SCIENCE IN WIRTSCHAFTSMATHEMATIK Julius-Maximilians-Universität Würzburg
2013 – 2014	ZWEISEMESTRIGES AUSTAUSCHPROGRAMM IN ECONOMICS University of Texas at Austin
2010 – 2012	BACHELOR OF SCIENCE IN WIRTSCHAFTSMATHEMATIK Julius-Maximilians-Universität Würzburg
2007 – 2010	ABITUR Freiherr-vom-Stein-Schule Fulda

Beruflicher Werdegang

Since 2016	AKTUARIN IM RISIKOMANAGEMENT Generali Deutschland AG, Köln
Since 2019	LEHRBEAUFTRAGTE FÜR MATHEMATIK Hochschule Ruhr West, Mülheim an der Ruhr
2015	PRAKTIKANTIN UND MASTERANDIN IM RISIKOMANAGEMENT Generali Deutschland AG, Köln
2014	PRAKTIKANTIN IM REVENUE MANAGEMENT Deutsche Lufthansa AG, Frankfurt am Main
2012 – 2013	STUDENTISCHE HILFSKRAFT UND TUTORIN Julius-Maximilians-Universität Würzburg

Publications

Part I	Krah, A.-S., Nikolić, Z. & Korn, R. (2018), 'A Least-Squares Monte Carlo Framework in Proxy Modeling of Life Insurance Companies', *Risks* **6**(2), 62. **URL:** https://www.mdpi.com/2227-9091/6/2/62
Part II	Krah, A.-S., Nikolić, Z. & Korn, R. (2020), 'Machine Learning in Least-Squares Monte Carlo Proxy Modeling of Life Insurance Companies', *Risks* **8**(1), 21. **URL:** https://www.mdpi.com/2227-9091/8/1/21
Tables 5–8	Krah, A.-S., Nikolić, Z. & Korn, R. (2020), 'Least-Squares Monte Carlo Proxy Modeling in Life Insurance: Neural Networks', *Risks* **8**(4), 116. **URL:** https://www.mdpi.com/2227-9091/8/4/116

Appendix

k	r_k^1	r_k^2	r_k^3	r_k^4	r_k^5	r_k^6	r_k^7	r_k^8	r_k^9	r_k^{10}	r_k^{11}	r_k^{12}	r_k^{13}	r_k^{14}	$\widehat{\beta}_k^{(N)}$	AIC	MSE 1	MSE 2
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8,048.89	405,920	279,972.87	41,918.48
1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	5,796.16	347,122	22,744.22	32,479.31
2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	80.08	342,819	9,863.42	5,112.25
3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	−28.03	337,760	4,048.53	4,413.39
4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	26.61	336,373	3,003.24	4,339.50
5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	108.73	335,246	2,534.24	4,329.35
6	0	0	0	0	0	0	0	0	0	0	0	1	0	0	22.68	334,903	2,334.00	4,314.37
7	0	0	0	1	0	0	0	1	0	0	0	0	0	0	−614.79	334,579	2,257.41	4,308.28
8	0	0	0	2	0	0	0	0	0	0	0	0	0	0	297.50	334,300	1,341.89	1,970.71
9	0	0	0	0	0	0	0	0	0	0	1	0	0	0	32.93	334,044	1,376.76	1,945.95
10	0	0	0	0	0	0	1	0	0	0	0	0	0	0	71.60	333,840	1,084.23	1,501.77
11	1	0	0	0	0	0	0	1	0	0	0	0	0	0	57.32	333,769	1,134.52	1,843.88
12	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1,636.91	333,524	638.87	748.50
13	0	1	0	0	0	0	0	1	0	0	0	0	0	0	−58.12	333,420	637.63	700.10
14	0	1	0	1	0	0	0	0	0	0	0	0	0	0	15.67	333,368	645.54	699.57
15	0	0	0	0	0	0	0	0	0	1	0	0	0	0	29.69	333,317	327.84	172.71
16	0	0	0	1	0	0	0	0	0	0	0	1	0	0	42.48	333,266	313.90	172.91
17	0	0	0	0	0	0	0	0	0	0	0	0	1	0	13.64	333,223	321.78	169.47
18	0	0	0	0	0	0	0	0	0	0	0	0	0	1	6.81	333,182	280.15	167.30
19	1	0	0	1	0	0	0	0	0	0	0	0	0	0	−5.66	333,144	292.49	166.74
20	0	0	1	0	0	0	0	0	0	0	0	0	0	0	−9.91	333,118	285.39	167.79
21	0	0	0	2	0	0	0	1	0	0	0	0	0	0	−440.05	333,094	259.30	178.81
22	0	0	0	0	0	1	0	1	0	0	0	0	0	0	36.14	333,084	268.45	169.96
23	1	0	0	1	0	0	0	1	0	0	0	0	0	0	−35.76	333,074	256.88	166.85
24	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0.85	333,067	270.83	163.39
25	0	0	0	0	0	0	0	1	0	0	1	0	0	0	−82.22	333,059	264.46	163.59
26	1	0	0	0	0	0	0	0	0	0	1	0	0	0	−2.32	333,047	248.44	163.48
27	1	0	0	0	0	0	0	0	0	0	0	1	0	0	−2.21	333,040	254.61	163.62
28	0	1	0	0	0	0	0	0	0	0	1	0	0	0	3.69	333,033	251.67	163.51
29	0	1	0	1	0	0	0	1	0	0	0	0	0	0	−48.47	333,027	259.06	163.76
30	0	1	0	0	0	0	0	2	0	0	0	0	0	0	126.06	333,021	255.68	162.53
31	0	0	0	1	0	1	0	0	0	0	0	0	0	0	−20.94	333,015	247.72	169.26
32	2	0	0	0	0	0	0	0	0	0	0	0	0	0	−0.30	333,011	282.84	264.54
33	2	0	0	1	0	0	0	0	0	0	0	0	0	0	−0.98	333,001	290.84	273.88
34	0	0	0	2	0	0	0	0	0	0	0	1	0	0	51.52	332,997	285.59	270.99
35	1	0	0	0	0	0	0	2	0	0	0	0	0	0	−93.15	332,992	299.74	269.59
36	0	0	0	1	0	0	0	0	0	0	0	0	0	1	−3.99	332,987	294.01	269.92
37	0	0	0	0	0	0	0	0	0	0	0	2	0	0	9.48	332,984	244.46	193.39
38	0	0	0	0	0	0	0	0	0	0	0	3	0	0	17.52	332,981	246.77	193.69
39	0	0	0	0	0	0	0	1	0	0	0	1	0	0	−42.29	332,978	252.84	193.19
40	0	0	0	3	0	0	0	0	0	0	0	0	0	0	76.70	332,976	239.20	186.54
41	0	0	0	4	0	0	0	0	0	0	0	0	0	0	−279.08	332,972	203.89	133.62
42	0	1	0	0	0	0	0	0	0	0	0	1	0	0	2.57	332,970	208.28	133.57
43	0	0	0	0	0	2	0	0	0	0	0	0	0	0	−0.62	332,968	218.11	146.76
44	0	0	0	1	0	2	0	0	0	0	0	0	0	0	4.71	332,965	210.39	148.32
45	0	0	0	0	0	0	0	0	0	1	0	0	0	1	13.36	332,964	208.86	148.27
46	0	0	0	1	0	0	0	0	0	0	1	0	0	0	11.10	332,962	198.29	148.32
47	0	0	2	0	0	0	0	0	0	0	0	0	0	0	−5.85	332,961	226.05	193.90
48	0	0	1	0	0	0	0	0	0	0	0	0	0	1	5.69	332,960	222.46	193.81
49	0	0	0	0	0	0	0	1	0	1	0	0	0	0	−56.01	332,960	207.33	198.33
50	0	0	0	0	0	0	1	1	0	0	0	0	0	0	70.44	332,959	209.11	197.42
51	0	0	0	0	0	0	0	3	0	0	0	0	0	0	1,390.37	332,958	217.77	203.07
52	1	0	0	1	0	1	0	0	0	0	0	0	0	0	−1.06	332,958	219.21	203.00
53	0	0	0	2	0	1	0	0	0	0	0	0	0	0	−64.50	332,958	192.10	159.88
54	0	0	0	2	0	2	0	0	0	0	0	0	0	0	17.64	332,953	165.97	143.94
55	0	0	0	1	0	0	0	0	0	1	0	0	0	0	17.25	332,953	169.81	137.14
56	0	0	0	1	0	0	0	0	0	1	0	0	0	1	37.24	332,952	172.65	137.00
57	0	1	0	0	0	0	0	0	0	0	0	0	1	0	2.00	332,952	172.90	137.59
58	0	0	0	0	0	0	0	1	0	0	0	0	1	0	−26.38	332,951	171.80	137.55
59	0	0	0	0	0	1	1	0	0	0	0	0	0	0	6.21	332,951	182.94	149.27
60	0	0	0	1	0	0	0	1	0	0	0	1	0	0	−66.58	332,951	182.17	148.10

Table A1: Construction sequence of the proxy function of BEL in the adaptive algorithm with the final coefficients. Furthermore, AIC and out-of-sample mean squared errors after each iteration.

k	r_k^1	r_k^2	r_k^3	r_k^4	r_k^5	r_k^6	r_k^7	r_k^8	r_k^9	r_k^{10}	r_k^{11}	r_k^{12}	r_k^{13}	r_k^{14}	r_k^{15}	$\widehat{\beta}_{\mathrm{OLS},k}$	AIC	v.mae	ns.mae	cr.mae
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14,718.24	437,251	4.557	3.231	4.027
1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	7,850.17	386,722	2.474	0.845	0.913
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−269.33	375,144	2.065	2.139	1.831
3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	145.21	366,567	1.656	0.444	0.496
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	−5.36	358,894	1.647	1.006	0.556
5	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	434.04	355,732	1.635	0.853	0.469
6	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1,753.40	354,318	1.679	0.956	0.374
7	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	19,145.78	349,759	1.234	0.491	0.628
8	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	33.33	347,796	0.999	0.340	0.594
9	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	868.25	346,444	0.912	0.357	0.602
10	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	30.59	345,045	0.839	0.389	0.650
11	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1.65	341,083	0.759	0.398	0.465
12	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	86.79	339,360	0.718	0.394	0.390
13	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	33.35	337,731	0.574	0.653	0.512
14	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	49.59	336,843	0.589	0.658	0.518
15	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	71.25	335,980	0.628	0.678	0.512
16	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	2,667.92	335,351	0.609	0.671	0.503
17	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	96.43	334,876	0.579	0.701	0.545
18	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	−6.31	334,413	0.593	0.720	0.531
19	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	−47.09	333,904	0.562	0.621	0.474
20	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	48.93	333,447	0.565	0.597	0.454
21	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	−3,412.68	333,116	0.553	0.543	0.407
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0.02	332,806	0.562	0.478	0.358
23	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	−0.12	332,547	0.550	0.450	0.381
24	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	43.77	332,294	0.545	0.468	0.378
25	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	118.94	332,042	0.530	0.464	0.362
26	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	−1,288.45	331,687	0.522	0.453	0.355
27	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	−44.72	331,405	0.525	0.444	0.343
28	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	−24,908.99	331,136	0.499	0.405	0.327
29	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	−86.88	330,562	0.504	0.348	0.268
30	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0.55	330,361	0.518	0.418	0.264
31	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	77.26	330,163	0.512	0.443	0.272
32	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	24.78	329,988	0.508	0.443	0.264
33	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	14.33	329,834	0.477	0.491	0.286
34	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	−0.39	329,688	0.477	0.500	0.290
35	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	28.36	329,550	0.476	0.502	0.291
36	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	−370.92	329,442	0.472	0.499	0.288
37	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	−17.90	329,147	0.462	0.505	0.301
38	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	8,574.53	329,043	0.472	0.518	0.300
39	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	−2.17	328,935	0.474	0.510	0.295
40	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	223.91	328,832	0.475	0.509	0.291
41	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	−1,801.73	328,733	0.455	0.445	0.248
42	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	−102.10	327,927	0.372	0.345	0.237
43	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0.70	327,858	0.368	0.353	0.235
44	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0.56	327,792	0.366	0.352	0.233
45	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	−3,034.32	327,729	0.365	0.356	0.228
46	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	−13,127.81	327,659	0.368	0.364	0.227
47	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	−17.54	327,603	0.368	0.366	0.226
48	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	−187.07	327,537	0.374	0.367	0.226
49	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	−300.54	327,483	0.369	0.367	0.230
50	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	−0.09	327,432	0.368	0.391	0.221
51	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	−60.84	327,382	0.359	0.390	0.228
52	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	−20.91	327,331	0.352	0.390	0.225
53	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0.00	327,287	0.346	0.377	0.206
54	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	−0.09	327,149	0.339	0.357	0.185
55	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1.44	327,105	0.315	0.321	0.173
56	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	−0.50	327,064	0.315	0.322	0.173
57	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	−6.06	327,025	0.322	0.317	0.175
58	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	−6,600.49	326,986	0.317	0.310	0.172
59	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	−407.57	326,823	0.308	0.302	0.183
60	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	3,378.82	326,787	0.306	0.301	0.183
61	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	205.28	326,733	0.304	0.299	0.183
62	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	−18.73	326,700	0.306	0.299	0.182
63	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	175.39	326,668	0.304	0.296	0.182
64	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	−0.20	326,638	0.304	0.298	0.181
65	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	2.45	326,610	0.301	0.296	0.183
66	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0.11	326,572	0.297	0.299	0.180
67	2	0	0	0	0	1	0	1	0	0	0	0	0	0	0	−13.02	326,545	0.292	0.286	0.169
68	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	93.69	326,519	0.292	0.287	0.172
69	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	891.58	326,478	0.294	0.282	0.173
70	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	−6.21	326,453	0.291	0.281	0.175
71	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	−112.56	326,428	0.289	0.281	0.176
72	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	−5.27	326,398	0.284	0.282	0.173
73	1	0	0	0	0	0	0	3	0	0	0	0	0	0	0	1,129.77	326,374	0.276	0.264	0.162
74	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	−0.29	326,352	0.272	0.266	0.158
75	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	−56.54	326,331	0.269	0.266	0.157

Table A2: OLS proxy function of BEL derived under 150-443 in the adaptive algorithm with the final coefficients. Furthermore, AIC scores and out-of-sample MAEs in % after each iteration.

k	r_k^1	r_k^2	r_k^3	r_k^4	r_k^5	r_k^6	r_k^7	r_k^8	r_k^9	r_k^{10}	r_k^{11}	r_k^{12}	r_k^{13}	r_k^{14}	r_k^{15}	$\widehat{\beta}_{\text{OLS},k}$	AIC	v.mae	ns.mae	cr.mae
76	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	−3.02	326,313	0.271	0.266	0.155
77	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	−10.59	326,295	0.264	0.270	0.151
78	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	−6.99	326,278	0.264	0.275	0.153
79	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	−2.25	326,261	0.252	0.285	0.154
80	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	−14.77	326,245	0.263	0.309	0.157
81	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1.95	326,229	0.267	0.306	0.155
82	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	2,248.54	326,214	0.266	0.307	0.156
83	0	0	0	0	0	0	0	3	0	0	0	0	0	0	1	−111.77	326,201	0.263	0.302	0.158
84	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	−0.11	326,187	0.262	0.302	0.157
85	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	−0.18	326,174	0.263	0.305	0.156
86	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	45.58	326,161	0.265	0.303	0.157
87	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0	−83,291.89	326,149	0.267	0.308	0.156
88	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	−56.20	326,137	0.267	0.308	0.156
89	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	−5.32	326,126	0.267	0.310	0.156
90	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	−10.87	326,116	0.267	0.313	0.158
91	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	−32.75	326,106	0.265	0.317	0.158
92	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2	−0.09	326,097	0.265	0.308	0.151
93	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	10.87	326,089	0.265	0.308	0.151
94	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	−48.93	326,081	0.264	0.306	0.148
95	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	69.57	326,073	0.256	0.288	0.141
96	0	0	0	1	0	0	0	3	0	0	0	0	0	0	0	−542,688.19	326,066	0.256	0.289	0.141
97	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	10.44	326,058	0.248	0.275	0.136
98	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	−1.08	326,051	0.248	0.276	0.136
99	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	419.05	326,045	0.249	0.275	0.136
100	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	12.80	326,038	0.250	0.276	0.136
101	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	−3.94	326,033	0.250	0.276	0.136
102	1	0	0	0	0	0	0	2	0	0	0	0	0	0	1	−10.12	326,027	0.248	0.281	0.138
103	2	0	0	0	0	0	0	1	0	0	0	0	0	0	1	−0.36	326,017	0.244	0.283	0.135
104	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	1.74	326,012	0.244	0.282	0.136
105	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0.00	326,006	0.242	0.268	0.132
106	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	−7.09	326,001	0.238	0.265	0.131
107	2	0	0	0	0	0	1	1	0	0	0	0	0	0	0	−109.46	325,982	0.238	0.263	0.129
108	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	−0.10	325,977	0.237	0.263	0.128
109	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	5.76	325,972	0.235	0.263	0.129
110	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	54.51	325,968	0.237	0.264	0.129
111	1	0	0	0	0	0	1	2	0	0	0	0	0	0	0	−1,386.73	325,963	0.235	0.264	0.129
112	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0.00	325,959	0.237	0.265	0.130
113	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0.11	325,955	0.235	0.265	0.130
114	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0.05	325,951	0.234	0.266	0.130
115	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	4.30	325,948	0.236	0.265	0.127
116	1	0	0	0	0	2	0	1	0	0	0	0	0	0	0	−19.81	325,944	0.237	0.262	0.126
117	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	−0.87	325,938	0.241	0.267	0.124
118	0	1	0	0	0	1	0	1	0	0	0	0	0	0	1	−0.36	325,935	0.241	0.267	0.124
119	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	−80.29	325,931	0.241	0.267	0.125
120	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	−6.95	325,928	0.241	0.267	0.124
121	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0.00	325,925	0.243	0.259	0.121
122	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	436.56	325,923	0.241	0.259	0.121
123	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	−0.03	325,920	0.243	0.263	0.121
124	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	2.99	325,918	0.242	0.263	0.120
125	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	−0.59	325,916	0.241	0.261	0.119
126	2	0	0	0	0	1	0	0	0	0	0	0	0	0	1	−0.02	325,908	0.247	0.265	0.124
127	0	0	0	0	0	1	0	2	0	0	0	0	0	0	1	−4.66	325,902	0.249	0.279	0.123
128	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	−8,179.68	325,900	0.249	0.280	0.124
129	0	0	0	0	0	1	0	3	0	0	0	0	0	0	0	691.40	325,898	0.249	0.280	0.123
130	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0.04	325,896	0.250	0.281	0.122
131	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	7.04	325,894	0.246	0.264	0.120
132	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	−27.72	325,892	0.247	0.264	0.119
133	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1.26	325,891	0.247	0.264	0.119
134	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	−2.67	325,889	0.249	0.265	0.118
135	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1.53	325,887	0.250	0.266	0.119
136	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	−0.07	325,885	0.250	0.265	0.120
137	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	40.44	325,884	0.251	0.265	0.119
138	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0	434.50	325,878	0.249	0.264	0.119
139	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	−5.99	325,877	0.248	0.264	0.119
140	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	14.64	325,873	0.246	0.263	0.120
141	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	−119.42	325,871	0.247	0.270	0.121
142	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3	0.00	325,870	0.248	0.271	0.121
143	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0.07	325,868	0.248	0.271	0.121
144	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	1.06	325,861	0.246	0.271	0.121
145	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	−0.74	325,859	0.247	0.271	0.121
146	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	−5.61	325,858	0.246	0.271	0.121
147	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	−0.08	325,857	0.247	0.270	0.121
148	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	−37.16	325,855	0.247	0.271	0.122
149	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0.41	325,851	0.247	0.271	0.122
150	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	−7,290.99	325,850	0.247	0.271	0.122

Table A2: Cont.

k	r_k^1	r_k^2	r_k^3	r_k^4	r_k^5	r_k^6	r_k^7	r_k^8	r_k^9	r_k^{10}	r_k^{11}	r_k^{12}	r_k^{13}	r_k^{14}	r_k^{15}	$\widehat{\beta}_{\mathbf{OLS},k}$	AIC	v.mae	ns.mae	cr.mae
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	745.35	391, 375	60.620	97.518	257.762
1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	5, 766.61	382, 610	50.402	99.306	256.789
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	272.75	367, 667	35.285	38.124	99.902
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5.46	359, 997	30.739	18.210	72.719
4	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	128.41	356, 705	30.119	25.088	29.357
5	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	−1, 750.72	355, 354	30.867	28.173	21.870
6	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	−19, 127.27	351, 002	22.942	14.948	44.668
7	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−33.25	349, 147	19.030	12.142	42.535
8	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	307.32	347, 777	18.221	10.928	35.420
9	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	−868.05	346, 423	16.662	11.527	35.941
10	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	−87.54	345, 025	15.987	10.264	31.461
11	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	−30.51	343, 570	14.858	11.187	34.502
12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	−1.66	339, 282	13.092	12.669	23.174
13	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	−33.33	337, 648	10.427	20.976	30.402
14	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	−70.63	336, 840	11.087	21.598	29.972
15	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	−41.37	336, 120	11.436	21.764	30.408
16	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	−2, 666.44	335, 495	11.088	21.543	29.890
17	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	−96.48	335, 022	10.545	22.479	32.334
18	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	6.30	334, 563	10.804	23.095	31.519
19	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	47.02	334, 058	10.232	19.913	28.128
20	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	−48.77	333, 610	10.292	19.163	26.995
21	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	3, 412.54	333, 281	10.083	17.438	24.190
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	−0.02	332, 970	10.246	15.328	21.326
23	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.12	332, 714	10.020	14.436	22.671
24	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	−120.68	332, 457	9.834	14.283	21.608
25	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1, 287.63	332, 108	9.725	13.969	21.273
26	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	44.71	331, 832	9.755	13.661	20.501
27	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	24, 899.66	331, 569	9.275	12.462	19.873
28	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	87.04	331, 004	9.292	10.757	17.022
29	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	−43.38	330, 742	9.171	11.183	16.023
30	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	−0.55	330, 543	9.444	13.409	15.766
31	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	−77.35	330, 345	9.324	14.207	16.192
32	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	−25.20	330, 161	9.246	14.203	15.692
33	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	−14.37	330, 007	8.672	15.764	16.964
34	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0.39	329, 859	8.682	16.031	17.223
35	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	−27.80	329, 728	8.665	16.110	17.264
36	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	−8, 757.49	329, 619	8.871	16.530	17.005
37	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	2.17	329, 513	8.937	16.276	16.790
38	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	369.16	329, 408	8.842	16.169	16.738
39	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	17.97	329, 109	8.637	16.387	17.527
40	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	−222.55	329, 008	8.656	16.359	17.271
41	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	1, 791.70	328, 910	8.297	14.282	14.748
42	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	101.23	328, 111	6.783	11.112	14.144
43	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	−0.70	328, 041	6.713	11.355	14.013
44	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	−0.57	327, 972	6.683	11.325	13.867
45	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3, 083.05	327, 905	6.654	11.456	13.595
46	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	12, 863.79	327, 837	6.700	11.721	13.500
47	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	17.78	327, 780	6.710	11.777	13.450
48	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	190.46	327, 711	6.824	11.818	13.468
49	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	300.76	327, 657	6.724	11.793	13.716
50	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0.09	327, 607	6.718	12.565	13.182
51	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	60.83	327, 557	6.543	12.533	13.558
52	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	20.91	327, 507	6.415	12.530	13.394
53	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0.00	327, 463	6.314	12.118	12.252
54	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0.08	327, 327	6.176	11.486	11.049
55	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	−1.46	327, 284	5.751	10.339	10.295
56	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0.50	327, 242	5.746	10.367	10.287
57	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	6.08	327, 203	5.871	10.211	10.450
58	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	6, 593.98	327, 165	5.780	9.973	10.274
59	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	406.73	327, 003	5.618	9.722	10.897
60	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	−3, 364.02	326, 968	5.581	9.671	10.904
61	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	−204.12	326, 914	5.542	9.626	10.921
62	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	18.90	326, 881	5.588	9.611	10.837
63	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	−175.17	326, 849	5.546	9.514	10.817
64	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0.21	326, 818	5.540	9.597	10.799
65	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	−2.44	326, 791	5.494	9.532	10.896
66	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	−0.11	326, 753	5.413	9.616	10.708
67	2	0	0	0	0	1	0	1	0	0	0	0	0	0	0	12.99	326, 726	5.317	9.215	10.046
68	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	−93.57	326, 700	5.329	9.255	10.231
69	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	−890.62	326, 660	5.355	9.090	10.326
70	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	113.04	326, 635	5.313	9.095	10.357
71	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	5.23	326, 605	5.231	9.101	10.164
72	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	6.20	326, 581	5.186	9.068	10.265
73	1	0	0	0	0	0	0	3	0	0	0	0	0	0	0	−1, 133.83	326, 556	5.034	8.488	9.647
74	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0.29	326, 534	4.950	8.580	9.374
75	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	56.56	326, 513	4.908	8.559	9.323

Table A3: OLS proxy function of AC derived under 150-443 in the adaptive algorithm with the final coefficients. Furthermore, AIC scores and out-of-sample MAEs in % after each iteration.

k	r_k^1	r_k^2	r_k^3	r_k^4	r_k^5	r_k^6	r_k^7	r_k^8	r_k^9	r_k^{10}	r_k^{11}	r_k^{12}	r_k^{13}	r_k^{14}	r_k^{15}	$\widehat{\beta}_{\mathrm{OLS},k}$	AIC	v.mae	ns.mae	cr.mae
76	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3.02	326,495	4.936	8.573	9.223
77	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	10.61	326,477	4.824	8.705	8.996
78	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	6.97	326,461	4.821	8.849	9.071
79	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	2.25	326,444	4.602	9.170	9.162
80	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	−1.94	326,429	4.688	9.069	8.997
81	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	−2,257.40	326,414	4.676	9.099	9.070
82	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	14.06	326,399	4.853	9.831	9.278
83	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0.11	326,385	4.844	9.851	9.203
84	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0.18	326,372	4.861	9.935	9.174
85	0	0	0	0	0	0	0	3	0	0	0	0	0	0	1	111.58	326,358	4.796	9.769	9.270
86	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	−45.11	326,346	4.826	9.724	9.330
87	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0	82,935.66	326,334	4.871	9.865	9.284
88	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	56.00	326,322	4.867	9.862	9.267
89	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	5.35	326,311	4.857	9.938	9.258
90	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	10.88	326,301	4.870	10.043	9.414
91	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	32.81	326,291	4.833	10.156	9.394
92	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	48.96	326,283	4.812	10.085	9.185
93	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	−10.90	326,274	4.801	10.083	9.210
94	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2	0.09	326,266	4.803	9.818	8.787
95	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	−69.45	326,258	4.659	9.250	8.413
96	0	0	0	1	0	0	0	3	0	0	0	0	0	0	0	543,840.26	326,251	4.663	9.269	8.393
97	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	−10.31	326,244	4.510	8.841	8.101
98	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1.07	326,237	4.523	8.847	8.091
99	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	−417.88	326,231	4.531	8.840	8.101
100	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	−12.92	326,224	4.546	8.847	8.081
101	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	3.94	326,219	4.558	8.866	8.072
102	1	0	0	0	0	0	0	2	0	0	0	0	0	0	1	10.10	326,213	4.513	9.012	8.203
103	2	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0.36	326,204	4.453	9.084	8.035
104	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	−1.74	326,198	4.445	9.063	8.070
105	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	7.09	326,193	4.383	8.967	8.008
106	2	0	0	0	0	0	1	1	0	0	0	0	0	0	0	109.50	326,174	4.371	8.899	7.889
107	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0.00	326,169	4.332	8.454	7.669
108	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	−5.85	326,164	4.290	8.456	7.689
109	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0.10	326,159	4.282	8.457	7.657
110	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	−54.88	326,154	4.313	8.463	7.689
111	1	0	0	0	0	0	1	2	0	0	0	0	0	0	0	1,380.74	326,150	4.291	8.489	7.700
112	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0.00	326,146	4.315	8.498	7.751
113	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	−0.11	326,142	4.287	8.501	7.736
114	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	−4.30	326,138	4.320	8.461	7.558
115	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	−0.05	326,135	4.299	8.514	7.566
116	1	0	0	0	0	2	0	1	0	0	0	0	0	0	0	20.09	326,131	4.320	8.417	7.498
117	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0.87	326,125	4.393	8.561	7.371
118	0	1	0	0	0	1	0	1	0	0	0	0	0	0	1	0.36	326,122	4.389	8.564	7.409
119	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	79.51	326,118	4.394	8.560	7.411
120	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0.00	326,115	4.430	8.304	7.187
121	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	6.91	326,113	4.420	8.305	7.176
122	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	−435.81	326,110	4.390	8.301	7.212
123	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0.03	326,107	4.419	8.450	7.206
124	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	−2.99	326,105	4.407	8.434	7.163
125	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0.59	326,103	4.394	8.366	7.095
126	2	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0.02	326,096	4.502	8.499	7.382
127	0	0	0	0	0	1	0	2	0	0	0	0	0	0	1	4.66	326,089	4.543	8.962	7.340
128	0	0	0	0	0	1	0	3	0	0	0	0	0	0	0	−692.59	326,088	4.537	8.961	7.248
129	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	8,097.70	326,086	4.539	8.995	7.316
130	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	−0.04	326,084	4.555	9.024	7.285
131	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	2.73	326,082	4.590	9.065	7.246
132	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	−1.53	326,080	4.612	9.097	7.280
133	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	−1.28	326,078	4.616	9.086	7.251
134	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0.07	326,077	4.607	9.055	7.287
135	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	−6.96	326,075	4.533	8.527	7.230
136	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	27.74	326,073	4.556	8.520	7.115
137	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	122.08	326,071	4.571	8.746	7.171
138	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	6.00	326,070	4.556	8.745	7.190
139	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	−14.50	326,066	4.533	8.699	7.199
140	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	−0.07	326,064	4.532	8.722	7.227
141	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	−1.05	326,057	4.507	8.733	7.250
142	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0.74	326,056	4.515	8.719	7.238
143	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	5.71	326,054	4.503	8.706	7.263
144	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	−39.87	326,053	4.499	8.715	7.244
145	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0	−431.71	326,047	4.470	8.669	7.215
146	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3	0.00	326,046	4.488	8.698	7.207
147	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0.08	326,045	4.494	8.694	7.223
148	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	37.33	326,043	4.496	8.703	7.236
149	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	−0.42	326,039	4.508	8.706	7.253
150	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	7,224.25	326,038	4.512	8.712	7.265

Table A3: Cont.

k	r_k^1	r_k^2	r_k^3	r_k^4	r_k^5	r_k^6	r_k^7	r_k^8	r_k^9	r_k^{10}	r_k^{11}	r_k^{12}	r_k^{13}	r_k^{14}	r_k^{15}	$\widehat{\beta}_{\mathbf{OLS},k}$	**AIC**	**v.mae**	**ns.mae**	**cr.mae**
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14,689.75	437,251	4.557	3.231	4.027
1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	7,990.98	386,722	2.474	0.845	0.913
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−274.24	375,144	2.065	2.139	1.831
3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	145.73	366,567	1.656	0.444	0.496
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	−5.11	358,894	1.647	1.006	0.556
5	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	416.79	355,732	1.635	0.853	0.469
6	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2,332.91	354,318	1.679	0.956	0.374
7	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	24,914.36	349,759	1.234	0.491	0.628
8	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49.42	347,796	0.999	0.340	0.594
9	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	859.49	346,444	0.912	0.357	0.602
10	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	29.50	345,045	0.839	0.389	0.650
11	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1.71	341,083	0.759	0.398	0.465
12	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	91.65	339,360	0.718	0.394	0.390
13	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	36.34	337,731	0.574	0.653	0.512
14	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	51.78	336,843	0.589	0.658	0.518
15	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	68.02	335,980	0.628	0.678	0.512
16	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	2,661.47	335,351	0.609	0.671	0.503
17	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	109.14	334,876	0.579	0.701	0.545
18	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	−12.63	334,413	0.593	0.720	0.531
19	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	−114.48	333,904	0.562	0.621	0.474
20	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	35.40	333,447	0.565	0.597	0.454
21	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	−4,570.15	333,116	0.553	0.543	0.407
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0.02	332,806	0.562	0.478	0.358
23	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	−0.26	332,547	0.550	0.450	0.381
24	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	47.17	332,294	0.545	0.468	0.378
25	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	123.47	332,042	0.530	0.464	0.362
26	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	−1,240.44	331,687	0.522	0.453	0.355
27	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	−43.82	331,405	0.525	0.444	0.343
28	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	−32,661.61	331,136	0.499	0.405	0.327
29	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	−140.90	330,562	0.504	0.348	0.268
30	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0.56	330,361	0.518	0.418	0.264
31	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	87.33	330,163	0.512	0.443	0.272
32	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	25.31	329,988	0.508	0.443	0.264
33	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	14.22	329,834	0.477	0.491	0.286
34	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	−0.44	329,688	0.477	0.500	0.290
35	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	26.88	329,550	0.476	0.502	0.291
36	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	−391.81	329,442	0.472	0.499	0.288
37	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	−18.58	329,147	0.462	0.505	0.301
38	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	11,959.32	329,043	0.472	0.518	0.300
39	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	−2.15	328,935	0.474	0.510	0.295
40	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	228.32	328,832	0.475	0.509	0.291
41	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	−1,938.37	328,733	0.455	0.445	0.248
42	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	−112.83	327,927	0.372	0.345	0.237
43	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0.71	327,858	0.368	0.353	0.235
44	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0.72	327,792	0.366	0.352	0.233
45	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	−4,230.29	327,729	0.365	0.356	0.228
46	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	−10,720.30	327,659	0.368	0.364	0.227
47	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	−18.39	327,603	0.368	0.366	0.226
48	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	−212.78	327,537	0.374	0.367	0.226
49	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	−177.64	327,483	0.369	0.367	0.230
50	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	−0.09	327,432	0.368	0.391	0.221
51	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	−57.40	327,382	0.359	0.390	0.228
52	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	−23.55	327,331	0.352	0.390	0.225
53	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0.00	327,287	0.346	0.377	0.206
54	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	−0.08	327,149	0.339	0.357	0.185
55	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1.15	327,105	0.315	0.321	0.173
56	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	−0.65	327,064	0.315	0.322	0.173
57	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	−4.41	327,025	0.322	0.317	0.175
58	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	−6,095.97	326,986	0.317	0.310	0.172
59	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	−332.88	326,823	0.308	0.302	0.183
60	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	3,624.77	326,787	0.306	0.301	0.183
61	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	191.46	326,733	0.304	0.299	0.183
62	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	−17.49	326,700	0.306	0.299	0.182
63	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	183.68	326,668	0.304	0.296	0.182
64	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	−0.20	326,638	0.304	0.298	0.181
65	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	2.55	326,610	0.301	0.296	0.183
66	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0.13	326,572	0.297	0.299	0.180
67	2	0	0	0	0	1	0	1	0	0	0	0	0	0	0	−29.57	326,545	0.292	0.286	0.169
68	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	95.55	326,519	0.292	0.287	0.172
69	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	922.48	326,478	0.294	0.282	0.173
70	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	−6.22	326,453	0.291	0.281	0.175
71	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	−134.95	326,428	0.289	0.281	0.176
72	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	−4.47	326,398	0.284	0.282	0.173
73	1	0	0	0	0	0	0	3	0	0	0	0	0	0	0	−26,186.72	326,374	0.276	0.264	0.162
74	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	−0.29	326,352	0.272	0.266	0.158
75	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	−58.01	326,331	0.269	0.266	0.157

Table A4: OLS proxy function of BEL derived under 300-886 in the adaptive algorithm with the final coefficients. Furthermore, AIC scores and out-of-sample MAEs in % after each iteration.

k	r_k^1	r_k^2	r_k^3	r_k^4	r_k^5	r_k^6	r_k^7	r_k^8	r_k^9	r_k^{10}	r_k^{11}	r_k^{12}	r_k^{13}	r_k^{14}	r_k^{15}	$\widehat{\beta}_{\mathrm{OLS},k}$	AIC	v.mae	ns.mae	cr.mae
76	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	−3.11	326,313	0.271	0.266	0.155
77	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	−2.10	326,295	0.264	0.270	0.151
78	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	−8.73	326,278	0.264	0.275	0.153
79	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	−1.93	326,261	0.252	0.285	0.154
80	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	−14.90	326,245	0.263	0.309	0.157
81	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	−1.22	326,229	0.267	0.306	0.155
82	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	3,341.29	326,214	0.266	0.307	0.156
83	0	0	0	0	0	0	0	3	0	0	0	0	0	0	1	−43.84	326,201	0.263	0.302	0.158
84	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	−0.12	326,187	0.262	0.302	0.157
85	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	−0.18	326,174	0.263	0.305	0.156
86	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	67.19	326,161	0.265	0.303	0.157
87	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0	−432,954.98	326,149	0.267	0.308	0.156
88	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	−34.58	326,137	0.267	0.308	0.156
89	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	−5.10	326,126	0.267	0.310	0.156
90	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	−10.78	326,116	0.267	0.313	0.158
91	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	−66.99	326,106	0.265	0.317	0.158
92	0	0	0	0	0	0	0	2	0	0	0	0	0	0	2	−0.09	326,097	0.265	0.308	0.151
93	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0.35	326,089	0.265	0.308	0.151
94	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	−93.83	326,081	0.264	0.306	0.148
95	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	70.45	326,073	0.256	0.288	0.141
96	0	0	0	1	0	0	0	3	0	0	0	0	0	0	0	−1,073,454.04	326,066	0.256	0.289	0.141
97	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	−21.59	326,058	0.248	0.275	0.136
98	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	−1.10	326,051	0.248	0.276	0.136
99	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	398.94	326,045	0.249	0.275	0.136
100	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	22.03	326,038	0.250	0.276	0.136
101	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	−4.12	326,033	0.250	0.276	0.136
102	1	0	0	0	0	0	0	2	0	0	0	0	0	0	1	1.30	326,027	0.248	0.281	0.138
103	2	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0.20	326,017	0.244	0.283	0.135
104	1	0	0	0	0	0	0	3	0	0	0	0	0	0	1	351.11	326,009	0.245	0.289	0.138
105	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	1.09	326,003	0.244	0.288	0.139
106	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0.00	325,997	0.242	0.274	0.136
107	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	−7.78	325,992	0.239	0.271	0.134
108	2	0	0	0	0	0	1	1	0	0	0	0	0	0	0	−126.28	325,973	0.238	0.269	0.132
109	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	−0.10	325,968	0.238	0.269	0.131
110	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	57.61	325,963	0.239	0.269	0.132
111	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	9.91	325,959	0.237	0.269	0.132
112	1	0	0	0	0	0	1	2	0	0	0	0	0	0	0	−1,698.92	325,954	0.236	0.270	0.132
113	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	−0.01	325,950	0.237	0.270	0.133
114	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0.10	325,946	0.236	0.271	0.133
115	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0.05	325,942	0.234	0.272	0.132
116	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	5.00	325,939	0.236	0.271	0.129
117	1	0	0	0	0	2	0	1	0	0	0	0	0	0	0	−17.60	325,935	0.238	0.268	0.127
118	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	−0.79	325,929	0.242	0.273	0.128
119	0	1	0	0	0	1	0	1	0	0	0	0	0	0	1	−0.55	325,925	0.241	0.273	0.128
120	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	−119.81	325,922	0.242	0.273	0.129
121	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	−7.16	325,919	0.241	0.273	0.128
122	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0.00	325,916	0.243	0.265	0.124
123	0	0	0	0	0	0	0	2	0	0	1	0	0	0	0	497.02	325,914	0.241	0.265	0.125
124	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	−0.03	325,911	0.243	0.269	0.125
125	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	−0.58	325,909	0.242	0.267	0.123
126	2	0	0	0	0	1	0	0	0	0	0	0	0	0	1	−0.02	325,901	0.248	0.271	0.129
127	0	0	0	0	0	1	0	2	0	0	0	0	0	0	1	−4.48	325,895	0.251	0.286	0.129
128	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	2.93	325,893	0.250	0.285	0.128
129	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	−5,069.15	325,891	0.250	0.286	0.128
130	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0.03	325,889	0.251	0.287	0.127
131	0	0	0	0	0	1	0	3	0	0	0	0	0	0	0	2,631.07	325,887	0.251	0.287	0.125
132	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	30.03	325,885	0.246	0.270	0.124
133	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	−27.79	325,883	0.248	0.270	0.123
134	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	−2.68	325,881	0.249	0.271	0.122
135	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	2.18	325,879	0.251	0.272	0.123
136	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	−0.07	325,878	0.250	0.271	0.124
137	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	52.06	325,876	0.251	0.272	0.123
138	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0	507.79	325,870	0.250	0.270	0.123
139	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0.09	325,869	0.248	0.270	0.123
140	0	0	0	0	0	0	0	0	2	0	0	1	0	0	0	14.53	325,865	0.246	0.269	0.123
141	0	0	0	0	0	0	0	1	0	0	0	0	0	0	3	0.00	325,864	0.247	0.270	0.122
142	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1.48	325,862	0.247	0.269	0.121
143	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	−98.06	325,861	0.248	0.276	0.122
144	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	−0.68	325,859	0.248	0.276	0.122
145	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0.08	325,858	0.248	0.276	0.122
146	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	1.10	325,850	0.247	0.277	0.122
147	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	−5.64	325,849	0.247	0.276	0.123
148	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	−0.08	325,847	0.247	0.276	0.123
149	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	20.58	325,846	0.246	0.277	0.123
150	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	−60.89	325,841	0.242	0.274	0.123

Table A4: Cont.

k	r_k^1	r_k^2	r_k^3	r_k^4	r_k^5	r_k^6	r_k^7	r_k^8	r_k^9	r_k^{10}	r_k^{11}	r_k^{12}	r_k^{13}	r_k^{14}	r_k^{15}	$\widehat{\beta}_{\mathrm{OLS},k}$	AIC	v.mae	ns.mae	cr.mae
151	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	−26.95	325,840	0.242	0.275	0.123
152	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0.42	325,835	0.243	0.275	0.123
153	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	−10,592.62	325,834	0.243	0.275	0.123
154	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0.93	325,833	0.243	0.275	0.125
155	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	2.96	325,832	0.244	0.275	0.124
156	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	−3.87	325,830	0.244	0.275	0.125
157	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	−68.29	325,829	0.243	0.277	0.125
158	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	−9,773.54	325,828	0.243	0.278	0.125
159	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	120.51	325,822	0.242	0.278	0.125
160	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0.03	325,821	0.243	0.278	0.127
161	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	−19.68	325,820	0.243	0.278	0.127
162	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	−24.62	325,819	0.240	0.261	0.127
163	0	0	0	0	0	0	0	0	1	0	0	0	0	0	3	0.00	325,818	0.239	0.261	0.128
164	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	−5.28	325,817	0.239	0.262	0.128
165	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	2.36	325,816	0.240	0.262	0.129
166	1	1	0	0	0	1	0	0	0	0	0	0	0	0	1	−0.02	325,814	0.238	0.264	0.129
167	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	−5.06	325,813	0.238	0.264	0.129
168	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	20.18	325,812	0.238	0.263	0.129
169	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	−461.05	325,812	0.239	0.264	0.130
170	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	6.14	325,811	0.238	0.265	0.130
171	0	0	0	1	0	0	0	2	0	0	0	0	0	0	1	2,708.64	325,810	0.237	0.265	0.130
172	0	0	0	1	0	0	0	3	0	0	0	0	0	0	1	9,307.25	325,805	0.239	0.265	0.129
173	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	−0.17	325,805	0.238	0.265	0.129
174	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	5.94	325,804	0.238	0.264	0.128
175	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	−0.07	325,804	0.238	0.264	0.127
176	0	0	1	0	0	0	1	2	0	0	0	0	0	0	0	−1,367.33	325,803	0.238	0.264	0.128
177	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1,133.78	325,803	0.237	0.264	0.128
178	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	−1.86	325,802	0.237	0.264	0.128
179	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.99	325,802	0.241	0.274	0.131
180	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	−0.01	325,766	0.241	0.300	0.149
181	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	−0.68	325,744	0.248	0.335	0.172
182	3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	−70.02	325,727	0.245	0.326	0.157
183	2	0	0	0	0	0	0	2	0	0	0	0	0	0	0	−1,883.77	325,700	0.238	0.313	0.144
184	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	−1.21	325,672	0.231	0.327	0.173
185	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	−157,391.76	325,655	0.225	0.309	0.175
186	0	0	0	0	0	0	0	4	0	0	0	0	0	0	1	2,127.74	325,644	0.221	0.303	0.176
187	2	0	0	0	0	0	0	2	0	0	0	0	0	0	1	21.17	325,583	0.206	0.296	0.190
188	3	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0.62	325,524	0.198	0.268	0.164
189	0	0	0	1	0	0	0	4	0	0	0	0	0	0	0	5,216,336.05	325,515	0.199	0.270	0.166
190	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	−0.54	325,506	0.201	0.275	0.173
191	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.01	325,500	0.195	0.281	0.184
192	2	0	0	0	0	0	1	2	0	0	0	0	0	0	0	136.68	325,499	0.193	0.279	0.182
193	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	−526.83	325,498	0.194	0.280	0.182
194	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	−32.63	325,494	0.192	0.270	0.178
195	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	−2,791.14	325,492	0.190	0.261	0.176
196	2	0	0	0	0	0	2	0	0	0	0	0	0	0	0	11.06	325,491	0.191	0.265	0.178
197	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0.09	325,491	0.190	0.265	0.179
198	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	13.23	325,490	0.186	0.258	0.178
199	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	143.48	325,488	0.187	0.261	0.179
200	2	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0.46	325,488	0.186	0.262	0.181
201	2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0.98	325,487	0.185	0.262	0.181
202	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	8.97	325,487	0.185	0.263	0.180
203	0	0	0	1	0	0	0	4	0	0	0	0	0	0	1	−33,222.10	325,487	0.184	0.263	0.179
204	2	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0.01	325,487	0.184	0.264	0.180
205	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	−0.32	325,487	0.184	0.263	0.178
206	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.20	325,486	0.183	0.264	0.177
207	2	0	0	0	0	1	1	0	0	0	0	0	0	0	0	−2.44	325,486	0.185	0.265	0.179
208	3	0	0	0	0	1	1	0	0	0	0	0	0	0	0	−1.76	325,485	0.184	0.261	0.173
209	2	0	0	0	0	1	1	1	0	0	0	0	0	0	0	−12.48	325,482	0.183	0.260	0.173
210	2	0	0	0	0	2	0	1	0	0	0	0	0	0	0	3.93	325,482	0.184	0.258	0.170
211	0	0	0	0	0	2	0	3	0	0	0	0	0	0	0	−495.92	325,481	0.184	0.257	0.168
212	0	0	0	0	0	1	1	2	0	0	0	0	0	0	0	−434.12	325,481	0.185	0.260	0.169
213	0	0	0	0	0	1	1	3	0	0	0	0	0	0	0	−2,854.58	325,479	0.185	0.260	0.167
214	2	0	0	0	0	0	0	1	0	0	1	0	0	0	0	6.58	325,479	0.184	0.261	0.167
215	1	0	0	0	0	0	0	0	0	0	0	0	0	2	0	7.08	325,479	0.183	0.257	0.167
216	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	−20.06	325,479	0.184	0.257	0.167
217	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	11.90	325,468	0.186	0.257	0.166
218	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0.20	325,468	0.186	0.257	0.166
219	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	18.33	325,468	0.186	0.257	0.165
220	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	9.56	325,468	0.185	0.258	0.165
221	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	37.24	325,463	0.194	0.265	0.168
222	0	1	0	0	0	0	0	0	0	0	0	0	0	3	0	17.46	325,460	0.196	0.265	0.168
223	1	0	0	0	0	0	0	0	0	0	0	0	0	3	0	−5.47	325,460	0.194	0.266	0.166
224	1	0	0	0	0	0	0	0	0	0	0	0	0	4	0	−11.21	325,459	0.194	0.268	0.168

Table A4: Cont.

k	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
0	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	0.839	0.802	0	21.468	104	0.389	0.376	23	21.659	113	0.650	0.636	89	27.112	179
20	0.565	0.540	−10	16.780	82	0.597	0.577	−75	8.274	2	0.454	0.445	−40	10.083	38
30	0.518	0.496	1	17.501	100	0.418	0.404	−47	7.970	37	0.264	0.259	1	13.378	85
40	0.475	0.454	−10	16.888	98	0.509	0.492	−66	6.234	27	0.291	0.285	−26	10.497	68
50	0.368	0.352	−15	13.268	78	0.391	0.378	−50	6.060	29	0.221	0.217	−9	10.674	69
60	0.306	0.293	−17	10.760	62	0.301	0.290	−36	5.863	29	0.183	0.179	5	10.651	69
70	0.291	0.278	−18	10.451	60	0.281	0.272	−33	6.060	30	0.175	0.171	8	10.958	72
80	0.263	0.251	−23	9.389	54	0.309	0.298	−41	4.837	22	0.157	0.154	−4	8.945	59
90	0.267	0.256	−24	9.196	54	0.313	0.303	−42	4.689	22	0.158	0.155	−7	8.587	57
100	0.250	0.239	−18	9.152	53	0.276	0.266	−35	4.637	22	0.136	0.133	0	8.606	57
110	0.237	0.226	−18	8.494	48	0.264	0.255	−34	4.144	18	0.129	0.126	−2	7.634	50
120	0.241	0.230	−16	8.896	50	0.267	0.258	−34	4.153	18	0.124	0.122	−2	7.679	51
130	0.250	0.239	−18	9.839	57	0.281	0.272	−37	4.810	24	0.122	0.120	−1	8.900	59
140	0.246	0.235	−15	9.855	57	0.263	0.254	−33	4.809	24	0.120	0.117	1	8.822	58
150	0.247	0.237	−14	9.924	57	0.271	0.262	−35	4.612	22	0.122	0.120	−1	8.537	56

Table A5: Out-of-sample validation figures of the OLS proxy function of BEL under 150-443 after each tenth iteration. MAEs in %.

k	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
0	60.620	3.178	−296	100.000	−207	97.518	2.936	−453	100.000	−369	257.762	4.251	−653	100.000	−568
10	15.987	0.838	−1	29.161	−110	10.264	0.309	−6	32.492	−119	31.461	0.519	−67	31.704	−180
20	10.292	0.540	10	21.029	−82	19.163	0.577	75	12.240	−21	26.995	0.445	39	13.324	−57
30	9.444	0.495	−1	21.971	−100	13.409	0.404	47	15.583	−56	15.766	0.260	−1	18.759	−105
40	8.656	0.454	10	21.197	−98	16.359	0.492	67	12.740	−46	17.271	0.285	26	15.434	−87
50	6.718	0.352	15	16.655	−78	12.565	0.378	50	12.938	−47	13.182	0.217	9	15.666	−88
60	5.581	0.293	17	13.506	−62	9.671	0.291	36	12.985	−48	10.904	0.180	−5	15.640	−88
70	5.313	0.279	19	13.026	−59	9.095	0.274	34	13.289	−49	10.357	0.171	−8	15.975	−90
80	4.688	0.246	21	11.326	−51	9.069	0.273	36	11.131	−41	8.997	0.148	0	13.590	−77
90	4.870	0.255	24	11.525	−53	10.043	0.302	42	10.995	−41	9.414	0.155	7	13.285	−75
100	4.546	0.238	18	11.471	−53	8.847	0.266	35	11.041	−41	8.081	0.133	0	13.308	−76
110	4.313	0.226	18	10.650	−48	8.463	0.255	34	9.999	−37	7.689	0.127	2	12.181	−69
120	4.430	0.232	16	11.350	−51	8.304	0.250	33	10.596	−39	7.187	0.119	−1	12.763	−73
130	4.555	0.239	18	12.345	−57	9.024	0.272	37	11.491	−42	7.285	0.120	1	13.663	−78
140	4.532	0.238	15	12.470	−57	8.722	0.263	35	11.282	−42	7.227	0.119	0	13.448	−76
150	4.512	0.237	14	12.459	−57	8.712	0.262	35	11.136	−41	7.265	0.120	1	13.242	−75

Table A6: Out-of-sample validation figures of the OLS proxy function of AC under 150-443 after each tenth iteration. MAEs in %.

k	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
0	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	0.839	0.802	0	21.468	104	0.389	0.376	23	21.659	113	0.650	0.636	89	27.112	179
20	0.565	0.540	−10	16.780	82	0.597	0.577	−75	8.274	2	0.454	0.445	−40	10.083	38
30	0.518	0.496	1	17.501	100	0.418	0.404	−47	7.970	37	0.264	0.259	1	13.378	85
40	0.475	0.454	−10	16.888	98	0.509	0.492	−66	6.234	27	0.291	0.285	−26	10.497	68
50	0.368	0.352	−15	13.268	78	0.391	0.378	−50	6.060	29	0.221	0.217	−9	10.674	69
60	0.306	0.293	−17	10.760	62	0.301	0.290	−36	5.863	29	0.183	0.179	5	10.651	69
70	0.291	0.278	−18	10.451	60	0.281	0.272	−33	6.060	30	0.175	0.171	8	10.958	72
80	0.263	0.251	−23	9.389	54	0.309	0.298	−41	4.837	22	0.157	0.154	−4	8.945	59
90	0.267	0.256	−24	9.196	54	0.313	0.303	−42	4.689	22	0.158	0.155	−7	8.587	57
100	0.250	0.239	−18	9.152	53	0.276	0.266	−35	4.637	22	0.136	0.133	0	8.606	57
110	0.239	0.229	−18	9.132	52	0.269	0.260	−35	4.577	22	0.132	0.129	−1	8.358	55
120	0.242	0.231	−16	9.519	54	0.273	0.263	−35	4.569	21	0.129	0.126	−1	8.380	55
130	0.251	0.240	−18	10.506	61	0.287	0.277	−37	5.421	27	0.127	0.125	0	9.724	64
140	0.246	0.235	−15	10.530	61	0.269	0.260	−34	5.329	27	0.123	0.120	2	9.526	63
150	0.242	0.232	−14	10.556	61	0.274	0.265	−35	5.119	26	0.123	0.120	0	9.261	61
160	0.243	0.232	−15	10.483	60	0.278	0.268	−36	5.018	25	0.127	0.124	0	9.144	60
170	0.238	0.228	−13	10.140	58	0.265	0.256	−33	4.968	24	0.130	0.127	2	8.884	59
180	0.241	0.230	−12	10.128	57	0.300	0.290	−37	4.552	18	0.149	0.146	2	8.716	58
190	0.201	0.192	−13	6.458	32	0.275	0.266	−33	4.124	−2	0.173	0.169	−4	4.721	27
200	0.186	0.178	−9	6.111	29	0.262	0.254	−29	4.460	−4	0.181	0.177	3	4.920	27
210	0.184	0.176	−9	6.210	30	0.258	0.249	−28	4.337	−3	0.170	0.167	3	4.846	28
220	0.185	0.177	−8	6.433	32	0.258	0.250	−28	4.286	−3	0.165	0.161	3	4.850	28
224	0.194	0.186	−9	6.659	34	0.268	0.259	−30	4.200	−2	0.168	0.165	1	5.007	29

Table A7: Out-of-sample validation figures of the OLS proxy function of BEL under 300-886 after each tenth and the final iteration. MAEs in %.

k	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
150-443 figures based on validation values minus 1.96 times standard errors															
150	0.286	0.273	−30	9.878	57	0.330	0.319	−46	3.915	16	0.151	0.148	−13	7.473	49
150-443 figures based on validation values															
150	0.247	0.237	−14	9.924	57	0.271	0.262	−35	4.612	22	0.122	0.120	−1	8.537	56
150-443 figures based on validation values plus 1.96 times standard errors															
150	0.231	0.221	1	9.977	57	0.219	0.212	−24	5.473	28	0.130	0.127	11	9.591	64
300-886 figures based on validation values minus 1.96 times standard errors															
224	0.236	0.225	−24	6.757	34	0.325	0.314	−41	4.610	−8	0.191	0.187	−11	4.307	22
300-886 figures based on validation values															
224	0.194	0.186	−9	6.659	34	0.268	0.259	−30	4.200	−2	0.168	0.165	1	5.007	29
300-886 figures based on validation values plus 1.96 times standard errors															
224	0.184	0.177	7	6.625	35	0.218	0.211	−19	3.982	4	0.173	0.169	13	5.813	37

Table A8: Out-of-sample validation figures of the derived OLS proxy functions of BEL under 150-443 and 300-886 after the final iteration based on three different sets of validation value estimates. Thereby emerges the first set of validation value estimates from pointwise subtraction of 1.96 times the standard errors from the original set of validation values. The second set is the original set. The third set is the addition counterpart of the first set.

k	AIC	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
Normal with identity link																
0	437, 251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	345, 045	0.839	0.802	0	21.468	104	0.389	0.376	23	21.659	113	0.650	0.636	89	27.112	179
20	333, 447	0.565	0.540	−10	16.780	82	0.597	0.577	−75	8.274	2	0.454	0.445	−40	10.083	38
30	330, 361	0.518	0.496	1	17.501	100	0.418	0.404	−47	7.970	37	0.264	0.259	1	13.378	85
40	328, 832	0.475	0.454	−10	16.888	98	0.509	0.492	−66	6.234	27	0.291	0.285	−26	10.497	68
50	327, 432	0.368	0.352	−15	13.268	78	0.391	0.378	−50	6.060	29	0.221	0.217	−9	10.674	69
60	326, 787	0.306	0.293	−17	10.760	62	0.301	0.290	−36	5.863	29	0.183	0.179	5	10.651	69
70	326, 453	0.291	0.278	−18	10.451	60	0.281	0.272	−33	6.060	30	0.175	0.171	8	10.958	72
80	326, 245	0.263	0.251	−23	9.389	54	0.309	0.298	−41	4.837	22	0.157	0.154	−4	8.945	59
90	326, 116	0.267	0.256	−24	9.196	54	0.313	0.303	−42	4.689	22	0.158	0.155	−7	8.587	57
100	326, 038	0.250	0.239	−18	9.152	53	0.276	0.266	−35	4.637	22	0.136	0.133	0	8.606	57
110	325, 968	0.237	0.226	−18	8.494	48	0.264	0.255	−34	4.144	18	0.129	0.126	−2	7.634	50
120	325, 928	0.241	0.230	−16	8.896	50	0.267	0.258	−34	4.153	18	0.124	0.122	−2	7.679	51
130	325, 896	0.250	0.239	−18	9.839	57	0.281	0.272	−37	4.810	24	0.122	0.120	−1	8.900	59
140	325, 873	0.246	0.235	−15	9.855	57	0.263	0.254	−33	4.809	24	0.120	0.117	1	8.822	58
150	325, 850	0.247	0.237	−14	9.924	57	0.271	0.262	−35	4.612	22	0.122	0.120	−1	8.537	56
Normal with inverse link																
0	437, 251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	343, 426	1.036	0.990	1	33.705	192	0.650	0.628	−63	21.481	114	0.391	0.382	44	33.482	221
20	334, 985	0.689	0.659	−6	21.313	118	0.515	0.498	−62	10.319	49	0.324	0.317	−4	16.493	107
30	331, 426	0.512	0.490	−16	18.836	109	0.393	0.380	−45	12.277	65	0.248	0.243	15	18.960	125
40	328, 875	0.433	0.414	−5	14.354	82	0.317	0.306	−26	9.312	47	0.294	0.288	26	15.188	99
50	327, 877	0.383	0.366	−8	12.959	76	0.285	0.276	−24	8.961	46	0.271	0.265	25	14.592	95
60	327, 274	0.337	0.323	−16	12.572	73	0.328	0.316	−37	7.636	38	0.219	0.215	10	13.087	85
70	326, 875	0.290	0.277	−14	11.248	64	0.271	0.261	−32	6.233	31	0.156	0.153	6	10.588	70
80	326, 603	0.259	0.248	−16	9.976	58	0.287	0.278	−38	5.042	22	0.158	0.155	−8	8.014	52
90	326, 390	0.254	0.243	−20	8.462	47	0.392	0.379	−51	4.451	1	0.220	0.215	−17	5.676	36
100	326, 225	0.270	0.258	−21	8.884	49	0.393	0.379	−51	4.454	5	0.219	0.215	−12	6.732	44
110	326, 152	0.272	0.260	−20	8.558	47	0.375	0.363	−48	4.441	4	0.208	0.204	−10	6.545	42
120	326, 094	0.267	0.255	−19	8.418	47	0.380	0.367	−49	4.414	3	0.209	0.205	−12	6.194	40
130	326, 058	0.266	0.254	−19	8.638	48	0.379	0.367	−49	4.329	4	0.203	0.199	−11	6.362	41
140	325, 982	0.258	0.247	−17	8.353	45	0.363	0.351	−46	4.380	2	0.197	0.193	−10	6.059	38
150	325, 952	0.258	0.247	−16	8.468	45	0.353	0.341	−44	4.282	3	0.192	0.188	−8	6.088	39
Normal with log link																
0	437, 251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	342, 325	0.879	0.840	26	25.171	132	0.422	0.408	−17	15.628	74	0.530	0.519	52	22.034	143
20	334, 417	0.661	0.632	−5	22.474	125	0.532	0.514	−64	10.764	51	0.330	0.323	−3	17.317	112
30	330, 901	0.560	0.536	−3	21.780	126	0.474	0.458	−55	11.199	59	0.266	0.261	3	17.802	117
40	328, 444	0.411	0.393	−10	13.639	78	0.315	0.304	−29	8.610	44	0.264	0.258	19	14.162	92
50	327, 574	0.341	0.326	−16	12.936	75	0.334	0.323	−35	8.294	42	0.262	0.257	12	13.642	89
60	327, 029	0.315	0.302	−17	11.991	69	0.312	0.301	−36	7.024	36	0.192	0.188	10	12.465	82
70	326, 637	0.279	0.267	−16	10.620	61	0.266	0.257	−31	6.142	31	0.162	0.158	9	10.797	71
80	326, 449	0.266	0.254	−21	10.069	59	0.304	0.294	−40	5.195	25	0.153	0.149	−4	9.234	61
90	326, 287	0.273	0.261	−22	9.742	57	0.300	0.290	−40	5.082	25	0.141	0.138	−5	8.990	59
100	326, 082	0.269	0.257	−23	8.052	45	0.370	0.358	−48	4.094	6	0.210	0.205	−13	6.314	41
110	326, 021	0.258	0.247	−19	8.043	44	0.343	0.331	−43	4.102	5	0.198	0.193	−7	6.381	41
120	325, 950	0.252	0.241	−17	7.891	42	0.329	0.318	−41	4.086	3	0.191	0.187	−7	5.883	37
130	325, 881	0.251	0.240	−18	8.049	45	0.359	0.347	−46	4.238	2	0.194	0.190	−10	5.924	38
140	325, 849	0.245	0.234	−17	7.978	44	0.340	0.328	−43	4.045	4	0.183	0.179	−7	6.131	40
150	325, 823	0.240	0.229	−15	7.980	44	0.316	0.305	−38	4.014	6	0.170	0.167	−2	6.434	42

Table A9: AIC scores and out-of-sample validation figures of the normal GLMs of BEL with identity, inverse and log link functions under 150-443 after each tenth iteration. MAEs in %.

k	AIC	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
Gamma with identity link																
0	437,243	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	345,605	0.872	0.834	1	23.485	114	0.315	0.304	6	19.861	105	0.530	0.519	68	25.266	167
20	333,911	0.553	0.529	−12	16.265	79	0.599	0.579	−76	8.268	0	0.464	0.454	−43	9.895	34
30	330,707	0.503	0.481	0	17.404	99	0.425	0.411	−49	7.754	35	0.267	0.262	−2	12.959	82
40	328,589	0.376	0.359	−13	13.317	76	0.341	0.330	−39	7.187	35	0.238	0.233	6	12.341	80
50	327,668	0.348	0.333	−15	13.173	77	0.356	0.344	−44	6.656	34	0.227	0.222	−4	11.348	74
60	327,135	0.305	0.292	−16	11.190	65	0.304	0.294	−37	6.059	30	0.175	0.172	3	10.843	71
70	326,686	0.273	0.261	−15	9.730	55	0.257	0.249	−30	5.364	26	0.165	0.161	9	9.928	65
80	326,461	0.268	0.257	−21	9.471	54	0.287	0.277	−36	5.151	25	0.149	0.146	2	9.549	63
90	326,328	0.259	0.248	−23	8.889	52	0.304	0.293	−40	4.373	20	0.148	0.145	−6	8.255	55
100	326,246	0.238	0.227	−20	8.321	48	0.262	0.253	−34	4.279	19	0.137	0.134	−1	7.845	52
110	326,184	0.233	0.223	−18	8.045	45	0.255	0.246	−33	3.907	16	0.130	0.127	−1	7.182	47
120	326,135	0.228	0.218	−16	8.191	46	0.253	0.245	−33	3.696	15	0.129	0.126	−2	6.870	45
130	326,093	0.244	0.233	−17	9.530	55	0.272	0.263	−35	4.628	22	0.124	0.122	0	8.596	57
140	326,068	0.238	0.228	−17	9.416	54	0.271	0.261	−35	4.523	22	0.125	0.123	−1	8.371	55
150	326,041	0.236	0.226	−14	9.329	53	0.260	0.251	−33	4.321	20	0.121	0.118	1	8.206	54
Gamma with inverse link																
0	437,243	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	343,969	1.037	0.991	0	33.818	193	0.661	0.639	−64	21.601	115	0.397	0.389	44	33.752	223
20	335,495	0.679	0.649	−7	20.888	115	0.530	0.512	−65	9.637	43	0.335	0.328	−9	15.410	99
30	332,646	0.627	0.600	−9	26.098	152	0.621	0.600	−82	12.361	64	0.346	0.339	−24	18.470	122
40	329,192	0.409	0.391	−10	14.061	81	0.317	0.306	−27	9.719	50	0.289	0.283	23	15.405	101
50	328,114	0.339	0.324	−12	12.599	73	0.313	0.302	−30	8.084	40	0.271	0.265	15	13.146	85
60	327,513	0.328	0.313	−16	12.247	71	0.294	0.284	−29	8.341	43	0.240	0.235	18	13.902	91
70	327,115	0.285	0.272	−12	11.127	64	0.251	0.243	−28	6.463	33	0.166	0.162	11	10.915	72
80	326,795	0.252	0.241	−17	8.376	45	0.315	0.305	−39	4.069	9	0.196	0.192	−8	6.416	40
90	326,615	0.250	0.239	−20	8.113	45	0.384	0.371	−51	4.414	0	0.218	0.213	−16	5.478	34
100	326,445	0.263	0.252	−20	8.724	48	0.382	0.369	−49	4.410	5	0.211	0.206	−11	6.595	43
110	326,370	0.266	0.255	−19	8.251	45	0.369	0.357	−47	4.494	2	0.205	0.201	−9	6.288	40
120	326,310	0.258	0.247	−17	8.003	44	0.357	0.345	−45	4.435	2	0.196	0.192	−8	6.087	39
130	326,277	0.259	0.248	−17	8.331	47	0.357	0.344	−45	4.356	4	0.187	0.183	−7	6.509	42
140	326,246	0.262	0.250	−17	8.583	48	0.357	0.345	−45	4.304	5	0.183	0.179	−7	6.620	43
150	326,222	0.254	0.243	−15	8.410	46	0.327	0.316	−40	4.111	7	0.171	0.167	−3	6.722	44
Gamma with log link																
0	437,243	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
1	388,234	2.365	2.261	−4	67.494	277	0.773	0.747	22	54.214	287	1.193	1.168	170	65.932	435
10	342,942	0.870	0.832	21	24.998	131	0.440	0.425	−24	15.145	71	0.505	0.494	43	21.396	138
20	334,881	0.649	0.621	−5	19.899	110	0.519	0.501	−65	8.283	36	0.312	0.306	−11	14.105	90
30	331,227	0.544	0.520	−4	21.752	126	0.479	0.463	−57	11.010	58	0.262	0.257	0	17.458	115
40	328,727	0.374	0.357	−10	14.009	81	0.329	0.318	−33	8.553	43	0.268	0.263	15	13.990	91
50	327,806	0.328	0.313	−16	12.750	74	0.327	0.316	−33	8.325	42	0.272	0.266	14	13.779	90
60	327,270	0.302	0.289	−15	11.825	68	0.297	0.287	−33	7.147	37	0.197	0.193	14	12.637	83
70	326,866	0.264	0.253	−15	10.159	58	0.249	0.241	−28	6.071	31	0.165	0.162	12	10.693	70
80	326,669	0.255	0.244	−19	9.819	57	0.288	0.279	−37	5.085	24	0.146	0.143	−2	9.090	60
90	326,433	0.266	0.254	−23	8.891	51	0.327	0.316	−45	4.079	15	0.171	0.167	−12	7.353	48
100	326,302	0.265	0.253	−23	7.839	44	0.361	0.349	−47	4.030	5	0.205	0.201	−12	6.246	40
110	326,224	0.256	0.244	−18	8.139	45	0.335	0.324	−41	4.211	8	0.191	0.187	−3	7.043	46
120	326,147	0.250	0.239	−18	7.817	43	0.340	0.328	−43	4.122	4	0.188	0.184	−6	6.247	41
130	326,111	0.247	0.236	−17	7.750	43	0.341	0.329	−43	4.115	3	0.186	0.183	−7	6.060	39
140	326,050	0.247	0.236	−17	7.730	43	0.336	0.324	−42	4.073	4	0.179	0.176	−6	6.117	40
150	326,022	0.243	0.232	−15	7.820	43	0.323	0.312	−40	4.040	3	0.174	0.170	−4	6.010	39

Table A10: AIC scores and out-of-sample validation figures of the gamma GLMs of BEL with identity, inverse and log link functions under 150-443 after each tenth iteration. MAEs in %.

k	AIC	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
inverse gaussian with identity link																
0	437,338	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	346,132	0.871	0.833	1	23.559	115	0.314	0.304	7	20.269	107	0.534	0.523	70	25.673	169
20	334,430	0.549	0.524	−13	15.996	77	0.599	0.579	−77	8.273	−1	0.468	0.458	−44	9.809	32
30	331,453	0.488	0.467	−4	15.939	89	0.517	0.499	−67	6.532	11	0.413	0.405	−40	9.280	38
40	328,985	0.370	0.354	−13	13.279	76	0.338	0.327	−39	7.193	35	0.238	0.233	6	12.301	80
50	328,064	0.332	0.317	−15	12.727	74	0.338	0.327	−40	6.871	35	0.232	0.227	1	11.664	76
60	327,533	0.298	0.285	−17	10.994	64	0.304	0.294	−37	5.868	29	0.172	0.168	3	10.646	69
70	327,082	0.274	0.262	−15	9.387	53	0.243	0.235	−27	5.535	27	0.171	0.167	13	10.253	67
80	326,849	0.267	0.255	−20	9.426	54	0.278	0.268	−34	5.271	25	0.152	0.148	5	9.783	65
90	326,715	0.247	0.236	−21	8.546	49	0.275	0.266	−35	4.399	20	0.140	0.137	−1	8.302	55
100	326,630	0.236	0.225	−20	7.879	45	0.262	0.253	−34	3.979	16	0.140	0.137	−2	7.249	48
110	326,564	0.225	0.215	−17	7.728	43	0.243	0.235	−31	3.850	15	0.129	0.126	0	6.958	46
120	326,507	0.237	0.226	−18	8.776	50	0.270	0.260	−35	4.120	19	0.130	0.127	−3	7.710	51
130	326,475	0.240	0.230	−17	9.225	53	0.265	0.256	−34	4.516	21	0.123	0.120	0	8.400	55
140	326,447	0.241	0.230	−16	9.415	54	0.270	0.261	−35	4.543	21	0.124	0.122	−1	8.426	56
150	326,352	0.249	0.238	−17	9.375	54	0.337	0.326	−44	4.224	12	0.150	0.146	−4	7.930	52
Inverse gaussian with inverse link																
0	437,338	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	344,458	1.129	1.079	−25	35.685	202	1.138	1.099	−150	14.423	63	0.639	0.626	−63	22.713	149
20	336,004	0.682	0.652	−5	21.011	117	0.534	0.516	−67	8.866	41	0.321	0.314	−12	14.895	95
30	333,060	0.626	0.598	−10	24.463	142	0.623	0.602	−83	10.859	55	0.376	0.369	−31	16.233	107
40	329,632	0.412	0.394	−14	15.912	93	0.345	0.333	−29	12.096	64	0.318	0.311	28	18.446	121
50	328,515	0.335	0.320	−12	12.387	71	0.305	0.295	−29	8.122	40	0.276	0.270	18	13.333	86
60	327,916	0.321	0.307	−15	11.970	70	0.286	0.276	−27	8.385	44	0.247	0.241	20	13.973	91
70	327,543	0.278	0.266	−12	10.488	60	0.246	0.238	−28	6.106	31	0.164	0.161	9	10.331	67
80	327,196	0.249	0.238	−17	8.227	45	0.308	0.297	−38	4.037	9	0.193	0.189	−7	6.381	40
90	327,012	0.247	0.236	−19	8.016	44	0.376	0.363	−49	4.390	−1	0.212	0.207	−15	5.407	33
100	326,837	0.261	0.250	−20	8.469	46	0.375	0.363	−48	4.428	4	0.208	0.204	−10	6.569	43
110	326,762	0.262	0.250	−18	8.090	44	0.365	0.353	−46	4.505	2	0.201	0.197	−8	6.242	40
120	326,699	0.259	0.248	−18	8.106	45	0.367	0.355	−47	4.402	2	0.192	0.188	−9	6.082	39
130	326,667	0.259	0.247	−17	7.987	44	0.352	0.340	−44	4.303	2	0.187	0.183	−8	5.958	38
140	326,642	0.258	0.246	−16	8.243	46	0.340	0.328	−42	4.228	6	0.173	0.169	−5	6.602	43
150	326,617	0.253	0.242	−15	8.152	44	0.324	0.313	−39	4.148	5	0.172	0.169	−3	6.476	42
Inverse gaussian with log link																
0	437,338	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	343,530	0.866	0.828	19	24.925	131	0.450	0.435	−28	14.940	69	0.494	0.484	39	21.122	136
20	335,355	0.644	0.616	−5	19.653	109	0.526	0.509	−67	7.947	33	0.318	0.311	−14	13.490	85
30	331,675	0.536	0.512	−4	21.697	125	0.482	0.465	−58	10.885	57	0.262	0.256	−2	17.245	113
40	329,140	0.366	0.350	−10	13.913	80	0.325	0.314	−32	8.604	44	0.269	0.264	16	14.011	91
50	328,190	0.324	0.310	−16	12.640	73	0.319	0.308	−32	8.482	43	0.274	0.268	16	13.966	91
60	327,666	0.296	0.283	−15	11.626	67	0.290	0.280	−31	7.181	37	0.201	0.197	15	12.695	83
70	327,263	0.261	0.250	−15	9.948	57	0.244	0.236	−27	6.042	30	0.172	0.168	12	10.531	69
80	327,061	0.251	0.240	−18	9.746	56	0.284	0.275	−37	4.988	24	0.145	0.142	−1	8.964	59
90	326,825	0.263	0.251	−23	8.769	51	0.321	0.310	−44	4.059	15	0.168	0.165	−11	7.316	48
100	326,695	0.261	0.249	−22	7.727	43	0.352	0.340	−45	4.048	6	0.203	0.199	−10	6.341	41
110	326,598	0.239	0.229	−17	7.408	40	0.343	0.332	−43	4.444	−1	0.185	0.181	−7	5.572	35
120	326,530	0.249	0.238	−18	7.520	41	0.343	0.331	−43	4.247	1	0.191	0.187	−7	5.928	38
130	326,494	0.246	0.235	−17	7.602	42	0.337	0.326	−43	4.108	2	0.183	0.179	−6	5.964	39
140	326,471	0.246	0.235	−17	7.772	43	0.332	0.321	−42	4.068	4	0.177	0.173	−6	6.092	39
150	326,413	0.247	0.237	−15	7.716	42	0.324	0.313	−40	4.095	2	0.172	0.168	−4	5.892	38
Inverse gaussian with $\frac{1}{\mu^2}$ link																
0	437,338	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	344,467	0.985	0.941	−14	31.473	176	0.993	0.959	−130	12.573	46	0.561	0.549	−52	18.986	124
20	336,815	0.668	0.639	−7	21.404	122	0.591	0.571	−75	9.506	38	0.372	0.364	−22	14.521	91
30	331,792	0.478	0.457	−5	15.821	90	0.367	0.354	−28	10.573	53	0.373	0.365	33	17.496	114
40	330,089	0.421	0.403	−1	15.183	89	0.295	0.285	−19	10.660	56	0.316	0.309	34	16.657	109
50	329,020	0.376	0.359	−10	14.443	85	0.300	0.290	−21	11.439	60	0.320	0.313	34	17.553	115
60	328,452	0.330	0.316	−12	12.905	75	0.290	0.280	−24	9.196	48	0.273	0.267	25	14.952	98
70	327,925	0.316	0.302	−16	11.733	69	0.301	0.291	−35	7.090	35	0.200	0.195	6	11.701	76
80	327,639	0.262	0.250	−18	8.128	43	0.298	0.288	−35	4.425	11	0.208	0.203	−1	7.205	45
90	327,265	0.278	0.266	−22	8.311	46	0.355	0.343	−44	4.383	9	0.202	0.197	−7	7.090	46
100	327,148	0.288	0.275	−22	8.166	44	0.357	0.345	−44	4.408	8	0.207	0.203	−6	7.039	46
110	327,078	0.274	0.262	−20	7.943	43	0.354	0.342	−44	4.451	4	0.196	0.192	−7	6.434	41
120	326,920	0.269	0.257	−18	8.350	46	0.374	0.361	−47	4.579	3	0.198	0.193	−9	6.419	41
130	326,887	0.270	0.258	−18	8.437	47	0.360	0.348	−44	4.544	6	0.196	0.192	−4	7.151	46
140	326,807	0.267	0.255	−18	8.193	45	0.345	0.333	−43	4.318	5	0.188	0.184	−5	6.661	43
150	326,778	0.262	0.250	−16	8.258	44	0.332	0.321	−41	4.238	5	0.177	0.174	−3	6.518	42

Table A11: AIC scores and out-of-sample validation figures of the inverse gaussian GLMs of BEL with identity, inverse, log and $\frac{1}{\mu^2}$ link functions under 150-443 after each tenth iteration. MAEs in %.

k	AIC	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
Normal with identity link																
0	437,251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	345,045	0.839	0.802	0	21.468	104	0.389	0.376	23	21.659	113	0.650	0.636	89	27.112	179
20	333,447	0.565	0.540	−10	16.780	82	0.597	0.577	−75	8.274	2	0.454	0.445	−40	10.083	38
30	330,361	0.518	0.496	1	17.501	100	0.418	0.404	−47	7.970	37	0.264	0.259	1	13.378	85
40	328,832	0.475	0.454	−10	16.888	98	0.509	0.492	−66	6.234	27	0.291	0.285	−26	10.497	68
50	327,432	0.368	0.352	−15	13.268	78	0.391	0.378	−50	6.060	29	0.221	0.217	−9	10.674	69
60	326,787	0.306	0.293	−17	10.760	62	0.301	0.290	−36	5.863	29	0.183	0.179	5	10.651	69
70	326,453	0.291	0.278	−18	10.451	60	0.281	0.272	−33	6.060	30	0.175	0.171	8	10.958	72
80	326,245	0.263	0.251	−23	9.389	54	0.309	0.298	−41	4.837	22	0.157	0.154	−4	8.945	59
90	326,116	0.267	0.256	−24	9.196	54	0.313	0.303	−42	4.689	22	0.158	0.155	−7	8.587	57
100	326,038	0.250	0.239	−18	9.152	53	0.276	0.266	−35	4.637	22	0.136	0.133	0	8.606	57
110	325,963	0.239	0.229	−18	9.132	52	0.269	0.260	−35	4.577	22	0.132	0.129	−1	8.358	55
120	325,922	0.242	0.231	−16	9.519	54	0.273	0.263	−35	4.569	21	0.129	0.126	−1	8.380	55
130	325,889	0.251	0.240	−18	10.506	61	0.287	0.277	−37	5.421	27	0.127	0.125	0	9.724	64
140	325,865	0.246	0.235	−15	10.530	61	0.269	0.260	−34	5.329	27	0.123	0.120	2	9.526	63
150	325,841	0.242	0.232	−14	10.556	61	0.274	0.265	−35	5.119	26	0.123	0.120	0	9.261	61
160	325,821	0.243	0.232	−15	10.483	60	0.278	0.268	−36	5.018	25	0.127	0.124	0	9.144	60
170	325,811	0.238	0.228	−13	10.140	58	0.265	0.256	−33	4.968	24	0.130	0.127	2	8.884	59
180	325,766	0.241	0.230	−12	10.128	57	0.300	0.290	−37	4.552	18	0.149	0.146	2	8.716	58
190	325,506	0.201	0.192	−13	6.458	32	0.275	0.266	−33	4.124	−2	0.173	0.169	−4	4.721	27
200	325,488	0.186	0.178	−9	6.111	29	0.262	0.254	−29	4.460	−4	0.181	0.177	3	4.920	27
210	325,482	0.184	0.176	−9	6.210	30	0.258	0.249	−28	4.337	−3	0.170	0.167	3	4.846	28
220	325,468	0.185	0.177	−8	6.433	32	0.258	0.250	−28	4.286	−3	0.165	0.161	3	4.850	28
224	325,459	0.194	0.186	−9	6.659	34	0.268	0.259	−30	4.200	−2	0.168	0.165	1	5.007	29
Normal with inverse link																
0	437,251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	343,426	1.036	0.990	1	33.705	192	0.650	0.628	−63	21.481	114	0.391	0.382	44	33.482	221
20	334,985	0.689	0.659	−6	21.313	118	0.515	0.498	−62	10.319	49	0.324	0.317	−4	16.493	107
30	331,426	0.512	0.490	−16	18.836	109	0.393	0.380	−45	12.277	65	0.248	0.243	15	18.960	125
40	328,875	0.433	0.414	−5	14.354	82	0.317	0.306	−26	9.312	47	0.294	0.288	26	15.188	99
50	327,877	0.383	0.366	−8	12.959	76	0.285	0.276	−24	8.961	46	0.271	0.265	25	14.592	95
60	327,274	0.337	0.323	−16	12.572	73	0.328	0.316	−37	7.636	38	0.219	0.215	10	13.087	85
70	326,875	0.290	0.277	−14	11.248	64	0.271	0.261	−32	6.233	31	0.156	0.153	6	10.588	70
80	326,603	0.259	0.248	−16	9.976	58	0.287	0.278	−38	5.042	22	0.158	0.155	−8	8.014	52
90	326,390	0.254	0.243	−20	8.462	47	0.392	0.379	−51	4.451	1	0.220	0.215	−17	5.676	36
100	326,224	0.269	0.257	−21	9.365	53	0.403	0.389	−52	4.500	7	0.225	0.220	−12	7.174	47
110	326,135	0.266	0.254	−19	8.894	49	0.377	0.364	−49	4.334	5	0.205	0.201	−12	6.497	42
120	326,069	0.266	0.254	−19	8.564	48	0.381	0.368	−50	4.271	4	0.204	0.200	−14	6.102	39
130	326,033	0.265	0.253	−19	8.498	47	0.386	0.373	−50	4.445	2	0.212	0.207	−14	5.917	38
140	325,950	0.253	0.242	−17	8.151	44	0.358	0.346	−46	4.345	1	0.189	0.185	−11	5.598	35
150	325,924	0.255	0.244	−17	8.485	46	0.364	0.352	−46	4.288	3	0.192	0.188	−11	5.894	38
160	325,886	0.258	0.247	−15	8.842	48	0.349	0.337	−44	4.199	5	0.178	0.174	−8	6.359	41
170	325,869	0.249	0.238	−14	8.503	46	0.331	0.320	−40	4.254	5	0.174	0.171	−5	6.182	40
180	325,850	0.248	0.237	−12	8.505	45	0.312	0.302	−37	4.099	6	0.164	0.161	−3	6.095	40
190	325,820	0.238	0.228	−12	8.240	43	0.313	0.303	−37	4.137	4	0.169	0.166	−3	5.825	38
200	325,803	0.244	0.234	−13	8.458	45	0.320	0.309	−38	4.073	6	0.171	0.167	−4	6.132	40
210	325,800	0.241	0.231	−13	8.376	45	0.313	0.302	−36	4.059	6	0.171	0.167	−2	6.248	41
213	325,797	0.241	0.230	−12	8.325	44	0.310	0.299	−36	4.063	6	0.171	0.167	−1	6.284	41
Normal with log link																
0	437,251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	342,325	0.879	0.840	26	25.171	132	0.422	0.408	−17	15.628	74	0.530	0.519	52	22.034	143
20	334,417	0.661	0.632	−5	22.474	125	0.532	0.514	−64	10.764	51	0.330	0.323	−3	17.317	112
30	330,901	0.560	0.536	−3	21.780	126	0.474	0.458	−55	11.199	59	0.266	0.261	3	17.802	117
40	328,444	0.411	0.393	−10	13.639	78	0.315	0.304	−29	8.610	44	0.264	0.258	19	14.162	92
50	327,574	0.341	0.326	−16	12.936	75	0.334	0.323	−35	8.294	42	0.262	0.257	12	13.642	89
60	327,029	0.315	0.302	−17	11.991	69	0.312	0.301	−36	7.024	36	0.192	0.188	10	12.465	82
70	326,637	0.279	0.267	−16	10.620	61	0.266	0.257	−31	6.142	31	0.162	0.158	9	10.797	71
80	326,449	0.266	0.254	−21	10.069	59	0.304	0.294	−40	5.195	25	0.153	0.149	−4	9.234	61
90	326,287	0.273	0.261	−22	9.742	57	0.300	0.290	−40	5.082	25	0.141	0.138	−5	8.990	59
100	326,082	0.269	0.257	−23	8.052	45	0.370	0.358	−48	4.094	6	0.210	0.205	−13	6.314	41
110	326,021	0.258	0.247	−19	8.043	44	0.343	0.331	−43	4.102	5	0.198	0.193	−7	6.381	41
120	325,950	0.252	0.241	−17	7.891	42	0.329	0.318	−41	4.086	3	0.191	0.187	−7	5.883	37
130	325,743	0.208	0.199	−13	6.208	30	0.310	0.299	−38	4.994	−10	0.191	0.187	−8	4.273	21
140	325,693	0.211	0.202	−13	6.620	34	0.302	0.292	−36	4.522	−3	0.186	0.182	−3	5.037	30
150	325,665	0.210	0.200	−13	6.729	35	0.298	0.288	−36	4.385	−2	0.180	0.176	−3	5.168	31
160	325,626	0.214	0.205	−14	6.549	33	0.302	0.292	−36	4.410	−3	0.183	0.179	−4	5.076	30
170	325,610	0.214	0.204	−14	6.590	33	0.291	0.281	−35	4.273	−3	0.173	0.169	−2	5.028	30
180	325,584	0.214	0.204	−13	6.587	33	0.296	0.286	−35	4.386	−4	0.176	0.172	−2	4.973	29
190	325,575	0.212	0.203	−12	6.502	32	0.283	0.273	−33	4.363	−4	0.173	0.170	0	4.950	29
200	325,567	0.201	0.192	−9	6.272	30	0.264	0.255	−29	4.491	−4	0.171	0.168	3	4.863	27
210	325,553	0.205	0.196	−9	6.655	32	0.267	0.258	−29	4.398	−2	0.176	0.173	3	5.165	30
214	325,552	0.206	0.197	−10	6.640	32	0.267	0.258	−29	4.402	−2	0.177	0.173	3	5.180	30

Table A12: AIC scores and out-of-sample validation figures of the normal GLMs of BEL with identity, inverse and log link functions under 300-886 after each tenth and the final iteration. MAEs in %.

k	AIC	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
Gamma with identity link																
0	437,243	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	345,605	0.872	0.834	1	23.485	114	0.315	0.304	6	19.861	105	0.530	0.519	68	25.266	167
20	333,911	0.553	0.529	−12	16.265	79	0.599	0.579	−76	8.268	0	0.464	0.454	−43	9.895	34
30	330,707	0.503	0.481	0	17.404	99	0.425	0.411	−49	7.754	35	0.267	0.262	−2	12.959	82
40	328,589	0.376	0.359	−13	13.317	76	0.341	0.330	−39	7.187	35	0.238	0.233	6	12.341	80
50	327,668	0.348	0.333	−15	13.173	77	0.356	0.344	−44	6.656	34	0.227	0.222	−4	11.348	74
60	327,135	0.305	0.292	−16	11.190	65	0.304	0.294	−37	6.059	30	0.175	0.172	3	10.843	71
70	326,686	0.273	0.261	−15	9.730	55	0.257	0.249	−30	5.364	26	0.165	0.161	9	9.928	65
80	326,461	0.268	0.257	−21	9.471	54	0.287	0.277	−36	5.151	25	0.149	0.146	2	9.549	63
90	326,328	0.259	0.248	−23	8.889	52	0.304	0.293	−40	4.373	20	0.148	0.145	−6	8.255	55
100	326,244	0.240	0.229	−20	9.273	54	0.282	0.273	−37	4.759	22	0.144	0.141	−2	8.662	57
110	326,178	0.236	0.225	−18	8.837	51	0.262	0.254	−34	4.454	20	0.135	0.132	0	8.139	54
120	326,117	0.237	0.226	−18	9.668	56	0.275	0.266	−36	4.845	24	0.129	0.126	−1	8.799	58
130	326,084	0.245	0.235	−17	10.148	59	0.270	0.260	−35	5.236	26	0.122	0.120	1	9.375	62
140	326,058	0.243	0.232	−17	10.153	58	0.273	0.264	−35	5.092	25	0.125	0.122	−1	9.122	60
150	326,031	0.239	0.229	−14	10.130	58	0.263	0.254	−33	4.914	24	0.121	0.118	2	9.014	60
160	325,871	0.232	0.222	−15	7.898	44	0.317	0.307	−39	3.918	5	0.174	0.170	−4	6.237	40
170	325,729	0.199	0.190	−13	6.235	30	0.280	0.271	−34	4.288	−5	0.176	0.172	−2	4.684	27
180	325,718	0.201	0.192	−13	6.171	30	0.279	0.270	−34	4.253	−5	0.172	0.169	−2	4.623	27
190	325,703	0.197	0.189	−12	6.158	30	0.278	0.268	−33	4.269	−5	0.171	0.168	−3	4.521	26
200	325,697	0.194	0.185	−11	5.943	28	0.264	0.255	−30	4.416	−5	0.169	0.165	0	4.470	25
210	325,689	0.190	0.181	−10	5.992	28	0.261	0.252	−29	4.381	−5	0.169	0.165	1	4.534	25
212	325,689	0.189	0.180	−11	5.975	28	0.261	0.252	−29	4.384	−5	0.169	0.165	1	4.545	25
Gamma with inverse link																
0	437,243	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	343,969	1.037	0.991	0	33.818	193	0.661	0.639	−64	21.601	115	0.397	0.389	44	33.752	223
20	335,495	0.679	0.649	−7	20.888	115	0.530	0.512	−65	9.637	43	0.335	0.328	−9	15.410	99
30	332,646	0.627	0.600	−9	26.098	152	0.621	0.600	−82	12.361	64	0.346	0.339	−24	18.470	122
40	329,192	0.409	0.391	−10	14.061	81	0.317	0.306	−27	9.719	50	0.289	0.283	23	15.405	101
50	328,114	0.339	0.324	−12	12.599	73	0.313	0.302	−30	8.084	40	0.271	0.265	15	13.146	85
60	327,513	0.328	0.313	−16	12.247	71	0.294	0.284	−29	8.341	43	0.240	0.235	18	13.902	91
70	327,115	0.285	0.272	−12	11.127	64	0.251	0.243	−28	6.463	33	0.166	0.162	11	10.915	72
80	326,795	0.252	0.241	−17	8.376	45	0.315	0.305	−39	4.069	9	0.196	0.192	−8	6.416	40
90	326,615	0.250	0.239	−20	8.113	45	0.384	0.371	−51	4.414	0	0.218	0.213	−16	5.478	34
100	326,445	0.263	0.252	−20	9.213	52	0.387	0.374	−50	4.469	8	0.219	0.214	−10	7.316	48
110	326,355	0.272	0.260	−21	8.812	49	0.384	0.371	−50	4.313	5	0.209	0.205	−14	6.489	42
120	326,297	0.267	0.255	−20	8.378	46	0.377	0.365	−48	4.470	2	0.206	0.202	−11	6.140	39
130	326,248	0.259	0.248	−17	8.210	45	0.365	0.352	−46	4.437	1	0.200	0.196	−10	5.933	38
140	326,214	0.258	0.247	−17	8.212	45	0.355	0.343	−45	4.404	3	0.192	0.188	−9	6.077	39
150	326,190	0.260	0.248	−17	8.701	49	0.349	0.337	−44	4.217	7	0.180	0.176	−7	6.781	44
160	326,147	0.247	0.236	−15	8.556	47	0.329	0.317	−40	4.091	7	0.174	0.170	−4	6.643	43
170	326,070	0.247	0.236	−15	8.355	46	0.332	0.321	−41	4.077	5	0.173	0.169	−6	6.182	40
180	326,045	0.243	0.233	−14	8.143	43	0.307	0.297	−37	4.001	6	0.164	0.160	−3	6.107	40
190	326,026	0.236	0.225	−13	7.996	42	0.305	0.295	−36	4.039	5	0.165	0.161	−2	5.973	39
200	325,979	0.239	0.229	−12	8.320	45	0.284	0.274	−31	4.162	11	0.154	0.151	5	7.110	47
208	325,969	0.234	0.223	−11	8.162	44	0.288	0.278	−31	4.185	9	0.158	0.154	5	6.832	45
Gamma with log link																
0	437,243	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	342,942	0.870	0.832	21	24.998	131	0.440	0.425	−24	15.145	71	0.505	0.494	43	21.396	138
20	334,881	0.649	0.621	−5	19.899	110	0.519	0.501	−65	8.283	36	0.312	0.306	−11	14.105	90
30	331,227	0.544	0.520	−4	21.752	126	0.479	0.463	−57	11.010	58	0.262	0.257	0	17.458	115
40	328,727	0.374	0.357	−10	14.009	81	0.329	0.318	−33	8.553	43	0.268	0.263	15	13.990	91
50	327,806	0.328	0.313	−16	12.750	74	0.327	0.316	−33	8.325	42	0.272	0.266	14	13.779	90
60	327,270	0.302	0.289	−15	11.825	68	0.297	0.287	−33	7.147	37	0.197	0.193	14	12.637	83
70	326,866	0.264	0.253	−15	10.159	58	0.249	0.241	−28	6.071	31	0.165	0.162	12	10.693	70
80	326,669	0.255	0.244	−19	9.819	57	0.288	0.279	−37	5.085	24	0.146	0.143	−2	9.090	60
90	326,433	0.266	0.254	−23	8.891	51	0.327	0.316	−45	4.079	15	0.171	0.167	−12	7.353	48
100	326,302	0.265	0.253	−23	7.839	44	0.361	0.349	−47	4.030	5	0.205	0.201	−12	6.246	40
110	326,224	0.256	0.244	−18	8.139	45	0.335	0.324	−41	4.211	8	0.191	0.187	−3	7.043	46
120	326,015	0.220	0.210	−17	6.898	36	0.317	0.306	−40	4.411	−1	0.194	0.190	−7	5.364	33
130	325,973	0.216	0.207	−15	6.654	33	0.307	0.296	−37	4.544	−4	0.196	0.192	−4	5.114	30
140	325,919	0.212	0.203	−15	6.334	31	0.302	0.292	−37	4.556	−5	0.191	0.187	−4	4.883	28
150	325,878	0.215	0.205	−14	6.486	33	0.297	0.287	−36	4.375	−3	0.181	0.177	−3	4.968	29
160	325,858	0.216	0.206	−14	6.619	34	0.299	0.289	−35	4.442	−2	0.181	0.177	−1	5.275	32
170	325,826	0.213	0.203	−14	6.485	33	0.302	0.292	−36	4.464	−4	0.183	0.180	−3	5.109	30
180	325,816	0.213	0.204	−14	6.505	33	0.300	0.290	−36	4.468	−3	0.179	0.176	−1	5.238	31
190	325,797	0.210	0.201	−14	6.580	33	0.295	0.285	−35	4.406	−3	0.179	0.176	−2	5.157	31
200	325,783	0.208	0.199	−13	6.496	32	0.290	0.280	−34	4.421	−3	0.178	0.174	−1	5.140	30
210	325,777	0.200	0.191	−10	6.260	30	0.263	0.254	−28	4.471	−3	0.176	0.173	4	5.107	30
220	325,774	0.199	0.190	−10	6.248	30	0.264	0.255	−28	4.541	−3	0.179	0.175	4	5.085	29
226	325,767	0.198	0.189	−8	6.256	29	0.249	0.241	−24	4.532	−1	0.184	0.180	8	5.417	32

Table A13: AIC scores and out-of-sample validation figures of the gamma GLMs of BEL with identity, inverse and log link functions under 300-886 after each tenth and the final iteration. MAEs in %.

k	AIC	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
Inverse gaussian with identity link																
0	437,338	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	346,132	0.871	0.833	1	23.559	115	0.314	0.304	7	20.269	107	0.534	0.523	70	25.673	169
20	334,430	0.549	0.524	−13	15.996	77	0.599	0.579	−77	8.273	−1	0.468	0.458	−44	9.809	32
30	331,453	0.488	0.467	−4	15.939	89	0.517	0.499	−67	6.532	11	0.413	0.405	−40	9.280	38
40	328,985	0.370	0.354	−13	13.279	76	0.338	0.327	−39	7.193	35	0.238	0.233	6	12.301	80
50	328,064	0.332	0.317	−15	12.727	74	0.338	0.327	−40	6.871	35	0.232	0.227	1	11.664	76
60	327,533	0.298	0.285	−17	10.994	64	0.304	0.294	−37	5.868	29	0.172	0.168	3	10.646	69
70	327,082	0.274	0.262	−15	9.387	53	0.243	0.235	−27	5.535	27	0.171	0.167	13	10.253	67
80	326,849	0.267	0.255	−20	9.426	54	0.278	0.268	−34	5.271	25	0.152	0.148	5	9.783	65
90	326,715	0.247	0.236	−21	8.546	49	0.275	0.266	−35	4.399	20	0.140	0.137	−1	8.302	55
100	326,627	0.234	0.224	−20	8.454	49	0.266	0.257	−34	4.414	20	0.144	0.141	−1	8.023	53
110	326,557	0.225	0.215	−17	8.350	47	0.246	0.238	−31	4.337	19	0.132	0.129	2	7.841	52
120	326,505	0.233	0.223	−17	8.897	51	0.256	0.247	−33	4.428	21	0.125	0.123	0	8.106	54
130	326,465	0.243	0.232	−16	9.965	58	0.265	0.256	−34	5.126	26	0.122	0.120	1	9.216	61
140	326,442	0.244	0.233	−16	10.175	59	0.273	0.264	−35	5.079	25	0.125	0.122	0	9.098	60
150	326,357	0.252	0.241	−16	10.133	58	0.352	0.340	−45	4.601	15	0.169	0.166	−1	8.831	58
160	326,130	0.206	0.197	−15	6.294	31	0.293	0.283	−36	4.360	−5	0.187	0.183	−4	4.711	26
170	326,112	0.204	0.195	−15	6.173	30	0.289	0.279	−35	4.284	−5	0.179	0.175	−4	4.688	27
180	326,099	0.203	0.194	−14	6.130	30	0.283	0.273	−34	4.277	−5	0.177	0.173	−3	4.654	26
190	326,088	0.204	0.195	−14	6.143	30	0.282	0.272	−34	4.280	−5	0.178	0.174	−3	4.699	27
200	326,076	0.204	0.195	−14	6.172	30	0.286	0.276	−34	4.347	−4	0.184	0.180	−3	4.823	27
210	326,071	0.199	0.190	−12	6.140	30	0.273	0.264	−32	4.277	−4	0.183	0.179	0	4.868	28
217	326,069	0.191	0.183	−11	5.967	28	0.261	0.252	−29	4.364	−5	0.178	0.175	2	4.779	27
Inverse gaussian with inverse link																
0	437,338	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	344,458	1.129	1.079	−25	35.685	202	1.138	1.099	−150	14.423	63	0.639	0.626	−63	22.713	149
20	336,004	0.682	0.652	−5	21.011	117	0.534	0.516	−67	8.866	41	0.321	0.314	−12	14.895	95
30	333,060	0.626	0.598	−10	24.463	142	0.623	0.602	−83	10.859	55	0.376	0.369	−31	16.233	107
40	329,632	0.412	0.394	−14	15.912	93	0.345	0.333	−29	12.096	64	0.318	0.311	28	18.446	121
50	328,515	0.335	0.320	−12	12.387	71	0.305	0.295	−29	8.122	40	0.276	0.270	18	13.333	86
60	327,916	0.321	0.307	−15	11.970	70	0.286	0.276	−27	8.385	44	0.247	0.241	20	13.973	91
70	327,543	0.278	0.266	−12	10.488	60	0.246	0.238	−28	6.106	31	0.164	0.161	9	10.331	67
80	327,196	0.249	0.238	−17	8.227	45	0.308	0.297	−38	4.037	9	0.193	0.189	−7	6.381	40
90	327,012	0.247	0.236	−19	8.016	44	0.376	0.363	−49	4.390	−1	0.212	0.207	−15	5.407	33
100	326,836	0.261	0.250	−20	9.073	51	0.382	0.369	−49	4.438	8	0.215	0.211	−9	7.237	47
110	326,750	0.268	0.257	−21	8.679	47	0.386	0.373	−50	4.510	4	0.217	0.212	−12	6.490	42
120	326,674	0.263	0.251	−19	8.191	45	0.378	0.365	−49	4.499	1	0.207	0.203	−12	6.011	38
130	326,636	0.261	0.250	−18	8.380	46	0.373	0.360	−48	4.402	2	0.198	0.193	−12	5.985	38
140	326,607	0.258	0.247	−17	8.253	46	0.349	0.337	−44	4.289	4	0.185	0.181	−8	6.277	40
150	326,581	0.258	0.246	−17	8.437	47	0.350	0.338	−44	4.228	6	0.183	0.179	−7	6.505	42
160	326,538	0.246	0.235	−15	8.445	47	0.326	0.315	−40	4.077	7	0.173	0.169	−4	6.572	43
170	326,522	0.249	0.238	−15	8.148	45	0.322	0.311	−39	4.119	6	0.175	0.172	−2	6.603	43
180	326,468	0.245	0.234	−14	8.583	47	0.298	0.288	−34	4.303	13	0.162	0.159	4	7.724	51
190	326,455	0.243	0.233	−14	8.506	47	0.299	0.289	−34	4.290	13	0.163	0.160	4	7.641	50
200	326,399	0.231	0.221	−12	7.918	42	0.286	0.277	−31	4.208	9	0.158	0.155	6	6.856	45
210	326,365	0.233	0.223	−12	7.983	43	0.288	0.279	−31	4.208	9	0.159	0.155	5	6.765	45
219	326,363	0.233	0.223	−11	8.040	43	0.283	0.274	−31	4.130	9	0.153	0.150	5	6.786	45

Table A14: AIC scores and out-of-sample validation figures of the inverse gaussian GLMs of BEL with identity, inverse, log and $\frac{1}{\mu^2}$ link functions under 300-886 after each tenth and the final iteration. MAEs in %.

k	AIC	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
Inverse gaussian with log link																
0	437,338	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	343,530	0.866	0.828	19	24.925	131	0.450	0.435	−28	14.940	69	0.494	0.484	39	21.122	136
20	335,355	0.644	0.616	−5	19.653	109	0.526	0.509	−67	7.947	33	0.318	0.311	−14	13.490	85
30	331,675	0.536	0.512	−4	21.697	125	0.482	0.465	−58	10.885	57	0.262	0.256	−2	17.245	113
40	329,140	0.366	0.350	−10	13.913	80	0.325	0.314	−32	8.604	44	0.269	0.264	16	14.011	91
50	328,190	0.324	0.310	−16	12.640	73	0.319	0.308	−32	8.482	43	0.274	0.268	16	13.966	91
60	327,666	0.296	0.283	−15	11.626	67	0.290	0.280	−31	7.181	37	0.201	0.197	15	12.695	83
70	327,263	0.261	0.250	−15	9.948	57	0.244	0.236	−27	6.042	30	0.172	0.168	12	10.531	69
80	327,061	0.251	0.240	−18	9.746	56	0.284	0.275	−37	4.988	24	0.145	0.142	−1	8.964	59
90	326,825	0.263	0.251	−23	8.769	51	0.321	0.310	−44	4.059	15	0.168	0.165	−11	7.316	48
100	326,695	0.261	0.249	−22	7.727	43	0.352	0.340	−45	4.048	6	0.203	0.199	−10	6.341	41
110	326,589	0.240	0.230	−19	7.484	41	0.342	0.330	−44	4.124	1	0.192	0.188	−11	5.484	35
120	326,409	0.216	0.207	−16	6.397	32	0.299	0.289	−37	4.534	−2	0.195	0.191	−4	5.170	30
130	326,363	0.216	0.207	−15	6.314	31	0.308	0.298	−37	4.693	−6	0.201	0.196	−4	4.957	28
140	326,331	0.218	0.208	−15	6.537	33	0.303	0.292	−36	4.505	−3	0.195	0.191	−1	5.362	32
150	326,270	0.216	0.207	−14	6.457	32	0.302	0.291	−36	4.524	−4	0.189	0.185	−2	5.049	30
160	326,249	0.217	0.208	−14	6.596	34	0.298	0.288	−36	4.418	−2	0.182	0.178	−1	5.291	32
170	326,231	0.217	0.207	−15	6.492	32	0.296	0.286	−35	4.391	−3	0.179	0.175	−2	5.189	31
180	326,206	0.214	0.205	−15	6.426	32	0.302	0.291	−36	4.466	−4	0.179	0.175	−3	4.950	29
190	326,191	0.206	0.197	−13	6.472	33	0.288	0.279	−34	4.422	−3	0.173	0.170	0	5.149	31
200	326,176	0.208	0.199	−13	6.545	33	0.286	0.276	−33	4.430	−2	0.179	0.175	0	5.288	31
210	326,161	0.208	0.199	−13	6.501	33	0.286	0.276	−33	4.439	−2	0.184	0.180	1	5.318	32
220	326,153	0.202	0.193	−10	6.280	30	0.260	0.251	−27	4.455	−2	0.178	0.174	5	5.190	31
222	326,153	0.201	0.192	−10	6.291	30	0.261	0.252	−28	4.494	−3	0.180	0.177	5	5.176	30
Inverse gaussian with $\frac{1}{\mu^2}$ link																
0	437,338	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	344,467	0.985	0.941	−14	31.473	176	0.993	0.959	−130	12.573	46	0.561	0.549	−52	18.986	124
20	336,815	0.668	0.639	−7	21.404	122	0.591	0.571	−75	9.506	38	0.372	0.364	−22	14.521	91
30	331,792	0.478	0.457	−5	15.821	90	0.367	0.354	−28	10.573	53	0.373	0.365	33	17.496	114
40	330,089	0.421	0.403	−1	15.183	89	0.295	0.285	−19	10.660	56	0.316	0.309	34	16.657	109
50	329,020	0.376	0.359	−10	14.443	85	0.300	0.290	−21	11.439	60	0.320	0.313	34	17.553	115
60	328,452	0.330	0.316	−12	12.905	75	0.290	0.280	−24	9.196	48	0.273	0.267	25	14.952	98
70	327,925	0.316	0.302	−16	11.733	69	0.301	0.291	−35	7.090	35	0.200	0.195	6	11.701	76
80	327,639	0.262	0.250	−18	8.128	43	0.298	0.288	−35	4.425	11	0.208	0.203	−1	7.205	45
90	327,265	0.278	0.266	−22	8.311	46	0.355	0.343	−44	4.383	9	0.202	0.197	−7	7.090	46
100	327,148	0.288	0.275	−22	8.166	44	0.357	0.345	−44	4.408	8	0.207	0.203	−6	7.039	46
110	327,077	0.275	0.262	−20	7.965	42	0.366	0.353	−45	4.676	2	0.207	0.202	−7	6.410	40
120	326,916	0.274	0.262	−18	8.313	45	0.393	0.380	−47	5.133	1	0.228	0.223	−5	6.790	43
130	326,876	0.269	0.257	−18	8.133	43	0.396	0.382	−47	5.217	0	0.234	0.229	−5	6.625	42
140	326,789	0.259	0.248	−18	8.149	44	0.395	0.381	−47	5.074	1	0.249	0.244	−6	6.697	42
150	326,576	0.227	0.217	−15	6.896	34	0.341	0.329	−39	5.291	−5	0.221	0.217	−3	5.510	31
160	326,479	0.214	0.205	−16	6.274	29	0.291	0.281	−35	4.571	−6	0.206	0.202	−8	4.617	22
170	326,451	0.210	0.201	−15	6.035	26	0.285	0.275	−34	4.611	−8	0.202	0.198	−8	4.441	19
180	326,426	0.196	0.187	−13	5.753	25	0.250	0.242	−28	4.373	−6	0.187	0.183	−2	4.426	21
190	326,408	0.195	0.187	−13	5.682	24	0.249	0.241	−28	4.360	−6	0.188	0.184	−2	4.464	21
200	326,397	0.193	0.184	−13	5.686	24	0.245	0.237	−27	4.252	−5	0.186	0.182	−3	4.382	20
210	326,305	0.187	0.179	−13	5.721	27	0.237	0.229	−26	3.811	0	0.162	0.159	2	4.510	27
220	326,172	0.176	0.168	−14	5.110	26	0.197	0.191	−22	3.346	4	0.146	0.143	6	4.919	31
230	326,160	0.175	0.168	−14	4.994	25	0.206	0.199	−21	3.583	3	0.159	0.155	8	5.114	32
240	326,141	0.166	0.159	−11	5.012	24	0.197	0.190	−16	3.909	5	0.182	0.178	14	5.560	35
250	326,124	0.174	0.166	−12	5.058	25	0.193	0.186	−15	3.833	9	0.188	0.184	17	6.266	41

Table A14: Cont.

k	AIC	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
Normal with identity link under 150-443																
150	325,850	0.247	0.237	−14	9.924	57	0.271	0.262	−35	4.612	22	0.122	0.120	−1	8.537	56
Normal with inverse link under 150-443																
150	325,952	0.258	0.247	−16	8.468	45	0.353	0.341	−44	4.282	3	0.192	0.188	−8	6.088	39
Normal with log link under 150-443																
150	325,823	0.240	0.229	−15	7.980	44	0.316	0.305	−38	4.014	6	0.170	0.167	−2	6.434	42
Gamma with identity link under 150-443																
150	326,041	0.236	0.226	−14	9.329	53	0.260	0.251	−33	4.321	20	0.121	0.118	1	8.206	54
Gamma with inverse link under 150-443																
150	326,222	0.254	0.243	−15	8.410	46	0.327	0.316	−40	4.111	7	0.171	0.167	−3	6.722	44
Gamma with log link under 150-443																
150	326,022	0.243	0.232	−15	7.820	43	0.323	0.312	−40	4.040	3	0.174	0.170	−4	6.010	39
Inverse gaussian with identity link under 150-443																
150	326,352	0.249	0.238	−17	9.375	54	0.337	0.326	−44	4.224	12	0.150	0.146	−4	7.930	52
Inverse gaussian with inverse link under 150-443																
150	326,617	0.253	0.242	−15	8.152	44	0.324	0.313	−39	4.148	5	0.172	0.169	−3	6.476	42
Inverse gaussian with log link under 150-443																
150	326,413	0.247	0.237	−15	7.716	42	0.324	0.313	−40	4.095	2	0.172	0.168	−4	5.892	38
Inverse gaussian with $\frac{1}{\mu^2}$ link under 150-443																
150	326,778	0.262	0.250	−16	8.258	44	0.332	0.321	−41	4.238	5	0.177	0.174	−3	6.518	42
Normal with identity link under 300-886																
224	325,459	0.194	0.186	−9	6.659	34	0.268	0.259	−30	4.200	−2	0.168	0.165	1	5.007	29
Normal with inverse link under 300-886																
213	325,797	0.241	0.230	−12	8.325	44	0.310	0.299	−36	4.063	6	0.171	0.167	−1	6.284	41
Normal with log link under 300-886																
214	325,552	0.206	0.197	−10	6.640	32	0.267	0.258	−29	4.402	−2	0.177	0.173	3	5.180	30
Gamma with identity link under 300-886																
212	325,689	0.189	0.180	−11	5.975	28	0.261	0.252	−29	4.384	−5	0.169	0.165	1	4.545	25
Gamma with inverse link under 300-886																
208	325,969	0.234	0.223	−11	8.162	44	0.288	0.278	−31	4.185	9	0.158	0.154	5	6.832	45
Gamma with log link under 300-886																
226	325,767	0.198	0.189	−8	6.256	29	0.249	0.241	−24	4.532	−1	0.184	0.180	8	5.417	32
Inverse gaussian with identity link under 300-886																
217	326,069	0.191	0.183	−11	5.967	28	0.261	0.252	−29	4.364	−5	0.178	0.175	2	4.779	27
Inverse gaussian with inverse link under 300-886																
219	326,363	0.233	0.223	−11	8.040	43	0.283	0.274	−31	4.130	9	0.153	0.150	5	6.786	45
Inverse gaussian with log link under 300-886																
222	326,153	0.201	0.192	−10	6.291	30	0.261	0.252	−28	4.494	−3	0.180	0.177	5	5.176	30
Inverse gaussian with $\frac{1}{\mu^2}$ link under 300-886																
250	326,124	0.174	0.166	−12	5.058	25	0.193	0.186	−15	3.833	9	0.188	0.184	17	6.266	41

Table A15: AIC scores and out-of-sample validation figures of the normal, gamma and inverse gaussian GLMs of BEL with identity, inverse, log and $\frac{1}{\mu^2}$ link functions under 150-443 and 300-886 after the final iteration. MAEs in %. Highlighted in green and red respectively the best and worst AIC scores and validation figures.

k	$K_{\max}$	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
4 Thin plate regression splines under normal with identity link in stagewise selection of length 5																
0	150	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	150	0.632	0.604	28	22.019	116	0.345	0.334	−8	13.247	65	0.479	0.469	66	21.072	139
20	150	0.406	0.388	0	11.330	44	0.375	0.362	−42	7.254	−12	0.341	0.334	−6	7.709	24
30	150	0.399	0.382	−11	12.268	59	0.465	0.449	−61	5.744	−6	0.314	0.307	−26	6.116	29
40	150	0.371	0.355	−8	11.415	53	0.480	0.463	−64	6.380	−16	0.340	0.332	−34	5.283	13
50	150	0.392	0.375	−13	12.079	59	0.520	0.503	−70	5.961	−12	0.365	0.358	−39	5.368	19
60	150	0.306	0.292	−15	9.833	48	0.405	0.391	−51	5.283	−2	0.273	0.267	−10	6.484	39
70	150	0.272	0.260	−15	9.896	56	0.321	0.310	−35	5.227	22	0.232	0.228	12	10.460	69
80	150	0.249	0.238	−17	8.627	49	0.308	0.297	−36	4.588	16	0.205	0.201	9	9.100	60
90	150	0.261	0.250	−17	9.262	54	0.325	0.314	−39	4.639	18	0.195	0.191	5	9.340	62
100	150	0.254	0.243	−18	9.593	55	0.340	0.328	−42	4.626	17	0.196	0.192	3	9.312	62
110	150	0.255	0.244	−18	9.407	54	0.336	0.324	−40	4.640	18	0.207	0.203	4	9.325	62
120	150	0.243	0.233	−16	8.474	48	0.307	0.296	−38	4.023	13	0.186	0.182	1	7.819	51
130	150	0.241	0.230	−16	8.481	49	0.308	0.298	−37	4.108	13	0.183	0.179	2	8.075	53
140	150	0.235	0.225	−15	8.018	45	0.295	0.285	−35	3.865	10	0.173	0.169	2	7.182	47
150	150	0.240	0.229	−15	8.192	46	0.291	0.281	−35	3.907	13	0.176	0.172	3	7.641	50
5 Thin plate regression splines under normal with identity link																
0	100	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	100	0.643	0.615	27	23.278	125	0.344	0.332	−6	15.238	78	0.493	0.483	69	23.151	153
20	100	0.387	0.370	1	10.371	35	0.364	0.352	−40	7.855	−20	0.335	0.328	−6	7.454	14
30	100	0.382	0.366	−10	11.235	50	0.454	0.439	−60	6.247	−14	0.317	0.310	−28	5.603	18
40	100	0.368	0.352	−11	10.931	48	0.463	0.447	−61	6.266	−16	0.337	0.329	−33	5.343	12
50	100	0.355	0.339	−11	10.086	40	0.481	0.465	−64	7.752	−28	0.351	0.344	−37	5.481	0
60	100	0.344	0.329	−9	10.015	40	0.490	0.474	−66	8.152	−30	0.364	0.356	−38	5.593	−3
70	100	0.339	0.324	−6	10.035	45	0.476	0.460	−64	7.578	−27	0.345	0.337	−37	5.078	0
80	100	0.295	0.282	−11	9.397	49	0.404	0.390	−51	5.513	−6	0.241	0.236	−11	5.820	34
90	100	0.296	0.283	−12	9.694	52	0.393	0.380	−49	5.155	0	0.206	0.202	−7	6.605	41
100	100	0.287	0.274	−11	9.431	48	0.397	0.383	−50	5.402	−5	0.202	0.198	−9	5.945	36
8 Thin plate regression splines under normal with identity link																
0	150	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	150	0.639	0.611	27	23.176	125	0.340	0.329	−3	15.517	80	0.516	0.505	73	23.627	156
20	150	0.375	0.359	3	9.604	26	0.334	0.322	−33	8.378	−24	0.341	0.333	1	7.711	10
30	150	0.361	0.345	−7	10.444	41	0.415	0.401	−52	6.961	−19	0.304	0.297	−21	5.871	13
40	150	0.356	0.340	−5	10.098	36	0.425	0.410	−54	7.920	−28	0.311	0.304	−27	5.647	−1
50	150	0.339	0.324	−7	9.712	33	0.418	0.404	−53	7.746	−27	0.311	0.304	−26	5.596	0
60	150	0.325	0.311	−6	9.037	26	0.411	0.397	−52	8.706	−34	0.310	0.304	−26	5.850	−8
70	150	0.325	0.311	−4	9.180	31	0.429	0.414	−55	8.773	−34	0.326	0.319	−30	5.912	−9
80	150	0.309	0.296	−5	8.618	29	0.430	0.415	−55	8.984	−35	0.336	0.329	−29	6.382	−9
90	150	0.313	0.299	−5	8.981	32	0.384	0.371	−48	7.390	−26	0.300	0.293	−26	5.430	−4
100	150	0.328	0.313	−6	9.910	47	0.400	0.387	−51	5.572	−12	0.291	0.285	−25	5.064	13
110	150	0.256	0.245	−10	7.985	38	0.326	0.315	−40	4.655	−6	0.201	0.197	−6	5.002	28
120	150	0.253	0.242	−9	7.340	30	0.321	0.310	−39	5.542	−14	0.209	0.204	−5	4.541	20
130	150	0.252	0.241	−9	7.767	34	0.326	0.315	−40	5.197	−11	0.205	0.201	−5	4.770	24
140	150	0.245	0.234	−8	7.592	33	0.322	0.311	−41	5.315	−15	0.197	0.193	−7	4.317	20
150	150	0.217	0.208	−11	6.477	32	0.239	0.231	−26	3.652	2	0.179	0.175	6	5.578	34
10 Thin plate regression splines under normal with identity link																
0	150	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	150	0.642	0.614	27	23.354	126	0.344	0.332	−5	15.463	80	0.509	0.499	71	23.654	156
20	150	0.382	0.365	2	10.101	33	0.341	0.329	−34	7.780	−18	0.338	0.331	1	7.728	18
30	150	0.370	0.354	−7	10.922	45	0.416	0.402	−52	6.497	−14	0.305	0.299	−20	6.103	18
40	150	0.354	0.338	−7	10.412	39	0.404	0.391	−51	6.747	−20	0.308	0.301	−24	5.600	8
50	150	0.347	0.331	−7	10.119	38	0.426	0.412	−54	7.258	−24	0.310	0.304	−27	5.467	4
60	150	0.342	0.327	−4	9.766	34	0.400	0.387	−50	7.600	−26	0.298	0.292	−23	5.615	0
70	150	0.334	0.319	−4	9.601	35	0.428	0.414	−55	8.158	−30	0.318	0.311	−29	5.618	−5
80	150	0.315	0.301	−5	9.093	35	0.432	0.418	−55	8.113	−29	0.334	0.327	−29	6.087	−3
90	150	0.323	0.309	−5	9.436	38	0.388	0.375	−49	6.558	−20	0.297	0.291	−26	5.194	2
100	150	0.309	0.296	−6	8.722	27	0.409	0.395	−54	8.780	−36	0.261	0.255	−27	4.994	−9
110	150	0.309	0.295	−6	8.542	26	0.411	0.397	−54	8.711	−37	0.284	0.278	−33	4.768	−15
120	150	0.206	0.197	−9	5.768	25	0.216	0.209	−23	3.806	−4	0.164	0.161	5	4.519	24
130	150	0.205	0.196	−10	5.759	24	0.226	0.218	−24	3.952	−5	0.175	0.172	4	4.579	24
140	150	0.214	0.205	−10	6.761	34	0.228	0.220	−25	3.363	5	0.167	0.163	6	5.762	36
150	150	0.212	0.203	−10	7.070	37	0.230	0.223	−24	3.575	8	0.173	0.170	8	6.337	40

Table A16: Out-of-sample validation figures of selected GAMs of BEL with varying spline function number per dimension and fixed spline function type under 150-443 after each tenth and the finally selected smooth function. MAEs in %.

	$J=4$, $k=50$			$J=4$, $k=100$			$J=4$, $k=150$			$J=10$, $k=50$			$J=10$, $k=100$			$J=10$, $k=150$		
k	df	p-val	sign	df	p-val	sign	df	p-val	sign	df	p-val	sign	df	p-val	sign	df	p-val	sign
1	2.858	2_{-16}	***	2.350	2_{-16}	***	1.948	2_{-16}	***	9.000	2_{-16}	***	8.941	2_{-16}	***	7.724	2_{-16}	***
2	3.000	2_{-16}	***	2.104	2_{-16}	***	1.000	2_{-16}	***	7.857	2_{-16}	***	4.436	2_{-16}	***	1.000	2_{-16}	***
3	3.000	2_{-16}	***	2.901	2_{-16}	***	2.922	2_{-16}	***	5.600	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***
4	2.997	2_{-16}	***	2.962	2_{-16}	***	2.998	2_{-16}	***	7.073	2_{-16}	***	6.791	2_{-16}	***	7.288	2_{-16}	***
5	2.729	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	8.679	2_{-16}	***	8.870	2_{-16}	***	8.210	2_{-16}	***
6	3.000	2_{-16}	***	3.000	2_{-16}	***	1.043	2_{-16}	***	3.417	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***
7	3.000	2_{-16}	***	2.806	2_{-16}	***	2.841	2_{-16}	***	7.990	2_{-16}	***	8.608	2_{-16}	***	1.000	2_{-16}	***
8	3.000	2_{-16}	***	2.956	2_{-16}	***	2.961	2_{-16}	***	8.282	2_{-16}	***	8.292	2_{-16}	***	8.122	2_{-16}	***
9	1.000	2_{-16}	***	1.000	2_{-16}	***	2.223	2_{-16}	***	7.710	2_{-16}	***	6.510	2_{-16}	***	6.549	2_{-16}	***
10	2.991	2_{-16}	***	2.924	2_{-16}	***	3.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***
11	2.587	2_{-16}	***	2.922	2_{-16}	***	2.889	2_{-16}	***	6.535	2_{-16}	***	7.014	2_{-16}	***	5.672	2_{-16}	***
12	2.645	2_{-16}	***	1.874	2_{-16}	***	1.000	2_{-16}	***	7.235	2_{-16}	***	7.284	2_{-16}	***	8.346	2_{-16}	***
13	2.244	2_{-16}	***	2.425	2_{-16}	***	1.000	2_{-16}	***	2.372	2_{-16}	***	2.531	2_{-16}	***	1.000	2_{-16}	***
14	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***
15	3.000	2_{-16}	***	1.000	2_{-16}	***	2.285	2_{-16}	***	5.430	2_{-16}	***	5.640	2_{-16}	***	4.437	2_{-16}	***
16	1.000	2_{-16}	***	1.000	2_{-16}	***	2.783	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***
17	2.344	2_{-16}	***	1.670	2_{-16}	***	1.646	2_{-16}	***	3.886	2_{-16}	***	1.610	2_{-16}	***	1.624	2_{-16}	***
18	3.000	2_{-16}	***	3.000	2_{-16}	***	3.000	2_{-16}	***	8.751	2_{-16}	***	8.620	1.4_{-5}	***	5.367	6.9_{-5}	***
19	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***
20	1.497	2_{-16}	***	1.501	2_{-16}	***	2.148	2_{-16}	***	1.754	2_{-16}	***	1.000	2_{-16}	***	3.141	8.1_{-16}	***
21	1.441	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***
22	1.770	2_{-16}	***	2.192	2_{-16}	***	1.400	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	3.985	1.9_{-9}	***
23	2.395	2_{-16}	***	2.746	2_{-16}	***	2.911	2_{-16}	***	2.057	2_{-16}	***	1.428	2_{-16}	***	2.663	2_{-16}	***
24	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	2.964	2_{-16}	***	1.000	3.3_{-13}	***	1.000	1.1_{-13}	***
25	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***
26	1.000	2_{-16}	***	1.485	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***
27	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2.2_{-10}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	1.6_{-10}	***
28	1.000	2_{-16}	***	2.607	2_{-16}	***	1.839	2_{-16}	***	1.000	2_{-16}	***	2.780	2_{-16}	***	1.914	2_{-16}	***
29	1.000	2_{-16}	***	1.000	2_{-16}	***	1.809	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***
30	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	6.740	2_{-16}	***	6.416	2_{-16}	***	6.508	2_{-16}	***
31	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2.4_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***
32	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***
33	1.000	2_{-16}	***	2.055	4.9_{-15}	***	1.893	2.2_{-15}	***	7.111	2_{-16}	***	7.175	6.3_{-12}	***	6.728	2_{-16}	***
34	1.000	3.2_{-16}	***	1.000	2.9_{-16}	***	1.000	8.7_{-11}	***	1.000	2_{-16}	***	1.213	2_{-16}	***	1.635	4.9_{-16}	***
35	3.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2.5_{-16}	***	4.780	2_{-16}	***	4.013	2_{-16}	***	4.224	2_{-16}	***
36	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	7.825	4.8_{-16}	***	7.867	1.1_{-15}	***	7.738	2.3_{-3}	**
37	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	2_{-16}	***	1.000	4.6_{-16}	***	1.000	7.5_{-16}	***	1.000	2_{-16}	***
38	2.512	1.1_{-14}	***	2.303	2_{-16}	***	2.057	2_{-16}	***	1.233	2_{-16}	***	1.000	2_{-16}	***	1.000	1.1_{-4}	***
39	1.000	2.7_{-12}	***	1.000	1.2_{-13}	***	1.000	1.9_{-13}	***	1.000	1.1_{-15}	***	1.000	2.6_{-16}	***	1.000	1.2_{-14}	***
40	1.826	6.4_{-11}	***	1.000	2_{-16}	***	1.915	3.6_{-15}	***	1.000	1.2_{-13}	***	1.514	2_{-16}	***	1.000	2_{-16}	***
41	2.668	7.5_{-16}	***	2.701	5.3_{-15}	***	1.787	9.8_{-7}	***	1.823	8.1_{-12}	***	1.319	9.4_{-15}	***	1.000	2_{-16}	***
42	1.000	1.1_{-15}	***	1.000	2_{-16}	***	1.000	2_{-15}	***	1.000	2.9_{-12}	***	1.000	8_{-12}	***	5.275	3.8_{-4}	***
43	1.000	3.8_{-10}	***	1.000	9.5_{-10}	***	1.000	2_{-9}	***	1.000	3.3_{-10}	***	1.000	7.7_{-11}	***	1.000	1.1_{-10}	***
44	1.713	1.3_{-8}	***	1.887	8.2_{-9}	***	1.892	6.2_{-9}	***	2.109	6_{-8}	***	1.779	5.3_{-8}	***	2.061	3.4_{-8}	***
45	1.000	5.7_{-9}	***	1.000	6.4_{-9}	***	1.000	1.9_{-8}	***	1.000	8_{-9}	***	1.000	2.1_{-8}	***	1.000	8.8_{-9}	***
46	1.917	3.5_{-9}	***	1.000	2_{-16}	***	1.000	1.3_{-15}	***	1.305	1.9_{-6}	***	1.610	1.1_{-6}	***	1.000	8.7_{-8}	***
47	1.451	1.2_{-6}	***	1.507	5.8_{-7}	***	1.234	1_{-6}	***	1.000	7.7_{-13}	***	1.000	5.5_{-13}	***	1.000	7.4_{-12}	***
48	2.753	3.2_{-7}	***	2.863	6.5_{-8}	***	2.804	2.1_{-8}	***	1.000	2.4_{-8}	***	1.000	7.8_{-8}	***	1.000	2.9_{-6}	***
49	1.000	5.5_{-7}	***	1.000	4.7_{-14}	***	1.000	1.6_{-11}	***	1.000	6.9_{-7}	***	1.000	9.6_{-12}	***	1.000	1.6_{-12}	***
50	1.000	9.2_{-7}	***	1.372	8.3_{-11}	***	1.000	1.1_{-12}	***	1.000	1.1_{-6}	***	1.000	2_{-10}	***	1.000	2_{-11}	***
51				1.004	2_{-16}	***	1.000	2_{-16}	***				1.000	1.1_{-6}	***	1.000	1.3_{-6}	***
52				2.839	2_{-16}	***	1.334	2_{-16}	***				1.000	4.3_{-13}	***	1.000	3_{-13}	***
53				2.640	2_{-16}	***	2.421	2_{-16}	***				1.000	4.7_{-10}	***	1.000	7.1_{-11}	***
54				2.664	2_{-16}	***	1.000	2_{-16}	***				3.237	2.8_{-6}	***	3.168	4.9_{-6}	***
55				1.000	9.2_{-9}	***	1.000	3.1_{-6}	***				3.906	5.8_{-8}	***	3.493	1_{-9}	***
56				1.000	2.8_{-9}	***	2.376	2.3_{-8}	***				1.098	3.5_{-5}	***	3.513	2_{-16}	***
57				1.000	3.3_{-15}	***	1.000	2.8_{-13}	***				5.574	5.1_{-3}	**	5.019	6.7_{-2}	.
58				1.000	2_{-16}	***	1.000	2_{-16}	***				1.000	7.3_{-5}	***	1.000	1_{-5}	***
59				1.000	1.2_{-11}	***	1.000	2_{-11}	***				1.000	1.8_{-6}	***	1.000	8.8_{-8}	***
60				1.000	2_{-16}	***	1.000	2_{-16}	***				3.717	5.2_{-4}	***	3.286	5.6_{-3}	**
61				1.000	7.5_{-11}	***	1.000	7.1_{-11}	***				1.000	6.7_{-5}	***	1.000	1.5_{-5}	***
62				2.613	4.2_{-4}	***	2.868	2_{-16}	***				1.000	1.1_{-5}	***	1.000	4.6_{-6}	***
63				1.000	7.9_{-15}	***	1.867	1.6_{-14}	***				4.210	6.6_{-3}	**	3.543	7.3_{-4}	***
64				1.000	2.4_{-6}	***	1.000	1.2_{-6}	***				1.000	1.7_{-4}	***	1.000	3.4_{-4}	***
65				2.960	2.3_{-13}	***	2.976	2_{-16}	***				2.799	7.1_{-3}	**	2.861	3_{-3}	**
66				1.904	2_{-16}	***	2.115	2_{-16}	***				3.054	1.7_{-3}	**	3.159	8.8_{-6}	***
67				2.859	9.1_{-14}	***	2.778	1.1_{-13}	***				3.671	7.6_{-3}	**	3.788	8.4_{-4}	***
68				1.000	2.9_{-1}		1.000	5.2_{-11}	***				1.000	4_{-4}	***	1.000	1.2_{-4}	***
69				2.797	2.8_{-3}	**	2.954	2.2_{-3}	**				1.000	2.8_{-3}	**	1.000	3.3_{-3}	**
70				1.000	2.4_{-6}	***	1.000	1.5_{-6}	***				1.000	6.7_{-3}	**	1.000	1.1_{-3}	**
71				2.957	6_{-14}	***	2.996	6.1_{-15}	***				1.000	8.6_{-3}	**	1.000	5_{-3}	**
72				2.612	1.4_{-13}	***	2.101	6.3_{-11}	***				1.000	1.2_{-2}	*	1.000	8.9_{-3}	**
73				1.196	2_{-16}	***	3.000	2_{-16}	***				1.000	1.5_{-2}	*	1.000	6.1_{-5}	***
74				2.994	3.8_{-6}	***	2.559	1.8_{-3}	**				3.644	1.2_{-1}		2.988	1.4_{-1}	
75				1.000	1.7_{-14}	***	1.000	3_{-14}	***				1.000	1.7_{-2}	*	1.000	1.8_{-2}	*

Table A17: Effective degrees of freedom, p-values and significance codes per dimension of GAMs of BEL built up of thin plate regression splines with normal random component and identity link function under 150-443 for spline function numbers $J \in \{4, 10\}$ per dimension at stages $k \in \{50, 100, 150\}$. The confidence levels corresponding to the indicated significance codes are '***' = 0.001, '**' = 0.01, '*' = 0.05, '.' = 0.1, ' ' = 1.

	$J=4$, $k=50$			$J=4$, $k=100$			$J=4$, $k=150$			$J=10$, $k=50$			$J=10$, $k=100$			$J=10$, $k=150$		
k	df	p-val	sign	df	p-val	sign	df	p-val	sign	df	p-val	sign	df	p-val	sign	df	p-val	sign
76				1.000	4.4_{-13}	***	2.334	3.8_{-14}	***				2.469	1_{-1}		2.077	1.8_{-1}	
77				1.353	4_{-9}	***	1.411	8.8_{-9}	***				1.000	2.5_{-2}	*	1.000	1.1_{-2}	*
78				1.000	1.5_{-5}	***	1.000	6.5_{-6}	***				1.000	2_{-16}	***	1.000	1.6_{-4}	***
79				1.000	3_{-5}	***	1.000	1.5_{-5}	***				5.186	1.5_{-6}	***	1.000	2_{-16}	***
80				1.000	1_{-7}	***	1.000	7.8_{-8}	***				1.892	2.2_{-2}	*	1.795	1.9_{-2}	*
81				2.725	1.3_{-4}	***	2.739	7.1_{-5}	***				1.000	5.2_{-6}	***	1.000	5.8_{-1}	
82				1.000	7.6_{-5}	***	2.175	1.4_{-5}	***				1.000	1.8_{-3}	**	1.000	5.1_{-1}	
83				2.240	1.3_{-3}	**	2.075	9_{-4}	***				7.020	2_{-16}	***	4.809	2.9_{-3}	**
84				1.000	6.8_{-5}	***	2.902	1.5_{-5}	***				4.003	1.5_{-1}		4.722	9.8_{-3}	**
85				1.000	7.5_{-5}	***	1.000	4_{-6}	***				1.000	1_{-9}	***	1.000	1.8_{-4}	***
86				1.000	3.7_{-4}	***	1.000	7.7_{-4}	***				3.115	1.2_{-1}		2.748	1.2_{-1}	
87				1.000	3.4_{-4}	***	1.000	9.1_{-5}	***				5.294	1.4_{-1}		5.598	1.3_{-1}	
88				1.000	1.9_{-4}	***	1.000	9.6_{-5}	***				2.263	1.5_{-1}		1.788	2.5_{-1}	
89				2.828	2.1_{-3}	**	3.000	6_{-5}	***				1.000	3.4_{-4}	***	1.000	3.3_{-4}	***
90				1.000	7.8_{-4}	***	1.000	5.6_{-4}	***				1.000	3.7_{-2}	*	1.000	3.8_{-2}	*
91				1.000	2.5_{-3}	**	1.000	2.9_{-3}	**				1.000	1.8_{-3}	**	1.000	1.2_{-3}	**
92				1.000	3.8_{-3}	**	1.000	3.5_{-3}	**				1.000	1.7_{-2}	*	1.000	1.2_{-2}	*
93				1.000	1.8_{-3}	**	1.000	1.1_{-3}	**				1.000	3.8_{-2}	*	1.000	2.8_{-2}	*
94				2.776	3.6_{-5}	***	1.000	1.8_{-7}	***				5.921	4.2_{-3}	**	3.962	2_{-16}	***
95				2.103	4.9_{-2}	*	1.974	1.3_{-1}					8.154	2_{-16}	***	2.290	2_{-16}	***
96				2.023	1.2_{-4}	***	1.000	4.6_{-10}	***				1.000	2.8_{-12}	***	1.000	1.6_{-5}	***
97				2.811	1.5_{-2}	*	2.873	5.9_{-3}	**				3.748	7.1_{-4}	***	1.000	1.2_{-6}	***
98				1.000	7.1_{-3}	**	1.000	1.1_{-2}	*				1.000	3.9_{-6}	***	7.349	2.8_{-1}	
99				1.000	1.4_{-2}	*	1.000	1.9_{-2}	*				2.149	1.2_{-3}	**	1.000	2.8_{-8}	***
100				2.764	2.9_{-2}	*	2.321	9_{-2}	.				1.000	3.1_{-3}	**	1.000	2.1_{-1}	
101							1.000	1.1_{-4}	***							1.000	8.2_{-10}	***
102							1.000	7.7_{-2}	.							1.000	1.6_{-2}	*
103							1.000	2.9_{-3}	**							4.084	5.8_{-4}	***
104							1.000	6.8_{-5}	***							1.000	3.2_{-2}	*
105							1.000	9.3_{-3}	**							1.000	6.8_{-2}	.
106							1.000	2.1_{-9}	***							1.000	5.2_{-3}	**
107							1.000	1.9_{-2}	*							3.397	1_{-1}	
108							2.187	9.6_{-2}	.							1.248	3.4_{-1}	
109							1.000	2.1_{-3}	**							3.079	3.9_{-1}	
110							1.000	4.6_{-2}	*							1.000	3.9_{-4}	***
111							1.000	2_{-16}	***							0.979	4.3_{-8}	***
112							1.000	2.9_{-2}	*							8.555	2_{-16}	***
113							1.000	9.5_{-1}								8.952	1.7_{-12}	***
114							1.644	9.6_{-2}	.							1.000	2_{-16}	***
115							1.000	2_{-2}	*							1.000	2_{-16}	***
116							1.000	1.8_{-2}	*							1.000	1.7_{-13}	***
117							1.000	4.8_{-3}	**							2.988	3.4_{-13}	***
118							1.000	2.4_{-2}	*							8.401	1.2_{-10}	***
119							2.704	8.3_{-2}	.							2.493	4.7_{-5}	***
120							1.000	1.8_{-2}	*							1.000	4.1_{-7}	***
121							1.413	6.7_{-1}								1.000	9_{-5}	***
122							1.886	6.2_{-1}								2.745	1.2_{-3}	**
123							1.000	1.4_{-5}	***							1.000	3.4_{-3}	**
124							2.499	1.8_{-1}								1.000	1.5_{-2}	*
125							1.000	3.6_{-2}	*							1.000	1.4_{-2}	*
126							2.416	1_{-1}								1.000	5.8_{-3}	**
127							1.000	5.1_{-5}	***							3.120	5.7_{-2}	.
128							1.000	3.8_{-2}	*							1.000	9.2_{-4}	***
129							1.000	1.3_{-3}	**							1.000	3.9_{-3}	**
130							1.000	5.7_{-2}	.							3.778	1.7_{-1}	
131							1.000	1.3_{-2}	*							2.752	2.7_{-2}	*
132							1.000	1.2_{-2}	*							1.000	6.9_{-3}	**
133							1.970	2.5_{-1}								1.000	4.8_{-3}	**
134							1.000	3.5_{-2}	*							1.000	5.5_{-2}	.
135							1.000	5.9_{-4}	***							1.000	3.8_{-2}	*
136							1.176	7.1_{-3}	**							5.289	1.4_{-1}	
137							2.357	3.4_{-1}								1.000	3.7_{-2}	*
138							1.000	6.7_{-2}	.							1.000	2_{-4}	***
139							1.000	7.9_{-2}	.							1.000	5.1_{-3}	**
140							1.000	6.9_{-2}	.							1.000	1.6_{-1}	
141							1.000	4.7_{-2}	*							8.453	2.5_{-3}	**
142							1.000	1.3_{-3}	**							1.000	4_{-2}	*
143							2.602	4.1_{-2}	*							3.975	1.4_{-1}	
144							1.631	4.6_{-1}								1.000	4.2_{-4}	***
145							1.000	8.3_{-2}	.							1.000	3.7_{-3}	**
146							1.000	1_{-2}	*							2.147	1.9_{-1}	
147							1.000	3.6_{-2}	*							1.000	5_{-2}	.
148							1.251	1.6_{-1}								1.000	4.1_{-2}	*
149							2.376	2.1_{-1}								1.000	5.4_{-2}	.
150							1.482	2_{-1}								1.000	6.3_{-2}	.

Table A17: Cont.

k	$K_{\max}$	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
5 Thin plate regression splines under normal with identity link																
0	100	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	100	0.643	0.615	27	23.278	125	0.344	0.332	−6	15.238	78	0.493	0.483	69	23.151	153
20	100	0.387	0.370	1	10.371	35	0.364	0.352	−40	7.855	−20	0.335	0.328	−6	7.454	14
30	100	0.382	0.366	−10	11.235	50	0.454	0.439	−60	6.247	−14	0.317	0.310	−28	5.603	18
40	100	0.368	0.352	−11	10.931	48	0.463	0.447	−61	6.266	−16	0.337	0.329	−33	5.343	12
50	100	0.355	0.339	−11	10.086	40	0.481	0.465	−64	7.752	−28	0.351	0.344	−37	5.481	0
60	100	0.344	0.329	−9	10.015	40	0.490	0.474	−66	8.152	−30	0.364	0.356	−38	5.593	−3
70	100	0.339	0.324	−6	10.035	45	0.476	0.460	−64	7.578	−27	0.345	0.337	−37	5.078	0
80	100	0.295	0.282	−11	9.397	49	0.404	0.390	−51	5.513	−6	0.241	0.236	−11	5.820	34
90	100	0.296	0.283	−12	9.694	52	0.393	0.380	−49	5.155	0	0.206	0.202	−7	6.605	41
100	100	0.287	0.274	−11	9.431	48	0.397	0.383	−50	5.402	−5	0.202	0.198	−9	5.945	36
5 Cubic regression splines under normal with identity link																
0	100	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	100	0.637	0.609	28	22.739	122	0.337	0.326	−4	14.733	75	0.505	0.494	71	22.781	150
20	100	0.388	0.371	2	10.094	32	0.358	0.346	−40	8.256	−25	0.319	0.313	−5	7.161	10
30	100	0.389	0.372	−6	11.426	50	0.436	0.421	−55	6.652	−14	0.289	0.283	−19	5.849	22
40	100	0.359	0.343	−9	10.508	41	0.448	0.433	−59	7.171	−23	0.310	0.303	−29	5.175	6
50	100	0.345	0.330	−9	9.906	35	0.476	0.460	−63	8.736	−34	0.328	0.321	−34	5.373	−5
60	100	0.338	0.323	−7	9.817	34	0.475	0.459	−63	9.192	−37	0.330	0.324	−34	5.491	−8
70	100	0.307	0.294	−8	9.341	47	0.430	0.416	−58	6.081	−18	0.234	0.229	−26	3.871	15
80	100	0.289	0.277	−13	10.157	55	0.410	0.396	−53	5.106	0	0.237	0.232	−11	6.939	43
90	100	0.283	0.271	−13	10.307	56	0.407	0.394	−53	5.067	1	0.229	0.224	−10	7.035	44
100	100	0.268	0.256	−12	9.903	52	0.399	0.386	−51	5.182	−2	0.226	0.221	−9	6.533	40
5 Duchon splines under normal with identity link																
0	100	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	100	0.753	0.720	−4	20.570	98	0.428	0.413	−39	11.806	49	0.408	0.399	6	15.241	93
20	100	0.704	0.673	−22	17.488	74	0.441	0.426	−51	8.606	31	0.380	0.372	−16	11.600	66
30	100	0.661	0.632	−32	19.699	95	0.376	0.363	−40	14.235	73	0.319	0.312	11	19.168	124
40	100	0.663	0.634	−21	18.426	84	0.292	0.282	−18	14.138	73	0.377	0.370	33	19.007	123
50	100	0.666	0.636	−17	18.534	86	0.287	0.277	−12	14.785	76	0.410	0.402	41	19.896	130
56	100	0.666	0.636	−18	18.532	86	0.288	0.279	−14	14.643	75	0.406	0.397	40	19.757	129
5 Eilers and Marx style P-splines under normal with identity link																
0	100	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	100	0.643	0.615	29	22.836	123	0.344	0.332	−9	13.951	70	0.471	0.461	65	21.854	144
20	100	0.389	0.372	1	10.496	37	0.365	0.353	−41	7.778	−20	0.336	0.329	−8	7.402	13
30	100	0.384	0.367	−9	11.377	53	0.459	0.444	−60	6.138	−13	0.320	0.313	−30	5.512	17
40	100	0.371	0.354	−10	10.977	49	0.454	0.439	−60	6.095	−16	0.327	0.320	−34	5.092	11
50	100	0.357	0.341	−9	10.459	45	0.467	0.451	−62	6.909	−22	0.335	0.328	−34	5.059	6
60	100	0.339	0.324	−10	9.932	43	0.492	0.476	−66	7.640	−28	0.365	0.357	−40	5.155	−2
70	100	0.343	0.328	−10	10.523	52	0.546	0.527	−75	7.681	−27	0.366	0.358	−46	4.576	2
80	100	0.334	0.319	−7	9.920	45	0.520	0.503	−67	8.655	−29	0.346	0.339	−36	5.036	1
90	100	0.228	0.218	−10	6.973	35	0.279	0.269	−31	4.299	0	0.208	0.204	3	5.810	34
100	100	0.225	0.215	−11	6.897	34	0.256	0.248	−30	3.716	2	0.164	0.161	1	5.212	32

Table A18: Out-of-sample validation figures of selected GAMs of BEL with varying spline function type and fixed spline function number of 5 per dimension under 100-443 after each tenth and the finally selected smooth function. MAEs in %.

k	K_{max}	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
10 Thin plate regression splines under normal with identity link																
0	150	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	150	0.642	0.614	27	23.354	126	0.344	0.332	−5	15.463	80	0.509	0.499	71	23.654	156
20	150	0.382	0.365	2	10.101	33	0.341	0.329	−34	7.780	−18	0.338	0.331	1	7.728	18
30	150	0.370	0.354	−7	10.922	45	0.416	0.402	−52	6.497	−14	0.305	0.299	−20	6.103	18
40	150	0.354	0.338	−7	10.412	39	0.404	0.391	−51	6.747	−20	0.308	0.301	−24	5.600	8
50	150	0.347	0.331	−7	10.119	38	0.426	0.412	−54	7.258	−24	0.310	0.304	−27	5.467	4
60	150	0.342	0.327	−4	9.766	34	0.400	0.387	−50	7.600	−26	0.298	0.292	−23	5.615	0
70	150	0.334	0.319	−4	9.601	35	0.428	0.414	−55	8.158	−30	0.318	0.311	−29	5.618	−5
80	150	0.315	0.301	−5	9.093	35	0.432	0.418	−55	8.113	−29	0.334	0.327	−29	6.087	−3
90	150	0.323	0.309	−5	9.436	38	0.388	0.375	−49	6.558	−20	0.297	0.291	−26	5.194	2
100	150	0.309	0.296	−6	8.722	27	0.409	0.395	−54	8.780	−36	0.261	0.255	−27	4.994	−9
110	150	0.309	0.295	−6	8.542	26	0.411	0.397	−54	8.711	−37	0.284	0.278	−33	4.768	−15
120	150	0.206	0.197	−9	5.768	25	0.216	0.209	−23	3.806	−4	0.164	0.161	5	4.519	24
130	150	0.205	0.196	−10	5.759	24	0.226	0.218	−24	3.952	−5	0.175	0.172	4	4.579	24
140	150	0.214	0.205	−10	6.761	34	0.228	0.220	−25	3.363	5	0.167	0.163	6	5.762	36
150	150	0.212	0.203	−10	7.070	37	0.230	0.223	−24	3.575	8	0.173	0.170	8	6.337	40
10 Cubic regression splines under normal with identity link																
0	125	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	125	0.638	0.610	27	23.397	127	0.341	0.329	−3	15.829	82	0.519	0.509	73	23.960	158
20	125	0.380	0.364	2	10.038	34	0.339	0.328	−34	7.650	−16	0.345	0.338	0	7.865	18
30	125	0.377	0.360	−6	11.458	53	0.411	0.397	−50	6.035	−5	0.309	0.302	−14	6.976	30
40	125	0.364	0.348	−10	10.929	47	0.421	0.407	−53	5.791	−10	0.315	0.308	−25	5.824	18
50	125	0.348	0.333	−11	10.437	44	0.436	0.421	−56	6.263	−15	0.319	0.312	−27	5.636	13
60	125	0.342	0.327	−5	9.791	36	0.403	0.389	−50	7.282	−23	0.308	0.302	−23	5.789	4
70	125	0.355	0.340	−3	10.502	48	0.442	0.427	−56	7.001	−20	0.327	0.320	−30	5.570	6
80	125	0.349	0.334	−2	10.275	46	0.434	0.419	−55	7.159	−22	0.326	0.319	−29	5.592	4
90	125	0.282	0.269	−5	7.978	37	0.275	0.266	−30	4.426	−3	0.215	0.210	−2	5.088	25
100	125	0.263	0.251	−5	7.109	29	0.301	0.291	−37	5.637	−17	0.200	0.196	−8	3.969	12
110	125	0.255	0.244	−7	6.999	30	0.303	0.292	−37	5.435	−15	0.202	0.198	−6	4.230	16
120	125	0.257	0.246	−7	7.052	30	0.304	0.294	−37	5.371	−14	0.200	0.196	−6	4.232	17
125	125	0.254	0.243	−7	7.139	31	0.299	0.289	−36	5.189	−13	0.197	0.192	−6	4.228	17
10 Duchon splines under normal with identity link																
0	100	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	100	0.786	0.752	−5	22.143	110	0.445	0.430	−44	12.588	57	0.406	0.397	1	16.238	102
20	100	0.783	0.749	−32	20.489	101	0.494	0.477	−62	11.319	58	0.357	0.350	−21	15.316	98
30	100	0.782	0.748	−39	21.134	98	0.538	0.520	−59	12.715	64	0.422	0.413	−3	18.621	121
40	100	0.816	0.780	−45	22.125	98	0.559	0.540	−63	13.071	65	0.450	0.440	−10	18.616	119
50	100	0.823	0.787	−45	21.473	96	0.555	0.536	−63	12.672	63	0.451	0.441	−10	18.114	116
53	100	0.821	0.785	−44	21.348	94	0.545	0.526	−61	12.593	62	0.446	0.437	−8	18.091	116
10 Eilers and Marx style P-splines under normal with identity link in stagewise selection of length 5																
0	150	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	150	0.648	0.619	27	23.688	128	0.349	0.337	−7	15.566	80	0.506	0.495	71	23.889	158
20	150	0.398	0.380	1	10.946	45	0.358	0.346	−37	7.063	−7	0.338	0.331	1	8.102	31
30	150	0.393	0.376	−9	11.983	59	0.435	0.421	−55	5.575	−2	0.299	0.293	−17	6.928	36
40	150	0.371	0.355	−8	11.374	55	0.449	0.434	−57	5.738	−9	0.314	0.308	−26	5.770	23
50	150	0.363	0.347	−9	10.956	50	0.460	0.444	−60	6.249	−14	0.315	0.308	−28	5.492	17
60	150	0.349	0.334	−8	10.479	46	0.443	0.428	−56	6.526	−17	0.305	0.298	−26	5.427	14
70	150	0.349	0.333	−6	10.629	51	0.464	0.449	−60	6.687	−17	0.325	0.318	−29	5.501	13
80	150	0.350	0.335	−7	10.465	48	0.468	0.452	−60	7.036	−19	0.335	0.328	−29	5.563	11
90	150	0.350	0.335	−7	10.639	51	0.470	0.454	−60	6.683	−17	0.330	0.323	−29	5.453	14
100	150	0.334	0.319	−8	9.960	46	0.468	0.452	−60	7.170	−20	0.339	0.332	−29	5.835	11
110	150	0.337	0.323	−9	10.249	48	0.450	0.435	−58	6.171	−15	0.329	0.322	−31	5.267	12
120	150	0.339	0.324	−7	10.283	45	0.433	0.419	−55	6.420	−17	0.320	0.313	−28	5.340	10
130	150	0.269	0.257	−13	8.912	43	0.365	0.352	−46	4.891	−4	0.244	0.238	−12	5.503	30
140	150	0.255	0.244	−12	8.157	36	0.356	0.344	−44	5.415	−10	0.246	0.241	−10	5.196	24
150	150	0.261	0.250	−12	8.514	39	0.368	0.355	−46	5.267	−9	0.245	0.240	−12	5.162	25

Table A19: Out-of-sample validation figures of selected GAMs of BEL with varying spline function type and fixed spline function number of 10 per dimension under between 100-443 and 150-443 after each tenth and the finally selected smooth function. MAEs in %.

k	$K_{\max}$	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
4 Thin plate regression splines under normal with identity link in stagewise selection of length 5																
0	150	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	150	0.632	0.604	28	22.019	116	0.345	0.334	−8	13.247	65	0.479	0.469	66	21.072	139
20	150	0.406	0.388	0	11.330	44	0.375	0.362	−42	7.254	−12	0.341	0.334	−6	7.709	24
30	150	0.399	0.382	−11	12.268	59	0.465	0.449	−61	5.744	−6	0.314	0.307	−26	6.116	29
40	150	0.371	0.355	−8	11.415	53	0.480	0.463	−64	6.380	−16	0.340	0.332	−34	5.283	13
50	150	0.392	0.375	−13	12.079	59	0.520	0.503	−70	5.961	−12	0.365	0.358	−39	5.368	19
60	150	0.306	0.292	−15	9.833	48	0.405	0.391	−51	5.283	−2	0.273	0.267	−10	6.484	39
70	150	0.272	0.260	−15	9.896	56	0.321	0.310	−35	5.227	22	0.232	0.228	12	10.460	69
80	150	0.249	0.238	−17	8.627	49	0.308	0.297	−36	4.588	16	0.205	0.201	9	9.100	60
90	150	0.261	0.250	−17	9.262	54	0.325	0.314	−39	4.639	18	0.195	0.191	5	9.340	62
100	150	0.254	0.243	−18	9.593	55	0.340	0.328	−42	4.626	17	0.196	0.192	3	9.312	62
110	150	0.255	0.244	−18	9.407	54	0.336	0.324	−40	4.640	18	0.207	0.203	4	9.325	62
120	150	0.243	0.233	−16	8.474	48	0.307	0.296	−38	4.023	13	0.186	0.182	1	7.819	51
130	150	0.241	0.230	−16	8.481	49	0.308	0.298	−37	4.108	13	0.183	0.179	2	8.075	53
140	150	0.235	0.225	−15	8.018	45	0.295	0.285	−35	3.865	10	0.173	0.169	2	7.182	47
150	150	0.240	0.229	−15	8.192	46	0.291	0.281	−35	3.907	13	0.176	0.172	3	7.641	50
4 Thin plate regression splines under normal with log link in stagewise selection of length 5																
0	40	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	40	0.788	0.754	8	23.011	114	0.423	0.408	26	22.471	118	0.700	0.685	94	28.248	186
20	40	0.452	0.432	−4	12.761	50	0.421	0.406	−48	7.626	−9	0.360	0.352	−11	8.166	29
30	40	0.462	0.442	−10	14.180	72	0.527	0.509	−68	6.209	−1	0.368	0.360	−32	7.116	36
40	40	0.438	0.419	−7	13.382	66	0.524	0.506	−69	6.189	−10	0.373	0.365	−39	5.913	20
4 Thin plate regression splines under gamma with identity link in stagewise selection of length 5																
0	70	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	70	0.625	0.598	31	21.068	110	0.332	0.321	−5	12.421	60	0.486	0.475	68	19.997	132
20	70	0.394	0.377	1	10.887	41	0.357	0.345	−39	7.283	−15	0.340	0.333	−6	7.641	19
30	70	0.383	0.367	−10	11.985	56	0.467	0.451	−62	5.853	−10	0.331	0.324	−30	5.742	22
40	70	0.289	0.277	−11	9.447	45	0.346	0.335	−41	5.159	0	0.256	0.250	−2	6.682	39
50	70	0.307	0.293	−11	10.339	53	0.389	0.376	−50	4.922	0	0.252	0.247	−11	6.294	38
60	70	0.308	0.295	−14	10.455	56	0.372	0.360	−49	4.377	7	0.222	0.218	−9	7.143	46
70	70	0.270	0.259	−16	9.999	57	0.325	0.314	−36	5.280	23	0.245	0.240	10	10.416	69
4 Thin plate regression splines under gamma with log link in stagewise selection of length 5																
0	120	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	120	0.780	0.745	12	22.104	101	0.436	0.421	35	21.150	110	0.736	0.720	101	26.692	175
20	120	0.497	0.475	−1	14.721	71	0.457	0.442	−55	6.794	2	0.360	0.352	−16	8.605	41
30	120	0.437	0.418	−7	13.581	66	0.483	0.467	−61	6.042	−3	0.364	0.357	−28	7.018	31
40	120	0.418	0.400	−7	12.575	58	0.505	0.488	−67	6.530	−16	0.382	0.374	−40	5.844	11
50	120	0.416	0.397	−11	12.456	58	0.522	0.505	−70	6.310	−15	0.392	0.384	−42	5.536	12
60	120	0.407	0.390	−11	12.201	59	0.547	0.529	−74	6.706	−19	0.411	0.403	−47	5.476	8
70	120	0.407	0.390	−7	12.104	59	0.480	0.464	−64	5.741	−13	0.356	0.349	−39	5.173	12
80	120	0.274	0.262	−9	10.461	60	0.319	0.309	−31	5.409	23	0.257	0.251	16	10.636	70
90	120	0.252	0.241	−10	9.362	52	0.289	0.279	−31	4.594	17	0.195	0.191	9	8.753	58
100	120	0.239	0.229	−13	8.404	46	0.254	0.245	−26	4.423	18	0.182	0.178	13	8.710	57
110	120	0.251	0.240	−15	8.307	46	0.256	0.248	−28	4.442	19	0.174	0.171	11	8.708	57
120	120	0.252	0.241	−16	8.368	47	0.263	0.254	−29	4.585	20	0.171	0.167	9	8.830	58
4 Thin plate regression splines under inverse gaussian with identity link in stagewise selection of length 5																
0	85	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	85	0.622	0.595	33	20.643	108	0.328	0.317	−3	12.034	57	0.488	0.478	68	19.473	129
20	85	0.443	0.423	0	13.176	63	0.412	0.398	−49	6.644	−1	0.336	0.329	−11	8.149	37
30	85	0.390	0.373	−10	12.087	60	0.481	0.465	−65	5.771	−9	0.334	0.327	−33	5.777	23
40	85	0.280	0.268	−9	9.655	48	0.339	0.327	−39	5.079	4	0.255	0.250	1	7.154	44
50	85	0.296	0.283	−10	9.742	48	0.374	0.362	−48	4.933	−3	0.242	0.237	−10	5.768	34
60	85	0.310	0.297	−14	10.405	54	0.367	0.354	−48	4.592	6	0.232	0.227	−8	7.165	46
70	85	0.272	0.260	−12	10.279	58	0.313	0.303	−34	5.205	22	0.249	0.244	12	10.286	67
80	85	0.247	0.236	−14	8.583	48	0.293	0.283	−33	4.594	15	0.217	0.213	10	8.776	58
85	85	0.250	0.239	−17	8.739	50	0.325	0.314	−38	4.585	14	0.218	0.213	6	8.871	58
4 Thin plate regression splines under inverse gaussian with log link in stagewise selection of length 5																
0	75	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	75	0.778	0.744	14	21.780	95	0.446	0.431	40	20.520	106	0.756	0.740	104	25.969	170
20	75	0.491	0.470	−1	14.542	69	0.452	0.437	−55	6.759	0	0.362	0.355	−17	8.423	38
30	75	0.425	0.407	−7	13.142	62	0.472	0.456	−60	6.123	−5	0.366	0.358	−27	6.854	27
40	75	0.406	0.388	−7	12.151	54	0.499	0.482	−66	6.757	−19	0.389	0.381	−41	5.920	7
50	75	0.412	0.394	−11	12.543	56	0.513	0.495	−69	6.309	−16	0.396	0.388	−42	5.655	10
60	75	0.298	0.285	−12	9.519	47	0.392	0.379	−50	5.298	−4	0.265	0.260	−10	6.172	36
70	75	0.263	0.251	−13	9.789	56	0.298	0.288	−31	5.406	23	0.227	0.222	16	10.673	70
75	75	0.258	0.246	−14	9.181	52	0.300	0.290	−33	5.049	19	0.223	0.219	13	9.837	65
4 Thin plate regression splines under inverse gaussian with $\frac{1}{\mu^2}$ link in stagewise selection of length 5																
0	55	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	55	0.803	0.768	2	23.425	117	0.383	0.370	−24	15.197	76	0.435	0.426	27	19.713	127
20	55	0.448	0.428	8	12.645	61	0.331	0.320	−29	7.088	10	0.330	0.323	18	9.983	56
30	55	0.387	0.370	1	12.458	64	0.331	0.320	−29	6.701	20	0.311	0.304	22	11.099	70
40	55	0.341	0.326	−5	11.661	61	0.339	0.328	−35	5.920	17	0.271	0.266	11	9.851	63
45	55	0.343	0.328	−9	10.928	55	0.361	0.349	−38	6.111	12	0.300	0.294	9	9.451	59
50	55	0.336	0.321	−7	10.645	55	0.355	0.343	−40	5.319	8	0.250	0.245	7	8.525	54
55	55	0.328	0.314	−9	10.595	56	0.328	0.317	−35	5.325	15	0.241	0.236	16	10.249	67

Table A20: Out-of-sample validation figures of selected GAMs of BEL with varying random component link function combination and fixed spline function number of 4 per dimension under between 40-443 and 150-443 after each tenth and the finally selected smooth function. MAEs in %.

k	$K_{\max}$	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
8 Thin plate regression splines under normal with identity link																
0	150	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	150	0.639	0.611	27	23.176	125	0.340	0.329	−3	15.517	80	0.516	0.505	73	23.627	156
20	150	0.375	0.359	3	9.604	26	0.334	0.322	−33	8.378	−24	0.341	0.333	1	7.711	10
30	150	0.361	0.345	−7	10.444	41	0.415	0.401	−52	6.961	−19	0.304	0.297	−21	5.871	13
40	150	0.356	0.340	−5	10.098	36	0.425	0.410	−54	7.920	−28	0.311	0.304	−27	5.647	−1
50	150	0.339	0.324	−7	9.712	33	0.418	0.404	−53	7.746	−27	0.311	0.304	−26	5.596	0
60	150	0.325	0.311	−6	9.037	26	0.411	0.397	−52	8.706	−34	0.310	0.304	−26	5.850	−8
70	150	0.325	0.311	−4	9.180	31	0.429	0.414	−55	8.773	−34	0.326	0.319	−30	5.912	−9
80	150	0.309	0.296	−5	8.618	29	0.430	0.415	−55	8.984	−35	0.336	0.329	−29	6.382	−9
90	150	0.313	0.299	−5	8.981	32	0.384	0.371	−48	7.390	−26	0.300	0.293	−26	5.430	−4
100	150	0.328	0.313	−6	9.910	47	0.400	0.387	−51	5.572	−12	0.291	0.285	−25	5.064	13
110	150	0.256	0.245	−10	7.985	38	0.326	0.315	−40	4.655	−6	0.201	0.197	−6	5.002	28
120	150	0.253	0.242	−9	7.340	30	0.321	0.310	−39	5.542	−14	0.209	0.204	−5	4.541	20
130	150	0.252	0.241	−9	7.767	34	0.326	0.315	−40	5.197	−11	0.205	0.201	−5	4.770	24
140	150	0.245	0.234	−8	7.592	33	0.322	0.311	−41	5.315	−15	0.197	0.193	−7	4.317	20
150	150	0.217	0.208	−11	6.477	32	0.239	0.231	−26	3.652	2	0.179	0.175	6	5.578	34
8 Thin plate regression splines under normal with log link in stagewise selection of length 5																
0	50	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	50	0.757	0.724	10	21.570	101	0.444	0.429	39	22.141	116	0.755	0.739	106	27.693	182
20	50	0.401	0.383	1	10.278	23	0.359	0.347	−35	9.154	−28	0.362	0.354	−1	8.110	7
30	50	0.396	0.379	−5	11.249	43	0.438	0.424	−53	7.692	−20	0.339	0.332	−19	6.803	14
40	50	0.382	0.365	−5	11.036	45	0.470	0.454	−60	7.846	−25	0.351	0.344	−31	6.234	4
50	50	0.370	0.353	−8	10.487	39	0.464	0.448	−60	8.000	−28	0.340	0.333	−32	5.901	0
8 Thin plate regression splines under gamma with identity link in stagewise selection of length 5																
0	100	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	100	0.637	0.609	29	22.743	123	0.334	0.323	−3	14.941	77	0.510	0.500	72	22.871	151
20	100	0.370	0.354	4	9.537	27	0.324	0.313	−31	8.076	−22	0.340	0.333	1	7.725	10
30	100	0.359	0.344	−8	10.558	44	0.414	0.400	−52	6.415	−15	0.305	0.298	−22	5.909	16
40	100	0.329	0.314	−9	9.643	37	0.402	0.388	−51	6.673	−21	0.321	0.314	−26	5.702	4
50	100	0.342	0.327	−7	9.631	33	0.409	0.395	−52	7.553	−27	0.326	0.320	−28	5.863	−3
60	100	0.324	0.310	−6	9.114	28	0.409	0.395	−52	8.421	−32	0.327	0.320	−28	6.067	−9
70	100	0.328	0.314	−6	9.617	41	0.451	0.435	−59	7.631	−26	0.349	0.342	−35	5.796	−2
80	100	0.270	0.258	−9	7.944	37	0.324	0.313	−38	5.068	−7	0.221	0.217	−2	5.461	29
90	100	0.279	0.267	−10	8.926	47	0.341	0.329	−40	4.595	2	0.224	0.219	−2	6.713	41
100	100	0.272	0.260	−11	8.654	44	0.335	0.324	−40	4.532	0	0.216	0.211	−2	6.397	38
8 Thin plate regression splines under gamma with log link in stagewise selection of length 5																
0	110	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	110	0.762	0.729	13	21.360	95	0.458	0.443	45	21.527	112	0.773	0.756	108	26.743	176
20	110	0.442	0.422	2	12.416	49	0.396	0.382	−44	7.515	−12	0.349	0.342	−8	8.083	24
30	110	0.387	0.370	−3	11.147	45	0.414	0.400	−49	7.058	−16	0.338	0.331	−18	6.847	16
40	110	0.372	0.356	−6	10.826	43	0.458	0.442	−59	7.546	−24	0.360	0.352	−34	6.225	1
50	110	0.357	0.342	−9	10.240	36	0.458	0.443	−60	7.977	−29	0.357	0.349	−36	6.073	−5
60	110	0.351	0.336	−5	9.866	30	0.439	0.424	−56	9.066	−36	0.353	0.346	−35	6.537	−15
70	110	0.354	0.339	−5	10.130	37	0.458	0.442	−59	8.442	−31	0.364	0.356	−37	6.271	−9
80	110	0.359	0.344	−6	10.122	37	0.463	0.447	−60	8.529	−32	0.371	0.363	−37	6.412	−9
90	110	0.282	0.270	−10	9.017	47	0.364	0.352	−44	4.991	−2	0.249	0.244	−6	6.286	36
100	110	0.268	0.256	−11	7.807	37	0.320	0.309	−38	4.748	−5	0.209	0.204	−1	5.604	32
110	110	0.259	0.247	−11	7.373	34	0.312	0.302	−37	4.801	−7	0.201	0.197	0	5.354	31

Table A21: Out-of-sample validation figures of selected GAMs of BEL with varying random component link function combination and fixed spline function number of 8 per dimension under between 50-443 and 150-443 after each tenth and the finally selected smooth function. MAEs in %.

k	$K_{\max}$	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
8 Thin plate regression splines under normal with log link																
0	25	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	25	0.663	0.634	26	23.298	123	0.341	0.330	1	16.218	84	0.547	0.536	78	24.370	161
20	25	0.398	0.381	2	10.221	23	0.361	0.349	−35	9.380	−28	0.375	0.367	−1	8.460	6
25	25	0.411	0.393	2	11.892	47	0.410	0.397	−47	7.709	−17	0.324	0.317	−11	7.120	19
8 Thin plate regression splines under normal with log link in stagewise selection of length 5																
0	50	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	50	0.757	0.724	10	21.570	101	0.444	0.429	39	22.141	116	0.755	0.739	106	27.693	182
20	50	0.401	0.383	1	10.278	23	0.359	0.347	−35	9.154	−28	0.362	0.354	−1	8.110	7
30	50	0.396	0.379	−5	11.249	43	0.438	0.424	−53	7.692	−20	0.339	0.332	−19	6.803	14
40	50	0.382	0.365	−5	11.036	45	0.470	0.454	−60	7.846	−25	0.351	0.344	−31	6.234	4
50	50	0.370	0.353	−8	10.487	39	0.464	0.448	−60	8.000	−28	0.340	0.333	−32	5.901	0
8 Thin plate regression splines under gamma with identity link																
0	71	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	71	0.637	0.609	29	22.743	123	0.334	0.323	−3	14.941	77	0.510	0.500	72	22.871	151
20	71	0.386	0.369	8	10.141	31	0.310	0.299	−26	7.904	−18	0.358	0.350	8	8.140	16
30	71	0.359	0.344	−8	10.558	44	0.414	0.400	−52	6.415	−15	0.305	0.298	−22	5.909	16
40	71	0.329	0.314	−9	9.643	37	0.402	0.388	−51	6.673	−21	0.321	0.314	−26	5.702	4
50	71	0.338	0.324	−7	9.543	32	0.412	0.399	−53	7.748	−28	0.324	0.318	−29	5.805	−4
60	71	0.324	0.310	−6	9.114	28	0.409	0.395	−52	8.421	−32	0.327	0.320	−28	6.067	−9
70	71	0.327	0.313	−5	9.417	36	0.434	0.419	−56	8.017	−29	0.342	0.335	−32	5.967	−5
71	71	0.291	0.278	−4	8.639	41	0.341	0.329	−43	5.205	−12	0.196	0.192	−17	3.898	14
8 Thin plate regression splines under gamma with identity link in stagewise selection of length 5																
0	100	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	100	0.637	0.609	29	22.743	123	0.334	0.323	−3	14.941	77	0.510	0.500	72	22.871	151
20	100	0.370	0.354	4	9.537	27	0.324	0.313	−31	8.076	−22	0.340	0.333	1	7.725	10
30	100	0.359	0.344	−8	10.558	44	0.414	0.400	−52	6.415	−15	0.305	0.298	−22	5.909	16
40	100	0.329	0.314	−9	9.643	37	0.402	0.388	−51	6.673	−21	0.321	0.314	−26	5.702	4
50	100	0.342	0.327	−7	9.631	33	0.409	0.395	−52	7.553	−27	0.326	0.320	−28	5.863	−3
60	100	0.324	0.310	−6	9.114	28	0.409	0.395	−52	8.421	−32	0.327	0.320	−28	6.067	−9
70	100	0.328	0.314	−6	9.617	41	0.451	0.435	−59	7.631	−26	0.349	0.342	−35	5.796	−2
80	100	0.270	0.258	−9	7.944	37	0.324	0.313	−38	5.068	−7	0.221	0.217	−2	5.461	29
90	100	0.279	0.267	−10	8.926	47	0.341	0.329	−40	4.595	2	0.224	0.219	−2	6.713	41
100	100	0.272	0.260	−11	8.654	44	0.335	0.324	−40	4.532	0	0.216	0.211	−2	6.397	38

Table A22: Out-of-sample validation figures of selected GAMs of BEL in adaptive forward stepwise and stagewise selection of length 5 under between 25-443 and 100-443 after each tenth and the finally selected smooth function. MAEs in %.

k	$K_{\max}$	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
5 Eilers and Marx style P-splines under normal with identity link																
0	100	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	100	0.643	0.615	29	22.836	123	0.344	0.332	−9	13.951	70	0.471	0.461	65	21.854	144
20	100	0.389	0.372	1	10.496	37	0.365	0.353	−41	7.778	−20	0.336	0.329	−8	7.402	13
30	100	0.384	0.367	−9	11.377	53	0.459	0.444	−60	6.138	−13	0.320	0.313	−30	5.512	17
40	100	0.371	0.354	−10	10.977	49	0.454	0.439	−60	6.095	−16	0.327	0.320	−34	5.092	11
50	100	0.357	0.341	−9	10.459	45	0.467	0.451	−62	6.909	−22	0.335	0.328	−34	5.059	6
60	100	0.339	0.324	−10	9.932	43	0.492	0.476	−66	7.640	−28	0.365	0.357	−40	5.155	−2
70	100	0.343	0.328	−10	10.523	52	0.546	0.527	−75	7.681	−27	0.366	0.358	−46	4.576	2
80	100	0.334	0.319	−7	9.920	45	0.520	0.503	−67	8.655	−29	0.346	0.339	−36	5.036	1
90	100	0.228	0.218	−10	6.973	35	0.279	0.269	−31	4.299	0	0.208	0.204	3	5.810	34
100	100	0.225	0.215	−11	6.897	34	0.256	0.248	−30	3.716	2	0.164	0.161	1	5.212	32
8 Eilers and Marx style P-splines under inverse gaussian with $\frac{1}{\mu^2}$ link in dynamically stagewise selection of prop. 0.25																
0	91	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
5	91	1.574	1.505	−18	41.688	233	0.732	0.708	−75	30.201	161	0.384	0.376	42	42.135	278
11	91	0.817	0.781	−3	22.381	113	0.396	0.383	−34	13.475	68	0.412	0.404	23	19.322	124
21	91	0.679	0.650	−9	24.203	138	0.763	0.738	−102	8.222	31	0.424	0.415	−44	13.548	89
37	91	0.525	0.502	1	15.485	79	0.521	0.504	−63	6.154	0	0.397	0.389	−30	7.461	33
62	91	0.505	0.482	−1	14.208	64	0.507	0.490	−61	6.842	−10	0.418	0.410	−33	7.405	18
91	91	0.309	0.296	−11	9.688	45	0.335	0.324	−36	5.239	6	0.279	0.273	2	7.420	43
10 Eilers and Marx style P-splines under normal with identity link in stagewise selection of length 5																
0	150	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	150	0.648	0.619	27	23.688	128	0.349	0.337	−7	15.566	80	0.506	0.495	71	23.889	158
20	150	0.398	0.380	1	10.946	45	0.358	0.346	−37	7.063	−7	0.338	0.331	1	8.102	31
30	150	0.393	0.376	−9	11.983	59	0.435	0.421	−55	5.575	−2	0.299	0.293	−17	6.928	36
40	150	0.371	0.355	−8	11.374	55	0.449	0.434	−57	5.738	−9	0.314	0.308	−26	5.770	23
50	150	0.363	0.347	−9	10.956	50	0.460	0.444	−60	6.249	−14	0.315	0.308	−28	5.492	17
60	150	0.349	0.334	−8	10.479	46	0.443	0.428	−56	6.526	−17	0.305	0.298	−26	5.427	14
70	150	0.349	0.333	−6	10.629	51	0.464	0.449	−60	6.687	−17	0.325	0.318	−29	5.501	13
80	150	0.350	0.335	−7	10.465	48	0.468	0.452	−60	7.036	−19	0.335	0.328	−29	5.563	11
90	150	0.350	0.335	−7	10.639	51	0.470	0.454	−60	6.683	−17	0.330	0.323	−29	5.453	14
100	150	0.334	0.319	−8	9.960	46	0.468	0.452	−60	7.170	−20	0.339	0.332	−29	5.835	11
110	150	0.337	0.323	−9	10.249	48	0.450	0.435	−58	6.171	−15	0.329	0.322	−31	5.267	12
120	150	0.339	0.324	−7	10.283	45	0.433	0.419	−55	6.420	−17	0.320	0.313	−28	5.340	10
130	150	0.269	0.257	−13	8.912	43	0.365	0.352	−46	4.891	−4	0.244	0.238	−12	5.503	30
140	150	0.255	0.244	−12	8.157	36	0.356	0.344	−44	5.415	−10	0.246	0.241	−10	5.196	24
150	150	0.261	0.250	−12	8.514	39	0.368	0.355	−46	5.267	−9	0.245	0.240	−12	5.162	25

Table A23: Out-of-sample validation figures of selected GAMs of BEL with varying spline function number per dimension and fixed spline function type under between 91-443 and 150-443 after each tenth and the finally selected smooth function or after each dynamically stagewise selected smooth function block. Furthermore, a variation in the random component link function combination. MAEs in %.

k	$K_{\max}$	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
4 Thin plate regression splines under normal with identity link																
150	150	0.240	0.229	−15	8.192	46	0.291	0.281	−35	3.907	13	0.176	0.172	3	7.641	50
5 Thin plate regression splines under normal with identity link																
100	100	0.287	0.274	−11	9.431	48	0.397	0.383	−50	5.402	−5	0.202	0.198	−9	5.945	36
8 Thin plate regression splines under normal with identity link																
150	150	0.217	0.208	−11	6.477	32	0.239	0.231	−26	3.652	2	0.179	0.175	6	5.578	34
10 Thin plate regression splines under normal with identity link																
150	150	0.212	0.203	−10	7.070	37	0.230	0.223	−24	3.575	8	0.173	0.170	8	6.337	40
5 Cubic regression splines under normal with identity link																
100	100	0.268	0.256	−12	9.903	52	0.399	0.386	−51	5.182	−2	0.226	0.221	−9	6.533	40
5 Duchon splines under normal with identity link																
56	100	0.666	0.636	−18	18.532	86	0.288	0.279	−14	14.643	75	0.406	0.397	40	19.757	129
5 Eilers and Marx style P-splines under normal with identity link																
100	100	0.225	0.215	−11	6.897	34	0.256	0.248	−30	3.716	2	0.164	0.161	1	5.212	32
10 Cubic regression splines under normal with identity link																
125	125	0.254	0.243	−7	7.139	31	0.299	0.289	−36	5.189	−13	0.197	0.192	−6	4.228	17
10 Duchon splines under normal with identity link																
53	100	0.821	0.785	−44	21.348	94	0.545	0.526	−61	12.593	62	0.446	0.437	−8	18.091	116
10 Eilers and Marx style P-splines under normal with identity link in stagewise selection of length 5																
150	150	0.261	0.250	−12	8.514	−39	0.368	0.355	−46	5.267	9	0.245	0.240	−12	5.162	−25
8 Thin plate regression splines under normal with log link																
25	25	0.411	0.393	2	11.892	47	0.410	0.397	−47	7.709	−17	0.324	0.317	−11	7.120	19
8 Thin plate regression splines under normal with log link in stagewise selection of length 5																
50	50	0.370	0.353	−8	10.487	39	0.464	0.448	−60	8.000	−28	0.340	0.333	−32	5.901	0
8 Thin plate regression splines under gamma with identity link																
71	71	0.291	0.278	−4	8.639	41	0.341	0.329	−43	5.205	−12	0.196	0.192	−17	3.898	14
8 Thin plate regression splines under gamma with identity link in stagewise selection of length 5																
100	100	0.272	0.260	−11	8.654	44	0.335	0.324	−40	4.532	0	0.216	0.211	−2	6.397	38
4 Thin plate regression splines under normal with identity link in stagewise selection of length 5																
150	150	0.240	0.229	−15	8.192	46	0.291	0.281	−35	3.907	13	0.176	0.172	3	7.641	50
4 Thin plate regression splines under normal with log link in stagewise selection of length 5																
40	40	0.438	0.419	−7	13.382	66	0.524	0.506	−69	6.189	−10	0.373	0.365	−39	5.913	20
4 Thin plate regression splines under gamma with identity link in stagewise selection of length 5																
70	70	0.270	0.259	−16	9.999	57	0.325	0.314	−36	5.280	23	0.245	0.240	10	10.416	69
4 Thin plate regression splines under normal with log link in stagewise selection of length 5																
120	120	0.252	0.241	−16	8.368	47	0.263	0.254	−29	4.585	20	0.171	0.167	9	8.830	58
4 Thin plate regression splines under inverse gaussian with identity link in stagewise selection of length 5																
85	85	0.250	0.239	−17	8.739	50	0.325	0.314	−38	4.585	14	0.218	0.213	6	8.871	58
4 Thin plate regression splines under inverse gaussian with log link in stagewise selection of length 5																
75	75	0.258	0.246	−14	9.181	52	0.300	0.290	−33	5.049	19	0.223	0.219	13	9.837	65
4 Thin plate regression splines under inverse gaussian with $\frac{1}{\mu^2}$ link in stagewise selection of length 5																
55	55	0.328	0.314	−9	10.595	56	0.328	0.317	−35	5.325	15	0.241	0.236	16	10.249	67
8 Thin plate regression splines under gamma with log link in stagewise selection of length 5																
110	110	0.259	0.247	−11	7.373	34	0.312	0.302	−37	4.801	−7	0.201	0.197	0	5.354	31
8 Eilers and Marx style P-splines under inverse gaussian with $\frac{1}{\mu^2}$ link in dynamic stagewise selection of proportion 0.25																
91	91	0.309	0.296	−11	9.688	45	0.335	0.324	−36	5.239	6	0.279	0.273	2	7.420	43

Table A24: Maximum allowed numbers of smooth functions and out-of-sample validation figures of all derived GAMs of BEL under between 25-443 and 150-443 after the final iteration. MAEs in %. Highlighted in green and red respectively the best and worst validation figures.

m	r_m^1	r_m^2	r_m^3	r_m^4	r_m^5	r_m^6	r_m^7	r_m^8	r_m^9	r_m^{10}	r_m^{11}	r_m^{12}	r_m^{13}	r_m^{14}	r_m^{15}	**BP.p-val**	**AIC**	**v.mae**	**ns.mae**	**cr.mae**
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1_{-20}	325,850	0.247	0.271	0.122
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1_{-20}	322,452	0.238	0.246	0.122
2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1_{-20}	315,980	0.239	0.255	0.153
3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1_{-20}	314,077	0.237	0.226	0.165
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1_{-20}	312,280	0.231	0.206	0.184
5	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1_{-20}	312,114	0.231	0.205	0.185
6	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1_{-20}	311,949	0.231	0.203	0.186
7	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1_{-20}	311,794	0.232	0.202	0.187
8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1_{-20}	311,700	0.235	0.200	0.190
9	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1_{-20}	311,610	0.233	0.198	0.190
10	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1_{-20}	311,363	0.227	0.194	0.195
11	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1_{-20}	311,293	0.229	0.194	0.197
12	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1_{-20}	311,237	0.228	0.193	0.198
13	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1_{-20}	311,196	0.230	0.193	0.198
14	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1.5_{-20}	311,161	0.231	0.193	0.200
15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	7.1_{-19}	311,136	0.231	0.191	0.202
16	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	5_{-15}	311,091	0.228	0.189	0.201
17	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	5.8_{-13}	311,067	0.228	0.188	0.203
18	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	8.3_{-13}	311,048	0.228	0.187	0.204
19	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	3.2_{-12}	311,030	0.228	0.188	0.204
20	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2.7_{-12}	311,003	0.230	0.188	0.205
21	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1.3_{-11}	310,988	0.230	0.188	0.206
22	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	9.4_{-11}	310,974	0.230	0.187	0.207

Table A25: FGLS variance models of BEL corresponding to $M_{\max} \in \{2, 6, 10, 14, 18, 22\}$ derived by adaptive selection from the set of basis functions of the 150-443 OLS proxy function given in Table A2 with exponents summing up to at max two. Furthermore, p-values of Breusch-Pagan test, AIC scores and out-of-sample MAEs in % after each iteration.

m	r_m^1	r_m^2	r_m^3	r_m^4	r_m^5	r_m^6	r_m^7	r_m^8	r_m^9	r_m^{10}	r_m^{11}	r_m^{12}	r_m^{13}	r_m^{14}	r_m^{15}	**BP.p-val**	**AIC**	**v.mae**	**ns.mae**	**cr.mae**
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1_{-20}	325,459	0.194	0.268	0.168
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1_{-20}	322,077	0.199	0.273	0.166
2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1_{-20}	315,615	0.196	0.275	0.175
3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1_{-20}	313,659	0.195	0.255	0.175
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1_{-20}	311,864	0.198	0.239	0.182
5	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1_{-20}	311,704	0.198	0.236	0.182
6	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1_{-20}	311,554	0.200	0.240	0.183
7	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1_{-20}	311,454	0.199	0.241	0.183
8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1_{-20}	311,360	0.199	0.238	0.186
9	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1_{-20}	311,318	0.201	0.236	0.188
10	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1_{-20}	311,287	0.203	0.234	0.189
11	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1_{-20}	311,260	0.203	0.233	0.189
12	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1_{-20}	311,237	0.203	0.232	0.189
13	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3.7_{-17}	311,001	0.200	0.223	0.192
14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1.7_{-16}	310,980	0.200	0.222	0.194
15	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	7.6_{-13}	310,934	0.200	0.220	0.196
16	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	4.2_{-11}	310,912	0.200	0.218	0.197
17	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1.3_{-10}	310,895	0.200	0.219	0.198
18	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	2.3_{-10}	310,881	0.200	0.217	0.198
19	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	7.6_{-10}	310,867	0.200	0.218	0.197
20	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	3.4_{-9}	310,854	0.200	0.218	0.196
21	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	9.9_{-9}	310,843	0.200	0.218	0.196
22	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	3.1_{-8}	310,832	0.200	0.217	0.196

Table A26: FGLS variance models of BEL corresponding to $M_{\max} \in \{2, 6, 10, 14, 18, 22\}$ derived by adaptive selection from the set of basis functions of the 300-886 OLS proxy function given in Table A4 with exponents summing up to at max two. Furthermore, p-values of Breusch-Pagan test, AIC scores and out-of-sample MAEs in % after each iteration.

m	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
0	0.247	0.237	−14	9.924	57	0.271	0.262	−35	4.612	22	0.122	0.120	−1	8.537	56
1	0.238	0.228	−15	8.668	49	0.246	0.238	−30	4.120	19	0.122	0.120	3	7.873	52
2	0.239	0.229	−16	8.147	46	0.255	0.246	−30	4.032	17	0.153	0.149	2	7.489	49
3	0.237	0.226	−15	7.789	43	0.226	0.218	−24	4.423	20	0.165	0.162	10	8.117	54
4	0.231	0.221	−13	7.684	42	0.206	0.199	−18	4.817	22	0.184	0.180	17	8.756	58
5	0.231	0.221	−13	7.666	42	0.205	0.198	−18	4.803	22	0.185	0.181	17	8.740	58
6	0.231	0.221	−13	7.577	41	0.203	0.196	−18	4.762	22	0.186	0.183	17	8.637	57
7	0.232	0.222	−12	7.661	42	0.202	0.195	−17	4.787	22	0.187	0.183	18	8.691	57
8	0.235	0.225	−12	7.774	42	0.200	0.193	−17	4.914	23	0.190	0.186	19	8.912	59
9	0.233	0.223	−11	7.692	42	0.198	0.191	−16	4.838	23	0.190	0.186	19	8.763	58
10	0.227	0.217	−10	7.460	40	0.194	0.188	−15	4.708	21	0.195	0.191	20	8.537	56
11	0.229	0.219	−10	7.447	40	0.194	0.187	−15	4.686	21	0.197	0.193	20	8.455	56
12	0.228	0.218	−10	7.426	40	0.193	0.186	−14	4.687	21	0.198	0.194	20	8.444	56
13	0.230	0.220	−9	7.513	41	0.193	0.187	−14	4.696	21	0.198	0.194	21	8.491	56
14	0.231	0.221	−9	7.527	41	0.193	0.186	−14	4.701	21	0.200	0.195	21	8.497	56
15	0.231	0.221	−9	7.523	41	0.191	0.185	−13	4.742	21	0.202	0.197	22	8.569	57
16	0.228	0.218	−9	7.437	40	0.189	0.182	−13	4.730	21	0.201	0.197	22	8.557	56
17	0.228	0.218	−9	7.421	40	0.188	0.182	−13	4.747	21	0.203	0.199	22	8.568	56
18	0.228	0.218	−9	7.433	40	0.187	0.181	−13	4.780	22	0.204	0.200	22	8.621	57
19	0.228	0.218	−9	7.435	40	0.188	0.182	−13	4.786	22	0.204	0.200	22	8.628	57
20	0.230	0.219	−9	7.442	40	0.188	0.182	−13	4.796	22	0.205	0.201	22	8.650	57
21	0.230	0.220	−9	7.466	40	0.188	0.181	−13	4.800	22	0.206	0.201	23	8.648	57
22	0.230	0.220	−8	7.436	40	0.187	0.180	−12	4.802	22	0.207	0.203	23	8.639	57

Table A27: Iteration-wise out-of-sample validation figures in adaptive variance model selection of BEL corresponding to $M_{\max} \in \{2, 6, 10, 14, 18, 22\}$ based on the 150-443 OLS proxy function given in Table A2 with exponents summing up to at max two. MAEs in %. Simultaneously type I FGLS regression results.

m	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
0	0.194	0.186	−9	6.659	34	0.268	0.259	−30	4.200	−2	0.168	0.165	1	5.007	29
1	0.199	0.190	−9	6.648	34	0.273	0.263	−31	4.272	−3	0.166	0.162	1	5.005	30
2	0.196	0.187	−9	6.527	33	0.275	0.266	−30	4.564	−3	0.175	0.171	5	5.401	32
3	0.195	0.186	−9	6.487	33	0.255	0.247	−27	4.350	1	0.175	0.171	9	5.916	37
4	0.198	0.189	−9	6.305	32	0.239	0.231	−23	4.262	4	0.182	0.178	13	6.303	40
5	0.198	0.190	−9	6.298	32	0.236	0.228	−22	4.252	4	0.182	0.178	14	6.336	40
6	0.200	0.191	−9	6.399	33	0.240	0.232	−23	4.292	4	0.183	0.179	13	6.389	40
7	0.199	0.190	−9	6.364	32	0.241	0.233	−23	4.304	4	0.183	0.179	13	6.324	40
8	0.199	0.190	−8	6.381	32	0.238	0.230	−22	4.313	4	0.186	0.182	14	6.407	40
9	0.201	0.193	−8	6.432	33	0.236	0.228	−22	4.313	5	0.188	0.184	15	6.521	41
10	0.203	0.194	−8	6.473	33	0.234	0.226	−21	4.310	5	0.189	0.185	16	6.621	42
11	0.203	0.195	−8	6.492	33	0.233	0.225	−21	4.303	5	0.189	0.185	16	6.628	42
12	0.203	0.194	−8	6.476	33	0.232	0.224	−21	4.294	5	0.189	0.186	16	6.641	42
13	0.200	0.191	−7	6.254	32	0.223	0.216	−19	4.252	5	0.192	0.188	17	6.615	42
14	0.200	0.191	−7	6.246	31	0.222	0.214	−19	4.257	6	0.194	0.190	18	6.697	42
15	0.200	0.191	−7	6.216	31	0.220	0.213	−18	4.243	6	0.196	0.192	19	6.773	43
16	0.200	0.191	−7	6.180	31	0.218	0.211	−18	4.239	6	0.197	0.193	19	6.753	43
17	0.200	0.192	−7	6.197	31	0.219	0.211	−18	4.249	6	0.198	0.194	19	6.804	43
18	0.200	0.191	−7	6.194	31	0.217	0.210	−18	4.250	6	0.198	0.194	19	6.801	43
19	0.200	0.191	−7	6.207	31	0.218	0.210	−18	4.238	6	0.197	0.193	19	6.787	43
20	0.200	0.191	−7	6.229	32	0.218	0.211	−18	4.226	6	0.196	0.192	19	6.793	43
21	0.200	0.192	−7	6.240	32	0.218	0.211	−18	4.224	7	0.196	0.192	19	6.814	43
22	0.200	0.192	−7	6.256	32	0.217	0.210	−18	4.223	7	0.196	0.192	19	6.844	44

Table A28: Iteration-wise out-of-sample validation figures in adaptive variance model selection of BEL corresponding to $M_{\max} \in \{2, 6, 10, 14, 18, 22\}$ based on the 300-886 OLS proxy function given in Table A4 with exponents summing up to at max two. MAEs in %. Simultaneously type I FGLS regression results.

k	AIC	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
$M_{\max} = 2$ in variance model selection																
0	437,251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	336,390	1.786	1.708	184	44.082	198	1.402	1.354	209	39.152	209	2.290	2.242	344	52.033	344
20	323,883	0.826	0.790	25	22.007	111	0.424	0.409	−28	10.764	44	0.437	0.428	28	16.424	99
30	319,958	0.465	0.445	3	12.876	55	0.288	0.278	2	9.650	40	0.467	0.457	57	15.234	96
40	318,945	0.401	0.384	−16	11.036	51	0.357	0.345	−37	7.158	16	0.330	0.323	3	10.127	55
50	318,206	0.355	0.339	−24	9.270	35	0.336	0.324	−36	6.611	8	0.339	0.332	−8	8.602	36
60	317,485	0.323	0.309	−25	8.407	36	0.309	0.298	−36	5.548	11	0.279	0.273	−11	7.244	36
70	317,197	0.306	0.293	−28	7.631	28	0.345	0.334	−43	5.405	−1	0.272	0.266	−17	5.899	25
80	316,263	0.272	0.260	−24	6.946	32	0.320	0.310	−42	4.051	0	0.227	0.222	−17	4.898	25
90	316,021	0.260	0.249	−23	7.143	39	0.298	0.288	−37	3.854	10	0.173	0.169	−5	6.461	42
100	315,871	0.256	0.245	−23	7.424	41	0.294	0.284	−35	4.078	14	0.186	0.182	0	7.443	49
110	315,784	0.256	0.245	−22	7.396	41	0.302	0.292	−37	3.962	12	0.189	0.185	−3	7.013	46
120	315,719	0.257	0.245	−23	6.923	38	0.296	0.286	−36	3.870	11	0.181	0.177	−2	6.872	45
130	315,675	0.258	0.247	−25	6.506	35	0.295	0.285	−36	3.760	9	0.188	0.184	−3	6.461	42
140	315,649	0.252	0.241	−23	6.424	34	0.283	0.274	−34	3.749	9	0.184	0.180	−1	6.399	42
150	315,629	0.239	0.229	−21	6.467	34	0.261	0.252	−30	3.796	10	0.177	0.173	3	6.654	44
$M_{\max} = 6$ in variance model selection																
0	437,251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	332,479	2.014	1.926	259	49.098	213	2.000	1.933	298	44.745	238	2.964	2.901	445	58.341	385
20	320,873	0.881	0.842	51	22.821	115	0.341	0.329	16	13.428	66	0.622	0.609	84	20.790	134
30	316,187	0.429	0.410	19	10.875	32	0.308	0.297	29	8.537	28	0.561	0.549	73	12.633	72
40	315,132	0.366	0.350	6	10.243	45	0.254	0.246	1	7.853	25	0.401	0.393	36	11.221	61
50	314,473	0.303	0.289	3	9.346	46	0.229	0.222	0	7.543	28	0.361	0.353	34	10.776	62
60	313,643	0.307	0.293	−18	7.567	28	0.251	0.242	−21	5.808	11	0.266	0.261	9	7.676	41
70	313,301	0.280	0.268	−17	7.768	30	0.222	0.214	−12	6.229	21	0.268	0.262	23	9.315	56
80	313,060	0.270	0.258	−20	7.092	28	0.230	0.222	−13	6.273	22	0.280	0.274	25	9.554	59
90	312,883	0.262	0.251	−22	6.754	29	0.239	0.231	−17	5.977	20	0.253	0.248	19	9.077	56
100	312,100	0.246	0.235	−19	6.177	29	0.202	0.195	−14	4.814	18	0.221	0.216	21	8.305	54
110	311,656	0.231	0.221	−16	6.446	33	0.189	0.182	−12	4.827	22	0.211	0.206	25	8.964	59
120	311,574	0.236	0.225	−16	6.545	34	0.209	0.202	−16	4.594	19	0.207	0.202	22	8.637	57
130	311,511	0.238	0.227	−17	6.551	35	0.207	0.200	−16	4.797	21	0.204	0.200	23	9.104	60
140	311,461	0.231	0.221	−16	6.026	31	0.189	0.183	−12	4.726	21	0.216	0.212	25	8.853	58
150	311,426	0.224	0.215	−14	5.904	31	0.177	0.171	−9	4.756	22	0.226	0.221	29	9.005	59
$M_{\max} = 10$ in variance model selection																
0	437,251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	328,519	2.120	2.027	288	50.524	221	2.206	2.132	329	46.563	248	3.194	3.127	480	60.396	399
20	319,481	0.971	0.928	95	24.185	105	0.439	0.424	53	11.839	49	0.821	0.803	117	18.086	112
30	316,529	0.655	0.627	56	16.560	74	0.420	0.406	57	12.301	61	0.780	0.764	113	18.285	117
40	314,460	0.379	0.362	19	10.089	42	0.268	0.259	19	8.120	28	0.473	0.463	54	11.608	63
50	313,842	0.324	0.310	2	8.422	33	0.229	0.221	−4	6.420	12	0.339	0.331	20	8.600	36
60	313,022	0.297	0.284	−13	7.619	31	0.223	0.215	−13	6.123	17	0.277	0.271	14	8.292	43
70	312,692	0.282	0.269	−17	7.494	26	0.221	0.213	−5	6.762	24	0.326	0.319	35	10.467	64
80	312,443	0.271	0.259	−19	7.171	27	0.218	0.211	−7	6.625	25	0.303	0.297	33	10.306	65
90	312,264	0.261	0.249	−21	6.610	27	0.222	0.215	−11	6.300	23	0.278	0.272	28	9.806	62
100	312,187	0.262	0.250	−21	6.568	26	0.216	0.208	−10	6.265	23	0.272	0.266	28	9.707	61
110	312,108	0.256	0.244	−21	6.031	23	0.203	0.196	−5	6.324	25	0.288	0.282	31	9.754	61
120	312,043	0.261	0.250	−23	5.989	20	0.200	0.194	−4	6.287	25	0.293	0.287	33	9.857	62
130	311,078	0.226	0.216	−18	5.466	25	0.160	0.155	−4	5.115	24	0.244	0.239	32	9.192	60
140	310,918	0.220	0.210	−16	5.451	25	0.153	0.148	−4	4.820	23	0.233	0.228	31	8.859	58
150	310,868	0.212	0.203	−14	5.375	25	0.148	0.143	0	5.098	25	0.256	0.250	36	9.296	61

Table A29: AIC scores and out-of-sample validation figures of type II FGLS proxy functions of BEL under 150-443 with variance models of varying complexity $M_{\max}$ after each tenth iteration. MAEs in %.

k	AIC	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
$M_{\max}=14$ in variance model selection																
0	437,251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	326,308	2.120	2.027	290	50.306	220	2.215	2.141	331	46.129	246	3.197	3.130	480	59.909	396
20	319,199	1.024	0.979	100	26.049	137	0.527	0.509	75	18.639	98	1.044	1.022	155	27.142	178
30	316,093	0.702	0.671	67	17.574	79	0.503	0.486	73	13.745	70	0.901	0.882	133	20.208	131
40	314,155	0.393	0.376	24	10.363	44	0.282	0.273	25	8.426	31	0.505	0.494	62	12.131	68
50	313,562	0.327	0.313	6	8.561	34	0.225	0.217	1	6.535	15	0.352	0.345	27	8.936	41
60	312,811	0.298	0.285	−10	7.608	29	0.203	0.196	4	7.086	29	0.336	0.329	37	10.283	62
70	312,455	0.289	0.276	−15	7.409	26	0.219	0.211	−2	6.863	25	0.343	0.335	38	10.612	65
80	312,235	0.273	0.261	−17	7.222	28	0.215	0.208	−4	6.738	26	0.322	0.316	37	10.662	67
90	312,057	0.264	0.253	−22	6.680	27	0.222	0.214	−10	6.406	24	0.283	0.277	28	9.981	63
100	311,953	0.255	0.244	−21	6.117	24	0.201	0.194	−5	6.381	25	0.290	0.284	31	9.780	61
110	311,898	0.252	0.241	−20	5.929	22	0.200	0.193	−4	6.236	24	0.293	0.287	32	9.583	60
120	311,832	0.263	0.251	−23	5.962	19	0.198	0.192	−3	6.300	25	0.303	0.296	34	9.878	62
130	310,916	0.223	0.213	−17	5.363	23	0.154	0.149	−1	5.233	25	0.263	0.257	36	9.305	61
140	310,757	0.215	0.206	−15	5.339	24	0.147	0.142	0	4.954	24	0.251	0.246	35	8.972	59
150	310,714	0.214	0.205	−14	5.368	25	0.146	0.141	−1	4.857	23	0.244	0.239	34	8.906	59
$M_{\max}=18$ in variance model selection																
0	437,251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	326,125	2.127	2.034	292	50.425	220	2.226	2.151	332	46.222	246	3.209	3.142	482	60.019	396
20	318,762	1.036	0.991	111	25.668	113	0.538	0.520	75	13.429	64	0.983	0.962	144	20.708	133
30	315,995	0.710	0.679	69	17.741	80	0.523	0.505	76	13.963	72	0.925	0.906	137	20.465	133
40	314,060	0.401	0.383	27	10.529	45	0.292	0.282	28	8.560	33	0.521	0.510	66	12.341	70
50	313,483	0.329	0.315	9	8.687	35	0.225	0.217	4	6.620	16	0.362	0.354	31	9.120	43
60	312,938	0.316	0.302	−5	7.840	30	0.209	0.202	5	6.855	26	0.347	0.340	41	10.297	62
70	312,363	0.270	0.258	−10	6.960	21	0.215	0.207	11	7.089	28	0.389	0.381	48	10.795	65
80	312,166	0.259	0.248	−12	6.558	22	0.204	0.198	9	7.008	29	0.369	0.361	47	10.718	67
90	311,963	0.234	0.223	−15	6.141	24	0.196	0.189	1	6.432	26	0.313	0.306	37	9.844	61
100	311,883	0.241	0.231	−18	6.031	24	0.194	0.187	−1	6.449	26	0.299	0.293	34	9.777	61
110	311,830	0.239	0.229	−18	5.836	22	0.193	0.187	0	6.298	25	0.303	0.296	35	9.610	60
120	311,766	0.244	0.234	−19	5.713	18	0.191	0.184	3	6.340	26	0.321	0.314	39	9.866	62
130	311,045	0.225	0.215	−15	5.396	23	0.148	0.143	0	5.061	24	0.259	0.254	35	8.950	59
140	310,694	0.213	0.204	−13	5.314	24	0.139	0.134	1	4.855	24	0.245	0.240	34	8.672	57
150	310,644	0.211	0.202	−14	5.131	23	0.139	0.135	1	4.816	23	0.250	0.245	35	8.618	57
$M_{\max}=22$ in variance model selection																
0	437,251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	325,988	2.127	2.034	292	50.414	220	2.226	2.151	332	46.259	246	3.210	3.143	482	60.061	397
20	318,926	1.034	0.988	105	26.160	137	0.569	0.550	83	19.043	101	1.098	1.075	163	27.621	181
30	315,805	0.712	0.681	71	17.763	79	0.537	0.519	78	14.063	72	0.943	0.923	140	20.603	134
40	313,973	0.409	0.391	29	10.730	46	0.301	0.291	31	8.709	34	0.539	0.527	70	12.589	72
50	313,411	0.349	0.334	7	8.950	34	0.223	0.216	3	6.618	16	0.357	0.349	30	9.081	42
60	312,873	0.308	0.295	−2	8.205	37	0.203	0.196	8	7.490	33	0.350	0.343	43	10.853	67
70	312,286	0.271	0.260	−9	6.950	21	0.217	0.210	12	7.124	28	0.398	0.389	50	10.856	66
80	312,091	0.261	0.249	−11	6.557	22	0.207	0.200	10	7.051	29	0.377	0.369	48	10.793	68
90	311,893	0.235	0.225	−15	6.043	23	0.196	0.189	1	6.367	25	0.314	0.307	36	9.683	60
100	311,815	0.238	0.228	−17	5.970	23	0.194	0.187	1	6.462	26	0.311	0.304	37	9.829	61
110	311,761	0.237	0.227	−17	5.780	21	0.194	0.188	2	6.364	25	0.313	0.307	37	9.694	60
120	311,697	0.243	0.232	−19	5.818	18	0.191	0.185	2	6.325	25	0.320	0.313	39	9.885	62
130	311,655	0.232	0.222	−17	5.688	18	0.195	0.188	8	6.714	29	0.353	0.346	46	10.509	67
140	310,748	0.215	0.206	−14	5.206	23	0.148	0.143	5	5.578	27	0.293	0.287	42	9.788	64
150	310,590	0.208	0.199	−13	5.209	23	0.139	0.134	5	5.193	26	0.275	0.270	40	9.256	61

Table A29: Cont.

k	AIC	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
$M_{\max} = 2$ in variance model selection																
0	437, 251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	336, 390	1.786	1.708	184	44.082	198	1.402	1.354	209	39.152	209	2.290	2.242	344	52.033	344
20	323, 883	0.826	0.790	25	22.007	111	0.424	0.409	−28	10.764	44	0.437	0.428	28	16.424	99
30	319, 958	0.465	0.445	3	12.876	55	0.288	0.278	2	9.650	40	0.467	0.457	57	15.234	96
40	318, 945	0.401	0.384	−16	11.036	51	0.357	0.345	−37	7.158	16	0.330	0.323	3	10.127	55
50	318, 206	0.355	0.339	−24	9.270	35	0.336	0.324	−36	6.611	8	0.339	0.332	−8	8.602	36
60	317, 485	0.323	0.309	−25	8.407	36	0.309	0.298	−36	5.548	11	0.279	0.273	−11	7.244	36
70	317, 197	0.306	0.293	−28	7.631	28	0.345	0.334	−43	5.405	−1	0.272	0.266	−17	5.899	25
80	316, 263	0.272	0.260	−24	6.946	32	0.320	0.310	−42	4.051	0	0.227	0.222	−17	4.898	25
90	316, 021	0.260	0.249	−23	7.143	39	0.298	0.288	−37	3.854	10	0.173	0.169	−5	6.461	42
100	315, 871	0.256	0.245	−23	7.424	41	0.294	0.284	−35	4.078	14	0.186	0.182	0	7.443	49
110	315, 784	0.256	0.245	−22	7.396	41	0.302	0.292	−37	3.962	12	0.189	0.185	−3	7.013	46
120	315, 719	0.257	0.245	−23	6.923	38	0.296	0.286	−36	3.870	11	0.181	0.177	−2	6.872	45
130	315, 675	0.258	0.247	−25	6.506	35	0.295	0.285	−36	3.760	9	0.188	0.184	−3	6.461	42
140	315, 641	0.250	0.239	−23	6.441	34	0.284	0.275	−34	3.741	9	0.182	0.178	−2	6.338	41
150	315, 622	0.238	0.228	−20	6.433	34	0.258	0.250	−29	3.821	11	0.177	0.174	4	6.740	44
160	315, 599	0.233	0.223	−20	6.578	35	0.256	0.247	−28	3.920	12	0.183	0.179	6	6.988	46
170	315, 573	0.232	0.222	−19	6.616	35	0.254	0.246	−28	3.880	12	0.181	0.178	5	6.927	45
180	315, 535	0.225	0.215	−19	6.502	35	0.252	0.243	−28	3.773	11	0.172	0.169	5	6.797	44
190	315, 523	0.229	0.219	−19	6.809	37	0.244	0.236	−26	4.020	15	0.164	0.161	9	7.607	50
200	315, 507	0.215	0.206	−18	6.738	36	0.243	0.235	−26	3.969	14	0.164	0.161	9	7.387	49
210	315, 500	0.214	0.205	−18	6.704	35	0.234	0.226	−24	3.989	14	0.162	0.159	10	7.323	48
220	315, 492	0.217	0.207	−18	6.769	35	0.239	0.231	−26	3.930	14	0.159	0.155	9	7.277	48
224	315, 491	0.209	0.199	−17	6.584	34	0.226	0.219	−22	3.999	14	0.165	0.161	12	7.290	48
$M_{\max} = 6$ in variance model selection																
0	437, 251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	332, 479	2.014	1.926	259	49.098	213	2.000	1.933	298	44.745	238	2.964	2.901	445	58.341	385
20	320, 873	0.881	0.842	51	22.821	115	0.341	0.329	16	13.428	66	0.622	0.609	84	20.790	134
30	316, 187	0.429	0.410	19	10.875	32	0.308	0.297	29	8.537	28	0.561	0.549	73	12.633	72
40	315, 132	0.366	0.350	6	10.243	45	0.254	0.246	1	7.853	25	0.401	0.393	36	11.221	61
50	314, 473	0.303	0.289	3	9.346	46	0.229	0.222	0	7.543	28	0.361	0.353	34	10.776	62
60	313, 643	0.307	0.293	−18	7.567	28	0.251	0.242	−21	5.808	11	0.266	0.261	9	7.676	41
70	313, 301	0.280	0.268	−17	7.768	30	0.222	0.214	−12	6.229	21	0.268	0.262	23	9.315	56
80	313, 060	0.270	0.258	−20	7.092	28	0.230	0.222	−13	6.273	22	0.280	0.274	25	9.554	59
90	312, 883	0.262	0.251	−22	6.754	29	0.239	0.231	−17	5.977	20	0.253	0.248	19	9.077	56
100	312, 100	0.246	0.235	−19	6.177	29	0.202	0.195	−14	4.814	18	0.221	0.216	21	8.305	54
110	311, 656	0.231	0.221	−16	6.446	33	0.189	0.182	−12	4.827	22	0.211	0.206	25	8.964	59
120	311, 574	0.236	0.225	−16	6.545	34	0.209	0.202	−16	4.594	19	0.207	0.202	22	8.637	57
130	311, 507	0.234	0.223	−16	6.706	36	0.206	0.199	−16	4.801	21	0.204	0.200	23	9.094	60
140	311, 456	0.226	0.216	−16	6.102	32	0.189	0.182	−12	4.717	21	0.215	0.211	25	8.827	58
150	311, 419	0.224	0.214	−15	5.899	31	0.178	0.172	−10	4.712	22	0.213	0.209	27	8.971	59
160	311, 355	0.217	0.207	−15	5.536	29	0.160	0.154	−4	5.013	25	0.246	0.241	33	9.420	62
170	311, 308	0.198	0.189	−13	5.090	23	0.141	0.137	−4	4.144	19	0.221	0.216	27	7.491	49
180	311, 266	0.202	0.193	−14	5.112	24	0.132	0.127	−3	4.433	22	0.218	0.213	27	7.868	52
190	311, 248	0.208	0.198	−16	5.287	23	0.143	0.138	−5	4.163	19	0.213	0.208	25	7.630	50
200	311, 228	0.202	0.193	−14	5.269	24	0.137	0.133	−4	4.148	20	0.213	0.209	27	7.639	50
210	311, 196	0.192	0.184	−14	5.032	20	0.125	0.121	4	4.655	23	0.253	0.248	32	7.919	52
220	311, 164	0.195	0.187	−15	5.079	21	0.122	0.118	1	4.620	23	0.237	0.232	31	8.070	53
230	311, 148	0.194	0.185	−15	5.146	22	0.122	0.118	1	4.571	23	0.236	0.231	29	7.949	52
237	311, 144	0.196	0.188	−15	5.342	23	0.125	0.121	0	4.765	24	0.235	0.230	30	8.243	54
$M_{\max} = 10$ in variance model selection																
0	437, 251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	331, 056	2.073	1.982	273	50.085	216	2.113	2.041	315	45.714	244	3.090	3.025	464	59.451	393
20	320, 199	0.924	0.884	76	23.133	101	0.375	0.362	25	10.921	35	0.655	0.641	82	15.999	92
30	316, 044	0.543	0.519	31	14.068	56	0.372	0.359	45	11.729	56	0.742	0.727	107	18.450	118
40	314, 821	0.385	0.368	11	10.626	47	0.256	0.248	6	8.118	28	0.424	0.415	43	11.685	65
50	314, 201	0.327	0.313	2	9.206	41	0.240	0.232	−8	6.713	17	0.336	0.329	21	9.103	45
60	313, 386	0.269	0.257	−5	7.831	34	0.220	0.213	6	7.506	31	0.365	0.357	46	11.223	71
70	312, 986	0.290	0.278	−17	7.316	26	0.210	0.203	−4	6.646	25	0.310	0.304	33	9.955	61
80	312, 722	0.280	0.268	−18	7.425	31	0.223	0.215	−8	6.792	27	0.300	0.293	33	10.652	68
90	312, 545	0.270	0.259	−22	7.110	32	0.233	0.225	−13	6.634	26	0.273	0.267	27	10.450	67
100	312, 469	0.265	0.253	−21	6.800	29	0.224	0.217	−11	6.420	25	0.274	0.268	29	10.128	64
110	312, 397	0.254	0.243	−19	6.136	25	0.202	0.195	−4	6.360	25	0.290	0.284	33	9.940	63
120	312, 346	0.247	0.236	−19	5.940	22	0.193	0.187	1	6.468	27	0.307	0.301	38	10.078	64
130	312, 299	0.240	0.230	−17	5.784	21	0.192	0.185	4	6.563	28	0.329	0.322	43	10.369	66
140	312, 274	0.247	0.236	−18	5.811	22	0.193	0.186	5	6.870	31	0.338	0.331	45	10.944	71
150	312, 243	0.249	0.238	−19	5.950	24	0.193	0.186	3	6.872	31	0.324	0.317	43	10.984	71
160	312, 222	0.255	0.244	−19	6.162	25	0.198	0.191	1	6.859	30	0.324	0.318	42	11.092	72
170	311, 204	0.228	0.218	−14	5.957	31	0.161	0.156	−1	5.874	30	0.276	0.270	40	10.703	71
180	311, 040	0.223	0.213	−13	6.021	31	0.154	0.149	−1	5.594	29	0.265	0.259	39	10.356	68
190	310, 996	0.222	0.213	−13	6.152	32	0.154	0.149	−2	5.584	28	0.258	0.253	38	10.311	68
200	310, 968	0.206	0.197	−10	6.163	32	0.144	0.139	3	5.924	31	0.285	0.279	42	10.568	70
210	310, 953	0.211	0.202	−10	5.930	30	0.143	0.138	3	5.615	29	0.276	0.270	41	10.153	67
220	310, 927	0.208	0.199	−11	6.353	33	0.147	0.142	−1	5.602	29	0.252	0.247	37	10.225	67
230	310, 919	0.211	0.202	−11	6.454	34	0.149	0.144	−1	5.702	29	0.259	0.253	38	10.376	69
240	310, 908	0.210	0.201	−11	6.559	35	0.152	0.147	−3	5.570	28	0.251	0.245	36	10.218	67
244	310, 905	0.208	0.199	−11	6.577	35	0.153	0.147	−2	5.617	29	0.252	0.247	37	10.259	68

Table A30: AIC scores and out-of-sample validation figures of type II FGLS proxy functions of BEL under 300-886 with variance models of varying complexity $M_{\max}$ after each tenth and the final iteration. MAEs in %.

k	AIC	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
$M_{\max} = 14$ in variance model selection																
0	437, 251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	327, 049	2.133	2.039	292	50.561	222	2.233	2.157	333	46.686	249	3.222	3.154	484	60.524	400
20	318, 965	1.020	0.976	108	25.288	111	0.507	0.490	69	12.759	57	0.931	0.912	136	19.634	124
30	316, 262	0.694	0.663	65	17.386	78	0.484	0.468	69	13.341	68	0.872	0.853	128	19.643	127
40	314, 272	0.392	0.375	23	10.373	44	0.277	0.268	23	8.322	30	0.493	0.483	59	11.941	66
50	313, 691	0.349	0.333	1	8.772	32	0.228	0.220	−5	6.440	12	0.335	0.328	19	8.633	36
60	312, 860	0.289	0.276	−10	7.475	30	0.204	0.197	−2	6.583	24	0.302	0.295	28	9.218	53
70	312, 542	0.286	0.273	−16	7.501	26	0.219	0.211	−3	6.802	24	0.334	0.327	37	10.548	64
80	312, 337	0.281	0.269	−18	7.254	27	0.215	0.207	−4	6.834	27	0.323	0.316	37	10.655	67
90	312, 126	0.261	0.250	−21	6.672	27	0.221	0.213	−10	6.384	23	0.286	0.280	29	9.942	62
100	312, 046	0.268	0.256	−22	6.695	27	0.222	0.215	−12	6.317	24	0.270	0.265	26	9.779	61
110	311, 961	0.257	0.245	−22	5.979	23	0.200	0.193	−5	6.316	25	0.284	0.278	31	9.695	61
120	311, 903	0.252	0.241	−21	5.892	19	0.193	0.186	1	6.411	26	0.311	0.304	37	9.977	63
130	311, 860	0.244	0.233	−19	5.886	20	0.190	0.184	3	6.562	28	0.322	0.315	41	10.344	66
140	311, 824	0.243	0.232	−20	5.880	19	0.190	0.183	5	6.758	30	0.335	0.328	44	10.696	69
150	311, 800	0.247	0.236	−21	6.011	20	0.185	0.179	2	6.452	28	0.309	0.303	40	10.365	66
160	310, 806	0.218	0.208	−16	5.451	25	0.140	0.135	0	5.234	27	0.255	0.249	37	9.596	63
170	310, 710	0.210	0.201	−15	5.473	25	0.137	0.132	0	5.077	26	0.249	0.244	36	9.359	62
180	310, 682	0.206	0.197	−14	5.303	24	0.136	0.131	2	5.064	26	0.266	0.260	39	9.492	63
190	310, 661	0.200	0.191	−13	5.285	23	0.144	0.139	5	5.163	26	0.298	0.292	44	9.843	65
200	310, 639	0.201	0.192	−13	5.413	22	0.143	0.138	4	5.088	25	0.293	0.287	44	9.726	64
210	310, 606	0.203	0.194	−13	5.599	23	0.145	0.141	6	5.459	27	0.314	0.307	47	10.294	68
220	310, 525	0.183	0.174	−13	4.672	12	0.148	0.143	−3	3.744	7	0.221	0.217	30	6.238	40
230	310, 513	0.179	0.171	−14	4.668	13	0.153	0.148	−6	3.729	7	0.206	0.202	27	6.113	40
240	310, 475	0.172	0.164	−14	4.347	10	0.130	0.126	−1	3.523	9	0.219	0.214	30	6.154	39
250	310, 462	0.171	0.163	−14	4.307	10	0.134	0.130	−2	3.480	8	0.211	0.206	28	5.958	38
258	310, 443	0.172	0.165	−14	4.371	10	0.134	0.129	−2	3.504	8	0.214	0.210	28	6.063	39
$M_{\max} = 18$ in variance model selection																
0	437, 251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	325, 846	2.112	2.020	290	50.142	221	2.201	2.127	328	46.153	246	3.183	3.116	478	59.925	396
20	318, 985	1.027	0.982	104	25.991	136	0.566	0.547	82	18.748	99	1.089	1.066	162	27.261	179
30	315, 896	0.705	0.674	69	17.595	79	0.526	0.508	76	13.871	71	0.928	0.908	137	20.356	132
40	314, 044	0.404	0.386	28	10.602	45	0.296	0.286	30	8.630	34	0.531	0.519	68	12.462	71
50	313, 483	0.330	0.316	9	8.715	35	0.225	0.217	5	6.643	17	0.365	0.358	32	9.177	44
60	312, 939	0.316	0.302	−5	7.833	31	0.210	0.203	5	6.895	26	0.352	0.345	42	10.382	63
70	312, 359	0.270	0.258	−10	6.927	21	0.216	0.208	11	7.084	27	0.393	0.385	49	10.781	65
80	312, 165	0.260	0.248	−12	6.555	22	0.206	0.199	10	7.018	29	0.373	0.365	48	10.721	67
90	311, 964	0.233	0.223	−15	6.130	24	0.196	0.189	1	6.433	26	0.313	0.307	37	9.838	61
100	311, 882	0.237	0.227	−17	5.756	20	0.190	0.183	2	6.218	24	0.305	0.298	36	9.431	58
110	311, 827	0.239	0.229	−18	5.733	21	0.190	0.184	1	6.305	25	0.303	0.296	36	9.588	60
120	311, 769	0.245	0.234	−20	5.762	18	0.189	0.183	3	6.425	27	0.319	0.313	39	9.924	62
130	311, 716	0.224	0.214	−16	5.502	15	0.190	0.183	10	6.403	27	0.350	0.342	46	9.993	63
140	311, 005	0.216	0.206	−13	5.222	21	0.142	0.137	6	5.361	26	0.291	0.285	42	9.416	62
150	310, 660	0.203	0.194	−12	5.094	21	0.133	0.129	7	5.158	26	0.284	0.278	42	9.129	60
160	310, 611	0.201	0.192	−12	5.033	21	0.137	0.133	8	5.360	27	0.303	0.297	45	9.568	63
170	310, 586	0.196	0.187	−11	4.994	21	0.136	0.132	10	5.548	28	0.316	0.310	47	9.821	65
180	310, 550	0.193	0.184	−12	4.987	21	0.135	0.130	1	4.264	20	0.241	0.236	35	8.200	54
190	310, 535	0.196	0.187	−14	5.087	21	0.139	0.135	−3	4.049	18	0.217	0.212	31	7.884	52
200	310, 511	0.182	0.174	−11	4.965	21	0.131	0.127	0	3.992	18	0.231	0.226	34	7.810	52
210	310, 467	0.185	0.177	−12	5.011	20	0.131	0.127	0	3.967	17	0.231	0.226	34	7.741	51
220	310, 463	0.181	0.173	−12	5.059	20	0.130	0.125	2	4.181	19	0.246	0.241	36	8.110	54
230	310, 454	0.181	0.173	−11	5.409	23	0.138	0.133	1	4.405	20	0.246	0.241	36	8.436	56
240	310, 440	0.182	0.174	−11	5.398	23	0.138	0.133	1	4.457	21	0.250	0.245	37	8.559	57
250	310, 431	0.181	0.173	−11	5.509	23	0.138	0.133	1	4.525	21	0.251	0.246	37	8.638	57
252	310, 425	0.185	0.176	−11	5.515	23	0.138	0.133	1	4.548	22	0.253	0.248	37	8.700	57
$M_{\max} = 22$ in variance model selection																
0	437, 251	4.557	4.357	−238	100.000	38	3.231	3.121	0	100.000	261	4.027	3.942	106	100.000	367
10	325, 796	2.115	2.023	290	50.203	222	2.206	2.131	329	46.238	246	3.189	3.121	479	60.021	396
20	318, 940	1.026	0.981	112	25.965	135	0.666	0.644	98	20.243	107	1.199	1.174	179	28.606	188
30	315, 849	0.708	0.677	70	17.681	79	0.532	0.514	77	14.005	72	0.936	0.917	139	20.526	133
40	314, 001	0.407	0.389	28	10.712	46	0.299	0.289	31	8.710	34	0.536	0.524	69	12.589	73
50	313, 413	0.348	0.332	10	9.025	36	0.223	0.216	5	6.616	17	0.364	0.356	32	9.225	44
60	312, 897	0.316	0.302	−4	7.866	31	0.211	0.203	6	6.983	27	0.358	0.351	44	10.549	65
70	312, 317	0.271	0.259	−9	6.969	22	0.217	0.210	12	7.185	28	0.399	0.391	50	10.961	67
80	312, 120	0.260	0.249	−11	6.565	23	0.207	0.200	10	7.119	30	0.379	0.371	49	10.896	69
90	311, 920	0.235	0.224	−15	6.091	24	0.196	0.189	1	6.427	26	0.313	0.306	37	9.791	61
100	311, 842	0.238	0.228	−16	6.034	23	0.194	0.187	1	6.531	27	0.311	0.304	37	9.949	63
110	311, 784	0.241	0.230	−18	5.900	24	0.192	0.185	1	6.554	28	0.304	0.297	36	10.004	63
120	311, 737	0.241	0.230	−18	5.809	21	0.189	0.182	2	6.395	27	0.310	0.303	38	9.924	63
130	311, 690	0.227	0.217	−16	5.653	18	0.187	0.181	8	6.468	28	0.339	0.332	45	10.100	64
140	310, 925	0.213	0.203	−13	5.206	22	0.140	0.136	7	5.430	27	0.293	0.286	43	9.548	63
150	310, 604	0.202	0.193	−11	5.131	22	0.133	0.129	7	5.286	27	0.289	0.283	42	9.321	61
160	310, 559	0.200	0.192	−11	5.063	22	0.139	0.134	9	5.507	28	0.310	0.304	46	9.791	65
170	310, 532	0.189	0.181	−10	4.999	22	0.134	0.129	8	5.194	26	0.297	0.291	44	9.438	62
180	310, 503	0.193	0.185	−12	5.222	24	0.132	0.128	4	5.137	26	0.270	0.264	40	9.462	62
190	310, 481	0.194	0.186	−13	5.113	22	0.140	0.136	−2	4.124	19	0.220	0.215	32	8.019	53
200	310, 454	0.189	0.181	−13	5.164	21	0.135	0.130	−1	4.033	18	0.224	0.220	33	7.836	52
210	310, 412	0.185	0.177	−12	5.038	20	0.132	0.128	0	4.019	18	0.231	0.226	34	7.805	52
220	310, 406	0.185	0.176	−12	5.067	20	0.132	0.128	1	4.062	18	0.239	0.234	35	7.981	53
224	310, 404	0.184	0.176	−12	5.112	20	0.132	0.128	1	4.076	18	0.239	0.234	35	7.934	52

Table A30: Cont.

k	$M_{\max}$	AIC	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
Type I algorithm under 150-443																	
150	2	315,980	0.239	0.229	−16	8.147	46	0.255	0.246	−30	4.032	17	0.153	0.149	2	7.489	49
150	6	311,949	0.231	0.221	−13	7.577	41	0.203	0.196	−18	4.762	22	0.186	0.183	17	8.637	57
150	10	311,363	0.227	0.217	−10	7.460	40	0.194	0.188	−15	4.708	21	0.195	0.191	20	8.537	56
150	14	311,161	0.231	0.221	−9	7.527	41	0.193	0.186	−14	4.701	21	0.200	0.195	21	8.497	56
150	18	311,048	0.228	0.218	−9	7.433	40	0.187	0.181	−13	4.780	22	0.204	0.200	22	8.621	57
150	22	310,974	0.230	0.220	−8	7.436	40	0.187	0.180	−12	4.802	22	0.207	0.203	23	8.639	57
Type I algorithm under 300-886																	
224	2	315,615	0.196	0.187	−9	6.527	33	0.275	0.266	−30	4.564	−3	0.175	0.171	5	5.401	32
224	6	311,554	0.200	0.191	−9	6.399	33	0.240	0.232	−23	4.292	4	0.183	0.179	13	6.389	40
224	10	311,287	0.203	0.194	−8	6.473	33	0.234	0.226	−21	4.310	5	0.189	0.185	16	6.621	42
224	14	310,980	0.200	0.191	−7	6.246	31	0.222	0.214	−19	4.257	6	0.194	0.190	18	6.697	42
224	18	310,881	0.200	0.191	−7	6.194	31	0.217	0.210	−18	4.250	6	0.198	0.194	19	6.801	43
224	22	310,832	0.200	0.192	−7	6.256	32	0.217	0.210	−18	4.223	7	0.196	0.192	19	6.844	44
Type II algorithm under 150-443																	
150	2	315,629	0.239	0.229	−21	6.467	34	0.261	0.252	−30	3.796	10	0.177	0.173	3	6.654	44
150	6	311,426	0.224	0.215	−14	5.904	31	0.177	0.171	−9	4.756	22	0.226	0.221	29	9.005	59
150	10	310,868	0.212	0.203	−14	5.375	25	0.148	0.143	0	5.098	25	0.256	0.250	36	9.296	61
150	14	310,714	0.214	0.205	−14	5.368	25	0.146	0.141	−1	4.857	23	0.244	0.239	34	8.906	59
150	18	310,644	0.211	0.202	−14	5.131	23	0.139	0.135	1	4.816	23	0.250	0.245	35	8.618	57
150	22	310,590	0.208	0.199	−13	5.209	23	0.139	0.134	5	5.193	26	0.275	0.270	40	9.256	61
Type II algorithm under 300-886																	
224	2	315,491	0.209	0.199	−17	6.584	34	0.226	0.219	−22	3.999	14	0.165	0.161	12	7.290	48
237	6	311,144	0.196	0.188	−15	5.342	23	0.125	0.121	0	4.765	24	0.235	0.230	30	8.243	54
244	10	310,905	0.208	0.199	−11	6.577	35	0.153	0.147	−2	5.617	29	0.252	0.247	37	10.259	68
258	14	310,443	0.172	0.165	−14	4.371	10	0.134	0.129	−2	3.504	8	0.214	0.210	28	6.063	39
252	18	310,425	0.185	0.176	−11	5.515	23	0.138	0.133	1	4.548	22	0.253	0.248	37	8.700	57
224	22	310,404	0.184	0.176	−12	5.112	20	0.132	0.128	1	4.076	18	0.239	0.234	35	7.934	52

Table A31: AIC scores and out-of-sample validation figures of all derived FGLS proxy functions of BEL under 150-443 and 300-886 after the final iteration. MAEs in %. Highlighted in green and red respectively the best and worst AIC scores and validation figures.

k	$K_{\max}$	$t_{\min}$	o	p	glm	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
Sobol set^2																				
148	206	0	6	s	inv.g, id	0.265	0.253	−24	10.317	55	0.575	0.555	−40	16.234	−56	0.822	0.805	80	17.657	64
49	50	0	3	n	inv.g, log	0.370	0.354	0	9.168	19	0.705	0.681	−12	29.477	−102	0.525	0.514	25	16.891	−65
60	66	0	4	s	inv.g, id	0.324	0.310	−11	8.517	16	1.712	1.654	151	44.504	132	0.917	0.897	102	19.877	83
45	50	0	4	b	inv.g, id	0.347	0.332	−2	8.686	11	0.447	0.431	−36	22.702	−125	0.511	0.500	35	15.785	−54
Sobol set and nested simulations set																				
45	50	0	4	b	inv.g, id	0.347	0.332	−2	8.686	11	0.447	0.431	−36	22.702	−125	0.511	0.500	35	15.785	−54
17	19	0	4	b	inv.g, id	0.834	0.797	25	24.673	124	0.480	0.464	−4	41.356	−243	0.763	0.747	108	21.398	−132
70	81	0	4	b	inv.g, id	0.335	0.320	−22	10.872	52	0.554	0.535	−35	14.073	−38	0.875	0.857	102	18.250	99
33	34	0	3	n	inv.g, id	0.426	0.407	−10	10.871	21	1.565	1.512	108	52.384	1	0.662	0.648	32	20.997	−75
Sobol set and capital region set																				
45	50	0	3	b	pois, log	0.379	0.362	0	9.556	28	0.480	0.464	−43	24.878	−139	0.510	0.500	28	16.938	−69
31	34	0	3	b	pois, log	0.476	0.455	−13	12.752	46	0.593	0.573	−54	31.148	−175	0.661	0.647	18	23.088	−103
45	50	0	4	b	inv.g, id	0.347	0.332	−2	8.686	11	0.447	0.431	−36	22.702	−125	0.511	0.500	35	15.785	−54
59	66	0	3	b	pois, log	0.428	0.439	40	16.674	98	0.760	0.734	−12	22.511	−41	0.809	0.792	68	18.403	39
Nested simulations set and Sobol set																				
134	144	1.6_{-5}	5	n	gaus, log	0.273	0.261	−22	10.255	54	1.025	0.990	−1	28.192	−23	1.515	1.484	179	32.616	157
45	50	0	4	s	inv.g, id	0.347	0.332	−2	8.686	11	0.447	0.431	−36	22.702	−125	0.511	0.500	35	15.785	−54
60	66	0	4	s	inv.g, id	0.324	0.310	−11	8.517	16	1.712	1.654	151	44.504	132	0.917	0.897	102	19.877	83
45	50	0	4	b	inv.g, id	0.347	0.332	−2	8.686	11	0.447	0.431	−36	22.702	−125	0.511	0.500	35	15.785	−54
Nested simulations set^2																				
45	50	0	4	b	inv.g, id	0.347	0.332	−2	8.686	11	0.447	0.431	−36	22.702	−125	0.511	0.500	35	15.785	−54
146	159	9.4_{-6}	5	n	gaus, log	0.279	0.267	−24	10.008	53	1.025	0.990	0	26.779	−11	1.498	1.467	174	31.702	163
76	97	3.8_{-5}	4	b	inv.g, log	0.344	0.329	−17	10.676	52	0.538	0.520	−37	11.874	−24	0.804	0.787	88	16.584	100
107	113	0	4	n	gaus, log	0.321	0.307	−20	11.976	63	0.997	0.963	8	25.694	0	1.529	1.496	191	32.148	182
Nested simulations set and capital region set																				
45	50	0	4	s	pois, id	0.353	0.338	−3	8.891	18	0.449	0.434	−36	23.634	−131	0.504	0.493	36	16.079	−58
31	34	0	4	s	pois, id	0.437	0.418	−11	11.254	32	0.548	0.530	−45	28.444	−157	0.648	0.634	29	21.374	−84
72	82	3.1_{-5}	4	b	inv.g, inv	0.365	0.349	−16	11.181	53	0.579	0.560	−49	14.528	−51	0.700	0.685	65	14.619	64
45	50	0	4	b	inv.g, id	0.347	0.332	−2	8.686	11	0.447	0.431	−36	22.702	−125	0.511	0.500	35	15.785	−54
Capital region set and Sobol set																				
125	144	0	5	f	inv.g, inv	0.283	0.271	−20	10.336	54	0.630	0.608	−63	17.245	−76	0.675	0.660	45	14.737	32
45	50	0	4	s	gaus, log	0.382	0.365	−1	9.916	32	0.469	0.453	−41	25.487	−144	0.495	0.485	32	16.868	−71
114	144	1.9_{-5}	5	s	inv.g,$1/\mu^2$	0.313	0.299	−12	9.414	40	0.708	0.684	−77	20.115	−97	0.626	0.612	36	14.095	17
45	50	0	4	b	gaus, log	0.382	0.365	−1	9.916	32	0.469	0.453	−41	25.487	−144	0.495	0.485	32	16.868	−71
Capital region set and nested simulations set																				
45	50	0	4	f	gaus, log	0.386	0.369	−1	10.095	34	0.468	0.452	−41	25.709	−145	0.496	0.486	32	17.077	−73
64	66	0	4	n	inv.g,$1/\mu^2$	0.420	0.401	−3	11.506	39	0.840	0.811	3	25.969	−38	1.298	1.271	146	29.110	105
148	175	0	6	s	inv.g,$1/\mu^2$	0.311	0.297	−16	10.447	52	0.576	0.556	−55	14.565	−57	0.611	0.598	30	12.844	27
77	81	0	4	n	inv.g,$1/\mu^2$	0.387	0.370	−11	11.519	52	1.029	0.994	−28	25.831	−32	1.279	1.252	148	26.700	145
Capital region set^2																				
45	50	0	4	s	gaus, log	0.382	0.365	−1	9.916	32	0.469	0.453	−41	25.487	−144	0.495	0.485	32	16.868	−71
33	34	0	3	n	inv.g,$1/\mu^2$	0.564	0.539	−14	15.693	64	0.827	0.800	−54	38.645	−185	0.745	0.729	−2	26.338	−134
148	175	0	6	s	inv.g,$1/\mu^2$	0.311	0.297	−16	10.447	52	0.576	0.556	−55	14.565	−57	0.611	0.598	30	12.844	27
148	175	4.7_{-6}	5	f	inv.g, inv	0.296	0.283	−20	10.416	53	0.549	0.530	−54	18.260	−87	0.664	0.650	32	16.307	−1

Table A32: Settings and out-of-sample validation figures of best performing MARS models of BEL derived in a two-step approach sorted by first and second step validation sets. MAEs in %. Highlighted in green and red respectively the best and worst validation figures.

k	$h_k(X)$	$\widehat{\beta}_{\mathrm{MARS},k}$
0	1	$15,397.13$
1	$h(X_8 - 0.104892)$	$7,901.89$
2	$h(0.104892 - X_8)$	$-8,165.64$
3	$h(0.205577 - X_1) \cdot h(0.104892 - X_8)$	688.83
4	$h(X_6 - 1.17224)$	265.08
5	$h(1.17224 - X_6)$	-280.94
6	$h(X_{15} - 53.8706)$	-2.11
7	$h(53.8706 - X_{15})$	1.16
8	$h(X_7 - -0.147599)$	-60.90
9	$h(-0.147599 - X_7)$	-334.77
10	$h(X_8 - -0.0456197)$	$3,183.07$
11	$h(0.205577 - X_1) \cdot h(0.104892 - X_8) \cdot h(X_{15} - 64.6262)$	-9.48
12	$h(0.205577 - X_1) \cdot h(0.104892 - X_8) \cdot h(64.6262 - X_{15})$	29.85
13	$h(X_1 - 0.945371)$	-64.88
14	$h(0.945371 - X_1)$	124.45
15	$h(X_6 - 1.56058) \cdot h(0.104892 - X_8)$	-815.20
16	$h(1.56058 - X_6) \cdot h(0.104892 - X_8)$	$1,085.80$
17	$h(1.44218 - X_2)$	-60.23
18	$h(X_1 - -1.61447) \cdot h(1.56058 - X_6) \cdot h(0.104892 - X_8)$	-233.14
19	$h(-1.61447 - X_1) \cdot h(1.56058 - X_6) \cdot h(0.104892 - X_8)$	415.92
20	$h(X_8 - 0.0159508) \cdot h(53.8706 - X_{15})$	8.94
21	$h(0.0159508 - X_8) \cdot h(53.8706 - X_{15})$	47.99
22	$h(X_9 - 0.247192)$	47.72
23	$h(0.247192 - X_9)$	-82.58
24	$h(0.993896 - X_{12})$	-63.61
25	$h(X_1 - 0.0195594) \cdot h(0.0159508 - X_8) \cdot h(53.8706 - X_{15})$	-12.58
26	$h(0.0195594 - X_1) \cdot h(0.0159508 - X_8) \cdot h(53.8706 - X_{15})$	-42.25
27	$h(X_7 - -0.147599) \cdot h(X_8 - -0.191689)$	$2,124.93$
28	$h(X_7 - -0.147599) \cdot h(-0.191689 - X_8)$	$1,510.41$
29	$h(X_3 - 0.323352) \cdot h(0.104892 - X_8)$	948.86
30	$h(0.323352 - X_3) \cdot h(0.104892 - X_8)$	-577.61
31	$h(X_1 - -1.26627) \cdot h(X_7 - -0.147599)$	101.15
32	$h(-1.26627 - X_1) \cdot h(X_7 - -0.147599)$	-10.00
33	$h(X_{14} - 0.684998)$	109.76
34	$h(0.684998 - X_{14})$	-37.89
35	$h(1.17224 - X_6) \cdot h(X_8 - -0.12538)$	216.62
36	$h(1.17224 - X_6) \cdot h(-0.12538 - X_8)$	$2,076.18$
37	$h(0.945371 - X_1) \cdot h(X_8 - 0.0019988)$	-156.79
38	$h(0.945371 - X_1) \cdot h(0.0019988 - X_8)$	$1,262.56$
39	$h(X_1 - -1.58818) \cdot h(X_6 - 1.56058) \cdot h(0.104892 - X_8)$	137.60
40	$h(1.56058 - X_6) \cdot h(0.104892 - X_8) \cdot h(X_{15} - 76.9327)$	-4.87
41	$h(1.56058 - X_6) \cdot h(0.104892 - X_8) \cdot h(76.9327 - X_{15})$	2.11
42	$h(0.205577 - X_1) \cdot h(X_2 - 1.43028) \cdot h(0.104892 - X_8)$	$24,003.07$
43	$h(0.205577 - X_1) \cdot h(1.43028 - X_2) \cdot h(0.104892 - X_8)$	-161.88
44	$h(X_1 - 0.945371) \cdot h(X_8 - -0.0165546)$	-224.18
45	$h(X_1 - 0.945371) \cdot h(-0.0165546 - X_8)$	-987.47

Table A33: Best MARS model of BEL derived in a two-step approach with the final coefficients.

k	r_k^1	r_k^2	r_k^3	r_k^4	r_k^5	r_k^6	r_k^7	r_k^8	r_k^9	r_k^{10}	r_k^{11}	r_k^{12}	r_k^{13}	r_k^{14}	r_k^{15}
$K_{\max} = 16$ in adaptive basis function selection															
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
5	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
6	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
8	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
11	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
12	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
13	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
16	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
$K_{\max} = 27$ in adaptive basis function selection															
17	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
18	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
19	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1
20	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
21	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
23	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1
24	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
25	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
27	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0

Table A34: Basis function sets of LC and LL proxy functions of BEL corresponding to $K_{\max} \in \{16, 27\}$ derived by adaptive OLS selection.

k	r_k^1	r_k^2	r_k^3	r_k^4	r_k^5	r_k^6	r_k^7	r_k^8	r_k^9	r_k^{10}	r_k^{11}	r_k^{12}	r_k^{13}	r_k^{14}	r_k^{15}
$K_{\max} = 15$ in risk factor wise basis function selection															
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
$K_{\max} = 22$ in combined risk factor wise and adaptive basis function selection															
16	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
17	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
19	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
20	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
22	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0

Table A35: Basis function sets of LC and LL proxy functions of BEL corresponding to $K_{\max} \in \{15, 22\}$ derived by risk factor wise or combined risk factor wise and adaptive OLS selection.

k	bw	o	v.mae	v.mae^a	v.res	v.mae^0	v.res^0	ns.mae	ns.mae^a	ns.res	ns.mae^0	ns.res^0	cr.mae	cr.mae^a	cr.res	cr.mae^0	cr.res^0
LC regression with gaussian kernel and LOO-CV																	
16	0.1	2	0.55	0.52	−44	13	50	0.70	0.68	−86	12	−7	0.55	0.54	−35	12	45
16	0.2	2	0.40	0.38	−26	11	47	0.52	0.50	−51	11	7	0.44	0.43	5	13	63
16	0.3	2	0.37	0.35	−25	11	45	0.45	0.44	−37	11	19	0.44	0.43	5	12	60
27	0.2	2	0.39	0.38	−26	11	43	0.51	0.49	−51	11	3	0.43	0.43	4	12	58
16	0.1	4	2.80	2.68	−155	84	−407	8.05	7.78	−558	247	−825	5.04	4.94	−96	128	−363
LL regression with gaussian kernel and LOO-CV																	
16	0.1	2	0.38	0.36	−11	12	57	0.57	0.55	−68	10	−15	0.41	0.40	−22	9	31
16	0.2	2	0.34	0.33	−6	11	59	0.45	0.43	−49	8	2	0.37	0.36	5	10	55
27	0.1	2	210.30	201.06	−30,682	5,209	−30,589	131.04	126.61	−18,981	3,670	−18,902	4.09	4.00	−82	92	−3
27	0.2	2	2,726.47	2,606.74	400,254	67,487	400,306	3,502.24	3,383.85	422,443	98,081	422,481	1.85	1.81	−25	41	13
LC regression with gaussian kernel and AIC																	
16	0.1	2	0.57	0.55	−43	14	55	0.65	0.62	−72	12	12	0.50	0.49	−12	14	72
16	0.2	2	1.63	1.55	38	41	73	1.94	1.88	266	57	286	2.57	2.51	384	61	404
27	0.1	2	0.56	0.54	−42	14	56	0.64	0.62	−72	12	12	0.50	0.49	−12	14	72
LC regression with Epanechnikov kernel and LOO-CV																	
15	0.1	2	0.53	0.50	−36	13	41	1.05	1.02	−38	22	24	0.51	0.50	−29	11	33
15	0.2	2	0.41	0.39	−31	10	33	1.14	1.10	3	26	53	1.18	1.16	97	27	146
15	0.3	2	0.40	0.38	−30	9	23	0.96	0.93	16	23	54	0.46	0.45	−6	11	33
15	0.4	2	0.35	0.33	−22	9	18	1.11	1.08	12	28	39	0.47	0.46	−2	11	25
15	0.5	2	0.34	0.33	−18	9	37	1.24	1.20	6	30	46	0.51	0.50	−22	11	18
15	0.6	2	0.33	0.32	−17	10	50	1.16	1.12	21	27	74	0.46	0.45	−2	11	50
15	0.7	2	0.33	0.32	−16	10	41	1.17	1.13	18	28	61	0.44	0.43	−14	9	28
15	0.8	2	0.33	0.31	−16	10	45	1.21	1.17	29	29	76	1.16	1.13	101	26	148
15	0.9	2	0.32	0.30	−20	12	61	1.14	1.10	40	27	107	1.14	1.11	111	29	178
15	1.0	2	0.32	0.31	−22	10	49	1.19	1.15	52	29	109	1.13	1.11	106	27	163
16	0.1	2	0.53	0.50	−40	13	43	1.20	1.16	2	28	71	0.51	0.50	−20	12	49
16	0.2	2	0.41	0.39	−26	11	50	1.16	1.12	27	28	88	0.44	0.43	2	12	64
16	0.3	2	0.36	0.34	−27	9	29	1.07	1.03	41	27	83	0.44	0.43	1	11	43
16	0.4	2	0.33	0.32	−19	8	22	1.16	1.12	27	30	53	0.45	0.44	4	10	30
16	0.5	2	0.32	0.31	−16	9	36	1.34	1.30	30	33	67	1.22	1.19	101	27	138
16	0.1	4	0.45	0.43	−26	13	34	0.74	0.71	−68	16	−23	0.59	0.57	5	15	51
16	0.2	4	3.29	3.15	−104	160	891	7.50	7.24	−14	329	966	8.06	7.89	176	295	1,157
16	0.1	6	3.31	3.16	−32	84	68	5.74	5.55	−96	158	−10	6.62	6.48	−53	148	32
16	0.2	6	3.32	3.18	−71	85	−217	9.37	9.06	73	268	−87	13.18	12.90	246	304	86
16	0.1	8	3.94	3.77	146	105	−119	10.71	10.35	−191	308	−470	8.84	8.65	−312	205	−591
16	0.2	8	8.53	8.16	397	286	−639	7.79	7.52	70	347	−980	12.37	12.11	1,365	390	315
22	0.1	2	0.50	0.48	−37	12	44	1.07	1.03	−41	22	25	0.52	0.50	−30	11	37
22	0.2	2	0.42	0.40	−28	10	39	1.07	1.03	−3	25	50	1.20	1.17	106	29	159
22	0.3	2	0.39	0.37	−29	9	23	0.89	0.86	6	22	43	0.45	0.44	−3	11	34
22	0.4	2	0.35	0.33	−21	8	16	1.05	1.02	3	27	26	0.49	0.48	−4	11	19
22	0.5	2	0.33	0.31	−14	9	32	1.17	1.13	−2	28	29	0.47	0.46	−15	10	16
22	0.6	2	0.33	0.32	−17	10	46	1.09	1.06	11	25	60	0.45	0.44	−1	11	48
22	0.7	2	0.32	0.31	−15	9	39	1.23	1.18	26	29	66	1.17	1.14	99	26	139
22	0.8	2	0.32	0.30	−15	10	46	1.19	1.15	32	28	78	1.12	1.10	106	26	152
22	0.9	2	0.31	0.30	−19	11	58	1.15	1.11	39	27	102	1.12	1.10	111	28	174
22	1.0	2	0.31	0.30	−21	10	48	1.13	1.09	41	27	96	1.12	1.10	107	27	162
27	0.2	2	0.40	0.38	−26	11	45	1.15	1.12	26	28	83	0.44	0.43	1	12	58
27	0.3	2	0.38	0.36	−28	9	24	0.90	0.87	7	22	45	0.46	0.45	−2	11	36
27	0.4	2	0.35	0.33	−21	9	17	1.05	1.02	2	27	26	0.48	0.47	−4	11	20
LL regression with Epanechnikov kernel and LOO-CV																	
15	0.1	2	0.45	0.43	−49	10	40	1.22	1.18	−100	22	−26	0.78	0.77	−104	11	−30
15	0.2	2	0.36	0.34	−34	8	13	1.59	1.53	−145	40	−112	0.60	0.58	−54	11	−21
15	0.3	2	0.32	0.31	−36	7	17	1.91	1.85	134	48	173	0.60	0.58	−36	11	3
15	0.4	2	0.34	0.33	−40	8	33	1.83	1.76	−164	42	−106	0.43	0.42	−49	6	9
15	0.5	2	0.33	0.31	−40	8	34	2.20	2.12	−219	53	−160	0.41	0.41	−45	6	15
15	0.6	2	0.30	0.29	−33	7	29	0.94	0.91	8	19	56	0.33	0.32	−28	5	21
15	0.7	2	0.31	0.30	−40	7	23	0.94	0.91	−13	19	36	0.36	0.35	−40	5	8
15	0.8	2	0.29	0.28	−38	5	8	0.86	0.83	4	19	36	0.32	0.32	−29	5	3
22	0.1	2	731.51	699.39	2,738	85,172	479,612	1,564.87	1,511.98	−111,628	127,410	365,231	492.49	482.11	−19,404	76,575	457,455
22	0.2	2	0.34	0.33	−34	8	0	0.83	0.80	−15	21	4	0.42	0.41	−25	8	−5
22	0.3	2	98.03	93.73	14,396	148	−250	101.69	98.25	15,174	147	513	100.00	97.89	15,028	100	367
22	0.4	2	98.05	93.75	14,399	147	−248	113.99	110.14	13,158	495	−1,503	100.00	97.89	15,028	100	367
22	0.5	2	100.00	95.61	14,685	100	38	118.95	114.93	14,984	651	323	100.00	97.89	15,028	100	367
22	0.6	2	99.72	95.34	14,644	106	−3	100.59	97.19	15,004	120	343	100.00	97.89	15,028	100	367
22	0.7	2	100.00	95.61	14,685	100	38	100.00	96.62	14,922	100	261	100.00	97.89	15,028	100	367
22	0.8	2	0.29	0.28	−39	5	9	152.43	147.27	22,622	4,264	22,655	0.31	0.30	−35	5	−2
LC regression with uniform kernel and LOO-CV																	
16	0.1	2	0.75	0.71	−56	18	46	1.53	1.48	−52	32	36	0.73	0.72	−59	15	29
16	0.5	2	1.22	1.17	−78	29	16	2.60	2.51	301	82	381	10.45	10.23	1,419	242	1,498
27	0.1	2	0.64	0.61	−38	16	31	1.30	1.26	13	32	68	0.59	0.58	−2	15	53
27	0.5	2	0.35	0.34	−16	12	53	1.34	1.30	25	33	79	1.40	1.37	117	32	171
16	0.1	4	0.71	0.68	−33	17	47	1.27	1.23	−1	31	65	0.67	0.65	−23	15	43
16	0.5	4	1.85	1.76	−139	39	50	2.29	2.22	18	51	193	7.09	6.94	769	157	943
27	0.1	4	0.66	0.63	−38	15	32	1.32	1.27	7	32	63	0.58	0.57	−15	14	40
27	0.5	4	0.39	0.37	−13	13	67	1.26	1.21	16	31	82	0.52	0.51	−10	13	56
16	0.1	6	1.83	1.75	−165	38	100	1.95	1.88	−178	29	72	1.55	1.51	−190	24	60
16	0.5	6	1.83	1.75	−6	56	271	1.08	1.04	80	65	344	1.66	1.63	225	74	488

Table A36: Settings and out-of-sample validation figures of LC and LL proxy functions of BEL using basis function sets from Tables A34 and A35. MAEs in %. Highlighted in green and red respectively the best and worst validation figures.

	Am. - Random Seed			Am. - Same Seed			Eu. - Same Seed		
Scenario	**excl.**	**ext.**	**%**	**excl.**	**ext.**	**%**	**excl.**	**ext.**	**%**
base	121.88	121.07	-0.67	120.75	121.10	0.29	122.25	121.40	-0.69
$r-$	124.60	124.76	0.13	124.04	124.00	-0.03	125.51	123.27	-1.78
$r+$	117.87	117.36	-0.44	118.05	118.19	0.12	119.51	119.52	0.01
$K-$	120.18	120.90	0.60	120.36	120.29	-0.06	121.74	121.73	-0.01
$K+$	121.71	121.34	-0.30	121.23	122.09	0.70	122.46	124.86	1.96
$S-$	97.78	95.44	-2.39	96.70	95.62	-1.11	98.03	97.68	-0.36
$S+$	146.77	146.35	-0.29	144.75	143.22	-1.06	146.69	143.84	-1.94
$\mu-$	121.56	120.18	-1.14	120.83	121.30	0.39	122.13	121.91	-0.18
$\mu+$	121.27	121.46	0.16	120.91	120.91	0.00	122.11	122.19	0.07
$S+, \mu-$	145.61	145.46	-0.10	144.91	143.42	-1.03	146.59	144.35	-1.53
$K-, \mu+$	120.35	121.29	0.78	120.22	120.09	-0.10	121.97	122.52	0.45
$r-, S+$	148.98	149.82	0.56	148.44	145.99	-1.65	150.46	145.03	-3.61
$r-, \mu-$	124.20	123.87	-0.27	123.78	124.20	0.34	125.56	123.78	-1.42
$r+, \mu-$	119.14	116.47	-2.24	118.22	118.39	0.14	119.60	120.03	0.36
$K-, S-$	120.96	121.29	0.27	120.29	120.09	-0.17	121.80	122.52	0.60
$r-, K+$	124.84	125.62	0.62	124.19	124.67	0.39	125.74	126.22	0.39
$K+, S-$	97.35	95.72	-1.68	96.81	96.61	-0.21	98.19	101.14	3.01
$K+, S+, \mu-$	145.94	145.74	-0.14	145.36	144.40	-0.66	147.08	147.81	0.50
$r-, S+, \mu+$	149.17	150.21	0.70	148.72	145.79	-1.97	150.51	145.83	-3.11
$r-, K-, S-, \mu-$	98.60	97.73	-0.89	98.84	98.35	-0.50	100.09	101.57	1.48
$r-, K-, S-, \mu+$	99.55	99.01	-0.54	98.85	97.96	-0.90	100.06	101.85	1.79
$r-, K-, S+, \mu-$	148.78	148.18	-0.40	148.20	145.68	-1.70	150.29	146.38	-2.60
$r-, K+, S-, \mu-$	100.52	99.32	-1.19	99.50	99.52	0.02	100.81	103.68	2.85
$r+, K-, S-, \mu-$	95.01	91.03	-4.18	94.07	91.65	-2.58	95.53	95.45	-0.08
$r+, K+, S+, \mu+$	143.47	142.96	-0.36	142.00	141.55	-0.32	143.70	147.41	2.58
$r+, K+, S+, \mu-$	143.23	141.68	-1.08	142.01	141.94	-0.05	143.72	147.13	2.37
$r+, K+, S-, \mu+$	94.66	91.59	-3.24	94.71	93.68	-1.08	95.87	99.90	4.20
$r+, K-, S+, \mu+$	142.54	143.68	0.80	140.89	139.13	-1.25	142.68	143.25	0.40
$r-, K+, S+, \mu+$	149.26	151.06	1.21	148.92	146.46	-1.65	150.82	148.78	-1.35

Table A37: Insurance contract values obtained from simulation with $\kappa = 0.04,\ \kappa_w = 0$.

	Am. - Random Seed			Am. - Same Seed			Eu. - Same Seed		
Scenario	**excl.**	**ext.**	**%**	**excl.**	**ext.**	**%**	**excl.**	**ext.**	**%**
base	135.03	134.65	-0.28	134.55	135.55	0.74	106.44	105.23	-1.14
$r-$	140.85	140.42	-0.30	139.85	140.25	0.28	107.28	104.81	-2.30
$r+$	130.30	128.86	-1.10	130.02	130.83	0.62	105.84	105.66	-0.17
$K-$	135.60	135.14	-0.34	134.21	135.43	0.91	106.47	106.07	-0.38
$K+$	134.88	136.54	1.23	135.14	137.12	1.46	106.73	108.49	1.65
$S-$	108.61	107.14	-1.35	107.86	108.63	0.71	85.33	84.57	-0.89
$S+$	162.56	159.07	-2.15	161.31	159.51	-1.12	127.91	125.02	-2.26
$\mu-$	135.84	135.04	-0.59	134.59	135.40	0.61	106.56	106.32	-0.23
$\mu+$	135.90	136.20	0.22	134.54	134.74	0.15	106.61	106.11	-0.46
$S+, \mu-$	163.13	159.46	-2.25	161.64	159.37	-1.40	127.93	126.11	-1.43
$K-, \mu+$	134.73	136.69	1.46	134.30	134.62	0.24	106.58	106.95	0.34
$r-, S+$	169.17	164.97	-2.48	167.42	164.15	-1.95	128.99	123.38	-4.35
$r-, \mu-$	141.06	140.81	-0.18	139.78	140.10	0.23	107.39	105.89	-1.40
$r+, \mu-$	130.90	129.25	-1.26	130.05	130.68	0.49	105.83	106.74	0.86
$K-, S-$	135.08	136.69	1.19	133.82	134.62	0.60	106.31	106.95	0.60
$r-, K+$	140.51	142.74	1.59	140.25	141.50	0.89	107.60	107.59	-0.01
$K+, S-$	109.18	109.03	-0.14	108.05	110.20	1.99	85.29	87.83	2.98
$K+, S+, \mu-$	162.79	161.35	-0.89	161.83	160.94	-0.55	127.83	129.36	1.20
$r-, S+, \mu+$	169.66	166.53	-1.85	167.77	163.34	-2.64	128.80	124.26	-3.52
$r-, K-, S-, \mu-$	112.64	113.22	0.51	111.79	113.45	1.48	85.98	87.75	2.06
$r-, K-, S-, \mu+$	113.05	114.38	1.17	111.76	112.79	0.92	85.87	87.55	1.96
$r-, K-, S+, \mu-$	168.13	165.41	-1.62	167.60	164.21	-2.03	128.57	125.78	-2.17
$r-, K+, S-, \mu-$	112.34	115.48	2.80	112.39	114.50	1.88	86.13	89.23	3.59
$r+, K-, S-, \mu-$	104.25	102.79	-1.40	103.81	103.26	-0.52	84.78	85.23	0.54
$r+, K+, S+, \mu+$	156.97	156.15	-0.52	156.20	155.95	-0.16	126.83	131.28	3.50
$r+, K+, S+, \mu-$	157.36	154.99	-1.50	156.44	156.61	0.11	126.80	131.48	3.69
$r+, K+, S-, \mu+$	104.90	104.49	-0.39	104.38	104.93	0.53	84.76	88.40	4.29
$r+, K-, S+, \mu+$	156.88	155.62	-0.81	155.68	153.61	-1.33	126.86	127.91	0.83
$r-, K+, S+, \mu+$	169.11	168.84	-0.16	167.70	164.59	-1.85	129.02	127.05	-1.53

Table A38: Insurance contract values obtained from simulation with $\kappa = 0,\ \kappa_w = 0.06$.