Dissertation

Calibration Methods in Financial Market Applications

Philipp Mahler

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

Betreuer und 1. Gutachter:	Prof. Dr. Ralf Korn
2. Gutachter:	Prof. Dr. Marcus C. Christiansen
Datum der Disputation:	11. September 2026

DE-386

Rheinland-Pfälzische Technische Universität
Kaiserslautern-Landau
Fachbereich Mathematik
Gottlieb-Daimler-Straße 47
67663 Kaiserslautern

Fraunhofer ITWM
Abteilung Finanzmathematik
Fraunhofer-Platz 1
67663 Kaiserslautern

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit mit dem Titel

Calibration Methods in Financial Market Applications

selbstständig und ohne unerlaubte fremde Hilfe angefertigt, keine anderen als die angegebenen Quellen und Hilfsmittel verwendet und die den verwendeten Quellen und Hilfsmitteln wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Ich erkläre, dass diese Dissertation weder in der vorliegenden noch in einer im Wesentlichen gleichen Form als Prüfungsarbeit bei einem anderen Fachbereich oder an einer anderen Hochschule eingereicht wurde.

Ort und Datum	Philipp Mahler

Liste eigener wissenschaftlicher Veröffentlichungen und Beiträge

Die folgenden eigenen wissenschaftlichen Veröffentlichungen und Beiträge stehen in Zusammenhang mit dem in dieser Dissertation verwendeten Material.

Veröffentlichte und angenommene Arbeiten

- Mark-Oliver Wolf, Benedict Nikolaus Meyer, Philipp Mahler und Maximilian Diehl. *`openIRM`: Publicly accessible internal risk model of an artificial life insurer for analyzing and benchmarking actuarial methods in the Solvency II setting.* European Actuarial Journal, 16(1), 225–282, 2026. DOI: 10.1007/s13385-025-00435-6.

- Mark-Oliver Wolf, Benedict Nikolaus Meyer, Philipp Mahler und Maximilian Diehl. *Publisher Correction: `openIRM`: Publicly accessible internal risk model of an artificial life insurer for analyzing and benchmarking actuarial methods in the Solvency II setting.* European Actuarial Journal, 16(1), 283–285, 2026. DOI: 10.1007/s13385-025-00440-9.

- Philipp Mahler, Mark-Oliver Wolf, Marcus Willems, Jörg Wenzel, Andreas Wagner und Ralf Korn. *Konstruktion von Chancen-Risiko-Klassen zur Klassifizierung von Altersvorsorgeprodukten.* DAV Journal, 2025.

Preprints und eingereichte Manuskripte

- Philipp Mahler und Peter Ruckdeschel. *Advanced Calibration Analysis and Tools: Identifying Influential Observations in Stochastic Interest Rate Model Calibration.* arXiv preprint arXiv:2606.20420, 2026. DOI: 10.48550/arXiv.2606.20420. Eingereichtes Manuskript.

Der Dissertationstext kennzeichnet, wo Material aus diesen Veröffentlichungen oder gemeinsamen Projekten reorganisiert, erweitert oder als Anwendungsfall innerhalb der Monographie verwendet wird.

Zusammenfassung

Die verlässliche Kalibrierung von Zinsmodellen ist wesentlich für marktkonsistente Bewertung und ökonomische Szenariengeneratoren. In vielen Anwendungen wird Kalibrierung jedoch vor allem als Anpassung von Modellpreisen an beobachtete Marktpreise durch Wahl geeigneter Modellparameter verstanden. Diese Sicht ist für langfristige aktuarielle Anwendungen zu eng: Ein niedriger Kalibrierungsfehler allein sagt noch nicht, ob die geschätzten Parameter lokal identifizierbar, zeitlich stabil und für anschließende Simulationen geeignet sind.

Diese Dissertation untersucht Kalibrierung deshalb als mehrstufiges Qualitäts- und Validierungsproblem. Im Mittelpunkt steht das G2++-Modell, ein stochastisches Zwei-Faktor-Short-Rate-Modell, das analytische Handhabbarkeit mit hinreichender Flexibilität für die Modellierung von Zinsstrukturbewegungen verbindet. Die Analyse beginnt mit der Konstruktion einer anwendungsadäquaten Anfangszinsstrukturkurve. Anschließend werden Werkzeuge aus der nichtlinearen Regressionsdiagnostik auf die Kalibrierung unter dem risikoneutralen Maß $\mathbb{Q}$ übertragen, darunter die gewichtete Hut-Matrix, Leverage, Influence-Diagnostik, effektive Freiheitsgrade und lokale Unsicherheitsquantifizierung. Für at-the-money Caps wird die analytische Struktur der G2++-Cappreise genutzt, um die lokale Geometrie effizient auszuwerten. Ergänzend werden Cap- und Swaption-Kalibrierungen verglichen und das G2++-Modell einem Ein-Faktor-Hull–White-Benchmark gegenübergestellt, um die Rolle von Instrumentenklasse und Modellstruktur empirisch einzuordnen.

Darauf aufbauend behandelt die Arbeit den Übergang vom risikoneutralen Maß $\mathbb{Q}$ zum physischen Maß $\mathbb{P}$. Dazu werden verschiedene Spezifikationen von Risikoprämien konstruiert und hinsichtlich der resultierenden $\mathbb{P}$-Dynamik, langfristiger Zentren und Eignung für aktuarielle Szenarien bewertet. Eine Fallstudie und eine strukturierte Ex-post-Validierung zeigen, wie lokale Kalibrierungsdiagnostik, Real-World-Dynamik und nachgelagerte Simulationsergebnisse gemeinsam beurteilt werden können. Der empirische Fokus liegt auf Euro-Zinsmarktdaten von 2016 bis 2025, einschließlich Zinsstruktur- und Derivatemarktdaten, sowie auf der kontrollierten Übertragung diagnostischer Aussagen zwischen Instrumentenklassen.

Die zentrale Schlussfolgerung lautet: Kalibrierung ist für ihre intendierte Verwendung erst dann verlässlich, wenn die Marktdatenbasis kohärent ist und Marktfit, lokale Identifizierbarkeit, Real-World-Dynamik und Ex-post-Validierung gemeinsam plausibel sind.

Abstract

Reliable calibration of interest-rate models is essential for market-consistent valuation and Economic Scenario Generators (ESGs). In many applications, however, calibration is still understood primarily as fitting model prices to observed market prices by choosing suitable model parameters. This view is too narrow for long-horizon actuarial applications: a low calibration error alone does not show whether the estimated parameters are locally identifiable, stable over time, and suitable for subsequent simulations.

This work therefore analyses calibration as a multi-stage quality and validation problem. The main model is the G2++ model, a stochastic two-factor short-rate model that combines analytical tractability with sufficient flexibility to represent term-structure movements.

The analysis starts with the construction of an application-adequate initial term structure. Tools from nonlinear regression diagnostics are then transferred to calibration under the risk-neutral measure $\mathbb{Q}$, including the weighted hat matrix, leverage, influence diagnostics, effective degrees of freedom, and local uncertainty quantification.

For at-the-money (ATM) caps, the analytical structure of G2++ cap prices is used to evaluate the local geometry efficiently. In addition, cap and swaption calibrations are compared, and the G2++ model is contrasted with a one-factor Hull–White benchmark to assess empirically the role of instrument class and model structure.

Building on this, the analysis turns to the shift from the risk-neutral measure $\mathbb{Q}$ to the physical measure $\mathbb{P}$. Different risk-premium specifications are constructed and evaluated with respect to the resulting $\mathbb{P}$-dynamics, long-run centres, and suitability for actuarial scenarios. A case study and a structured ex-post validation show how local calibration diagnostics, real-world dynamics, and subsequent simulation outputs can be assessed jointly. The empirical focus is on Euro interest-rate market data from 2016 to 2025, including term-structure and derivative-market data, and on the controlled transfer of diagnostic statements across instrument classes.

The main conclusion is that calibration is reliable for its intended use only if the market-data basis is coherent and market fit, local identifiability, real-world dynamics, and ex-post validation are jointly plausible.

Danksagung

Hinter dieser Dissertation liegt ein langer und schwieriger Weg für mich. Ich bin allen Menschen sehr dankbar, die mich dabei begleitet und mir geholfen haben. Zuallererst gilt mein Dank meinem Betreuer, Prof. Dr. Ralf Korn, für seine Unterstützung und sein stetes Vertrauen, das er in meine Arbeit gesetzt hat. Besonders dankbar bin ich für unsere Treffen, die mich nicht nur immer inhaltlich weitergebracht haben, sondern mir auch Perspektive und Sicherheit gegeben haben, dass ich das schon packe.

Unterstützung in besonderer Weise habe ich auch von Prof. Dr. Peter Ruckdeschel erfahren, der mir in kürzester Zeit viele mir neue Konzepte der Statistik nähergebracht hat, wofür ich sehr dankbar bin. Ohne unsere Treffen hätten zahlreiche Ideen nicht ihren Weg in die Dissertation gefunden. Dem Fraunhofer ITWM danke ich für die Möglichkeit, während meiner Promotion als wissenschaftlicher Mitarbeiter tätig zu sein, was mir erlaubt hat, aktuelle Fragen aus der Industrie in meine Forschung einzugliedern. Ebenso bedanke ich mich bei meinen aktuellen und ehemaligen Kolleg:innen der Abteilung Finanzmathematik sowohl für die moralische Unterstützung als auch für die Offenheit fachlich über dieses doch nischige Thema zu diskutieren.

Nennen möchte ich hier insbesondere Mark-Oliver Wolf, mit dem ich nicht nur am hier zitierten *openIRM*-Paper zusammengearbeitet habe, der sich darüber hinaus Woche für Woche meine Sorgen anhören durfte. Wir haben im gleichen Zeitraum an unseren Dissertationen gearbeitet, viele ähnliche Herausforderungen geteilt und uns gegenseitig durch schwierige Phasen geholfen. Das hat diesen Prozess deutlich weniger einsam gemacht.

Dankbar bin ich natürlich auch meiner Familie. Ich danke meinen Eltern, die mein Interesse an der Mathematik früh gefördert und so schon den Grundstein für diese Dissertation gelegt haben, bevor ich überhaupt wusste, was dieses Wort bedeutet. Ich wünschte, meine Mutter hätte den Abschluss dieser Arbeit miterleben können. Auch meinen Brüdern danke ich für ihre Unterstützung und die Perspektive, die sie mir außerhalb der Dissertation gegeben haben. Mein herzlicher Dank gilt auch meinen Schwiegereltern, die immer bereit waren zuzuhören und echtes Interesse an meiner Arbeit gezeigt haben. Ebenso danke ich meinen Freund:innen für ihre Geduld, ihre Ermutigung und ihren Humor. Sie haben mir zugehört, wenn ich mich beschweren musste, mich abgelenkt, wenn ich Abstand brauchte, und mich mehr als einmal daran erinnert, dass diese Dissertation nicht das Einzige im Leben ist. Nennen möchte ich hier vor allem Imelda, Timo, Gianna und Marcus. Vielen Dank, dass ich euch meine Freunde nennen kann.

Ich hätte diese Arbeit jedoch nicht ohne meine Frau Katharina zu Ende gebracht. Sie ist der Mensch, auf den ich mich immer verlassen kann. Ihre Liebe, ihr Verständnis, ihre Ermutigung und ihre Geduld haben mir die Stabilität gegeben, auch die Phasen zu überwinden, in denen ich ernsthaft daran dachte, aufzugeben. Dafür bin ich dankbarer, als ich jemals ausdrücken könnte. Diese Arbeit ist für dich.

Contents

Selbstständigkeitserklärung **i**

Liste eigener wissenschaftlicher Veröffentlichungen und Beiträge **iii**

Zusammenfassung **v**

Abstract **vii**

Danksagung **ix**

1 Introduction **1**
1.1 Background and motivation 1
1.2 Literature overview and research gaps 1
1.3 Key findings 3
1.4 Outline 5

2 Financial Market Modelling in Practice **7**
2.1 The German Pension Landscape and Regulatory Environment 7
2.2 The Chain of Dependency in Stochastic Modelling 8
2.2.1 From Ornstein–Uhlenbeck Processes to the Short Rate 9
2.2.2 From the Short Rate to Asset Classes 9
2.2.3 From Basic Assets to Investment Funds and Portfolios 10
2.2.4 From Portfolios to Reference Portfolios 10
2.2.5 From Reference Portfolios to Chance-Risk Classes (CRK) 10
2.3 Implication for the Calibration Chain 11

3 Fundamentals in Interest Rate Theory **13**
3.1 Risk-neutral valuation and actuarial scenario generation 13
3.2 The Arbitrage-Free Pricing Framework 13
3.2.1 Zero-Coupon Bonds 15
3.3 Measure Theory and Numeraire Change 15
3.3.1 The T-Forward Measure $\mathbb{Q}^T$ 15
3.4 Interest Rate Definitions and the HJM Framework 15
3.4.1 Continuously Compounded Spot and Forward Rates 16
3.4.2 Connection to the Heath-Jarrow-Morton (HJM) Framework 16
3.4.3 Expectations Hypothesis and Term Premia 16
3.5 Modern Market Context: Multi-Curve and Standard Formulas 17
3.5.1 Multi-Curve Discounting and Tenor Forwarding 17
3.5.2 Single-Curve Core Setup and Its Limits 18
3.5.3 Discrete Market Rates: LIBOR and Swaps 18
3.6 Calibration Instruments and Market Quoting Conventions 19
3.6.1 Caps and Floors 19
3.6.2 Swaptions 20

3.6.3 ATM Strikes, Swap Annuities, and Forward Swap Rates 20
3.6.4 Black-Type Market Quotes 21
3.6.5 Black volatilities as quoting devices 21
3.7 Affine Term Structure Models (ATSM) 22
3.8 One-Factor Short-Rate Models 22
3.8.1 The Ho-Lee Model 22
3.8.2 The Vasicek Model 23
3.8.3 The Hull-White Model (Extended Vasicek) 23
3.8.4 The Cox-Ingersoll-Ross (CIR) Model 23
3.9 Alternative Framework: The LIBOR Market Model 23
3.10 Motivation for the Two-Factor G2++ Model 23
3.10.1 The correlation problem 24
3.10.2 Volatility structures 24
3.10.3 The G2++ specification 25
3.10.4 Analytical tractability and model scope 25

4 The Interest Rate Model: Risk-Neutral and Real-World Dynamics 27
4.1 Introduction to capital market modelling 27
4.2 The G2++ model under the risk-neutral measure $\mathbb{Q}$ 27
4.3 Model-Implied Pricing Objects under $\mathbb{Q}$ 29
4.3.1 Bridging the Gap: From Market Formulas to G2++ Pricing 29
4.4 Switching to the real-world measure $\mathbb{P}$ 32
4.5 Spot Rates under the Real-World Measure 33
4.5.1 Representation of Maturity-Specific Spot Rates 33
4.5.2 Interpretation of the Two Gaussian Factors 34

5 Calibration: What Does It Mean? 37
5.1 Historical Estimation vs. Implied Calibration 37
5.1.1 Historical Estimation 37
5.1.2 Implied Calibration 38
5.2 Validation and Criteria for a Robust Calibration 38
5.3 The Challenge of Nonlinearity 39

6 Calibration Step I: The Term Structure 43
6.1 Why the initial term structure is part of calibration 43
6.2 Market objects and conventions 44
6.2.1 A numerical convention check: yields, discount factors, and dates 44
6.3 Constructing the initial term structure 45
6.3.1 Nelson-Siegel-Svensson method 46
6.3.2 Interpolation and extrapolation methods 47
6.3.3 Numerical reconstruction of the curve behind cap quotes 47
6.3.4 From volatility quotes to monetary prices 51
6.4 On the choice of the initial term structure 53
6.4.1 Published term structures 53
6.4.2 Swap-derived and OIS-derived term structures 54
6.4.3 Single-curve simplification and its limits 54
6.5 Connection to calibration instruments 55
6.6 Checklist for an application-adequate initial curve 55

7 Calibration Step II: The Risk-Neutral Parameters 57
7.1 Calibrating the risk-neutral parameters 57
7.1.1 The optimisation problem 57

7.2 Diagnostic Tools for Advanced Calibration Analysis . . . 58
7.2.1 Key Concepts: Nonlinear Regression . . . 59
7.2.2 Key Concepts: Asymptotic Statistics . . . 60
7.2.3 Key Concepts: Functional Delta Method . . . 62
7.2.4 RMSRE as a Weighted Least-Squares Problem . . . 64
7.2.5 Tangent-Space Geometry and the Weighted Hat Matrix . . . 65
7.2.6 Analytical Derivatives for Caps . . . 66
7.2.7 Exact Hessian and Gauss–Newton Adequacy . . . 69
7.2.8 Main Diagnostic Tools . . . 74
7.2.9 Empirical Diagnostic Study Design . . . 76
7.2.10 Goodness of Fit over Time . . . 77
7.2.11 Leverage and Influence . . . 77
7.2.12 Effective Degrees of Freedom and Active Constraints . . . 80
7.2.13 Regime Sensitivity and Parameter Uncertainty . . . 82
7.2.14 Boundary-Dominated Leverage . . . 84
7.2.15 From Analytical Cap Diagnostics to Numerical Swaption Diagnostics . . . 87
7.2.16 Why the Daily Calibration Is Kept Stand-Alone . . . 90
7.2.17 From Local Diagnostics to Validation Governance . . . 91
7.3 Empirical Comparison of Calibration Designs . . . 92
7.3.1 A Common Comparison Protocol . . . 92
7.3.2 Sensitivity to the Absolute-Error Objective . . . 92
7.3.3 Curve, Subset, Solver, and Parameter-Box Axes . . . 93
7.3.4 Caps as the Analytical Diagnostic Benchmark . . . 94
7.3.5 Swaptions as an Instrument-Class Extension . . . 94
7.3.6 The One-Factor Benchmark . . . 95
7.3.7 Scope of the Empirical Comparison . . . 95
7.4 Caps versus Swaptions . . . 96
7.4.1 Why Compare Caps and Swaptions? . . . 96
7.4.2 Common Sample and Calibration Protocol . . . 96
7.4.3 Swaption Instrument Subsets . . . 97
7.4.4 Fit and Convergence . . . 97
7.4.5 Parameter Regimes . . . 98
7.4.6 Boundary Behaviour of the Correlation Parameter . . . 98
7.4.7 Interpretation and Empirical Scope . . . 99
7.4.8 Representative Cross-Instrument Transfer Check . . . 99
7.5 The One-Factor Benchmark: Hull–White versus G2++ . . . 102
7.5.1 The Hull–White Model as a One-Factor Benchmark . . . 103
7.5.2 Hull–White Cap and Floor Prices . . . 104
7.5.3 Analytical Hull–White Cap Jacobian . . . 105
7.5.4 Diagnostic Geometry of the One-Factor Benchmark . . . 105
7.5.5 Endpoint Leverage in the Hull–White Cap Benchmark . . . 106
7.5.6 Hull–White Swaptions and the Jamshidian Route . . . 110
7.5.7 Empirical Evidence from the Paired Cap Sample . . . 111
7.5.8 Why Boundary Correlation Does Not Imply Hull–White . . . 116
7.5.9 Conclusion of the One-Factor Benchmark . . . 119

8 Calibration Step III: The Risk Premium 121
8.1 Constructing the risk premium . . . 121
8.1.1 Constant mean-reversion levels . . . 122
8.1.2 Step-function premium families . . . 122
8.2 Forecast information and target construction . . . 124

8.3 Calibrating the risk premium . . . 125
8.3.1 Calibration with constant mean-reversion levels . . . 125
8.3.2 Calibration with two-level premium functions . . . 126
8.3.3 Calibration with n-level premium functions . . . 127
8.4 Comparing the risk-premium specifications . . . 127
8.4.1 Forecast families and target construction . . . 128
8.4.2 Constant premium as the baseline . . . 128
8.4.3 Two-level step functions as the first extension . . . 128
8.4.4 General n-level step functions and iterative calibration . . . 128
8.4.5 A controlled common illustration . . . 129
8.5 Ergodic stationary distributions . . . 130
8.5.1 The short rate . . . 131
8.5.2 Spot rates of maturity S . . . 131
8.5.3 Numerical verification and interpretation . . . 132

9 Case Study: openIRM Calibration 135
9.1 Regulatory context: PRIIPs and Solvency II . . . 135
9.2 Calibrating for European market regulation . . . 136
9.2.1 Data preparation . . . 136
9.2.2 Model specifications . . . 137
9.2.3 Calibration results . . . 138
9.3 From calibration to available-capital outputs . . . 139

10 Ex-Post Validation 143
10.1 Ex-Post Validation versus Calibration Diagnostics . . . 143
10.2 A Documented Actuarial Validation Perspective . . . 144
10.2.1 Validation starts at the simulation origin . . . 144
10.2.2 Expected values, quantiles, and tail-sensitive summaries . . . 145
10.2.3 Yield-curve shape and the share of inverse curves . . . 145
10.2.4 Reference portfolios, CRK, and example products . . . 146
10.2.5 Broader actuarial validation style . . . 146
10.3 A reconstructed ECB-hike-date validation case . . . 146
10.4 A structured validation catalogue . . . 151
10.5 From validation findings to the final workflow . . . 152

11 Conclusion and Calibration Workflow 153
11.1 Answer to the research question . . . 153
11.2 The calibration-and-validation workflow . . . 153
11.3 Main contributions . . . 155
11.4 Boundaries of the evidence . . . 155
11.5 Extensions and future research . . . 156

References 157

Appendix 163

A Explicit Entries of the Hessian Matrix 163
A.1 Auxiliary Functions . . . 163
A.2 First-Order Variance Derivatives . . . 163
A.3 Second Derivatives with Respect to Volatility and Correlation . . . 164
A.4 Second Derivatives Involving Mean-Reversion Speeds . . . 165

B Real-World Measure Derivations 166

B.1 Girsanov Drift Shift and Real-World Factor Dynamics 166
B.2 Properties of the Ornstein-Uhlenbeck Processes x_t and y_t 167
B.3 Properties of the Short Rate r_t 168
B.3.1 Constant risk premiums 169
B.3.2 Step Function: Two Levels 170
B.3.3 Step Function: n Levels 170
B.4 Expected Spot Rates and Calibration Equations 171
B.5 Representative risk-premium calibration slices 173

C Technical Status of Swaption Diagnostics in the G2++ Model 174
C.1 Swaption map and parameter dependence 174
C.2 Why an analytical Jacobian is not used 175
C.3 Numerical Jacobian used for the diagnostic comparison 175

D Boundary-Dominated Leverage in the G2++ Cap Calibration 177
D.1 Exact reconstruction interpretation of leverage 177
D.2 Why no unconditional U-shape claim is made 178
D.3 Cumulative G2++ benchmark rows 179
D.4 Verification setup and boundary-gap definitions 180

E Statistical Background of the Local Diagnostic Approximation 182
E.1 Conditional daily design 182
E.2 M- and Z-estimator reading of the WLS optimum 183
E.3 Influence, covariance, and transformed uncertainty 184

1 Introduction

1.1 Background and motivation

Traditional interest-rate calibration is often presented as numerical curve fitting: a model is chosen, market prices are collected, and parameters are fitted by minimising an error criterion. In actuarial financial modelling, the calibrated model also serves as the starting point for long-horizon simulations, reference portfolios, risk-return classifications, solvency calculations, and validation reports. A calibration error or an unstable parameter history can therefore propagate far beyond the pricing layer.

The practical motivation comes from two closely related settings. The first is the German pension-product environment, where the *Altersvorsorge-Produktinformationsblattverordnung* and the resulting Chance-Risk Class framework require comparable simulation-based statements about subsidised pension products [19, 47, 72, 77]. The second is the insurance-internal-model environment, where Solvency II, actuarial guidance on economic scenario generators, and public benchmark models such as *openIRM* connect market-consistent valuation, real-world scenario generation, and solvency-relevant capital measures [31, 90, 46, 40]. In both cases, the interest-rate model is a component of a larger decision system.

Calibration is treated here as a sequence of linked modelling decisions: initial term structure, risk-neutral calibration and diagnostics, real-world risk premia, and output validation. This sequence begins with the construction of an application-adequate initial term structure. This first choice already fixes discount factors, forward rates, and the market environment in which the stochastic parameters are later interpreted. It then continues with risk-neutral calibration under $\mathbb{Q}$, where liquid market instruments are used to fit the model parameters. The resulting fit is assessed through the residual norm and through local diagnostics such as leverage, influence, effective degrees of freedom, active constraints, and local uncertainty. The calibration chain subsequently requires a transition to the real-world measure $\mathbb{P}$, where risk premia determine the scenario dynamics relevant for simulation. Finally, the resulting simulations have to be validated against actuarially relevant output quantities. The guiding question is whether the calibrated model fits market data and remains interpretable, stable, and acceptable for the actuarial purpose for which it is used.

1.2 Literature overview and research gaps

The literature review is organised around five areas, each corresponding to one stage of the calibration chain developed here.

The first strand covers curve construction and market conventions. Nelson–Siegel and Svensson provide the classical parsimonious curve families [79, 87]. Practical curve construction and interpolation issues are discussed in the market-convention literature, including central-bank curve

construction and shape-preserving interpolation methods [32, 55]. Supervisory long-end extrapolation, Smith–Wilson methods, and UFR-type specifications are covered by Smith–Wilson and the EIOPA risk-free-rate material [86, 42, 43, 41]. OIS discounting and multi-curve modelling provide the post-crisis market-convention background for the discounting and tenor structure used in derivative pricing [59]. These references are important because the initial curve is not a neutral input. It fixes discount factors, forward rates, and calibration targets before the stochastic parameters are optimised.

The second strand is the no-arbitrage term-structure and short-rate modelling literature. The HJM framework gives the forward-rate no-arbitrage benchmark [58]. Affine term-structure models explain why exponential-affine bond prices and low-dimensional factor structures lead to tractable pricing formulas [39]. LIBOR and swap market models describe an alternative market-rate modelling tradition in which the directly modelled quantities are forward or swap rates rather than the instantaneous short rate [15, 67]. The model used in this dissertation is the G2++ model, a tractable two-factor Gaussian short-rate model from the standard interest-rate modelling literature [16]. The calibration perspective also follows the applied G2++ calibration work of Berninger and Pfeiffer [11]. Recent actuarial applications of G2++ to guaranteed pension-type products show that this model family remains relevant outside a purely trading-oriented setting [54].

The third strand concerns calibration under the risk-neutral measure. The calibration literature shows that the choice of loss function and the choice of calibration instruments can affect the fitted parameters and the interpretation of the calibration result [21, 75, 88]. This is the starting point for the diagnostic part of the dissertation. The fitted model is examined locally using residuals, leverage, influence, effective degrees of freedom, local uncertainty, and active constraints. The gap addressed here is that these diagnostic quantities are rarely used systematically in interest-rate calibration for actuarial applications.

The fourth strand concerns the transition from risk-neutral pricing to real-world simulation. The expectations hypothesis provides a simple link between forward rates and future spot rates, but the empirical literature shows that this link is too restrictive without term premia [48, 20, 50]. Dynamic yield-curve models and forecasting approaches provide richer ways to describe the evolution of yields over time [36, 37]. Term-premium estimates and decompositions show how strongly the interpretation of forward rates can depend on market, model, and sample period [71, 2, 22, 10].

For this dissertation, the consequence is that the $\mathbb{Q}$-to-$\mathbb{P}$ step has to be treated as a separate modelling problem. The risk-premium specification determines how the calibrated interest-rate model moves from the current market state to the forecast targets and to its long-run behaviour. The relevant questions are which targets are used, how much forecast information is available, and how flexible the real-world risk premium path should be.

The fifth strand is actuarial validation and governance. In the German pension-product and PRIIPs-related product-information setting, the relevant validation question is whether the capital-market model produces simulation outputs that remain plausible for reference portfolios, chance-risk measures, costs, risk indicators, and product classifications [19, 47, 77, 30]. In the insurance setting, Solvency II, EIOPA internal-model material, and the DAV ESG guidance stress the intended use of the model, the quality of the calibration, the plausibility of the generated scenarios, documentation, and validation of model outputs [46, 40, 31]. Recent economic-scenario-generator literature points to related questions such as parameter uncertainty, model combination, and validation of data-driven generators [7, 68]. These works are not part of the core G2++ methodology developed here, but they show that ESG validation remains an

active research area. For this dissertation, the important distinction is that local calibration diagnostics under $\mathbb{Q}$ and ex-post validation of simulated actuarial outputs are related, but they are not the same task.

The contribution lies in connecting these strands into one coherent calibration chain. This dissertation does not propose a new general term-structure model. Instead, it studies how a tractable interest-rate model can be calibrated, diagnosed, transferred to the real-world measure, used in actuarial applications, and validated in a way that makes the modelling assumptions explicit at each stage.

Two later chapters also have a specific prior-work relation. The diagnostic material in Chapter 7 builds on joint work with Prof. Dr. Peter Ruckdeschel on advanced calibration diagnostics. In this dissertation, this material is reorganised and extended for the G2++ calibration chain. The extensions include the G2++ cap and swaption comparison on the daily empirical sample over 2016–2025, the Hull–White benchmark, the RMSE/RMSRE sensitivity check, the Gauss–Newton adequacy analysis in Section 7.2.7, and the connection to the real-world risk-premium calibration and validation layers.

Chapter 9 reuses the published *openIRM* benchmark material [90] as an application case for the calibration chain developed in this dissertation. My contribution to the openIRM work concerned mainly the calibration of the capital-market model, in particular the interest-rate model and the corresponding $\mathbb{Q}$- and $\mathbb{P}$-calibration steps. The Solvency-II internal-risk-model construction, the available-capital calculation, the nested-simulation setting, and the broader *openIRM* benchmark design were primarily developed and led by Mark-Oliver Wolf, with contributions from the other co-authors. The separate list of own scientific publications and contributions records the publication and contribution status explicitly. This section clarifies how prior work, joint work, and application material enter the dissertation.

Table 1.1 summarises how the dissertation is positioned relative to the main literature strands. The table separates established results from the gaps that motivate this dissertation.

1.3 Key findings

Four findings organise the argument. They follow the same order as the calibration chain itself.

First, the initial term structure is part of calibration. Curve choice, interpolation, extrapolation, OIS discounting, and the single-curve simplification influence the deterministic environment in which the stochastic model is later interpreted. These choices therefore cannot be treated as merely technical preprocessing steps.

Second, a risk-neutral calibration has to be assessed beyond the final objective value. Pricing fit is the entry point, but it is not the whole calibration result. Residual structure, leverage, influence, effective degrees of freedom, identifiability, boundary activity, local uncertainty, and regime sensitivity all carry information about the credibility of the fitted parameter vector. The empirical comparison layer in Chapter 7 further shows how this evidence changes across caps, swaptions, and the Hull–White benchmark without turning the analysis into a universal model ranking.

Third, the transition to $\mathbb{P}$ is a separate modelling layer. Constant, two-level, and more general step-function risk premium specifications use different amounts of forecast information and imply different transition paths, long-run levels, and stationary distributions. A richer risk premium path is useful only if it is supported by the available forecasts and remains interpretable for the

Literature strand	**Established contribution**	**Remaining gap here**	**Contribution of this dissertation**
Curve construction and conventions	Yield-curve families, interpolation, OIS discounting, multi-curve conventions, and supervisory long-end extrapolation.	The initial curve is often treated as an input rather than as a calibration decision with consequences for the later model fit.	Treats the initial curve and the quote-to-price translation as the first step of the calibration chain.
Short-rate and market models	No-arbitrage term-structure theory, affine tractability, Gaussian short-rate models, and market-rate model alternatives.	The same tractable model is rarely carried through curve construction, $\mathbb{Q}$-calibration, $\mathbb{P}$-calibration, and validation for actuarial use.	Uses G2++ as the chosen workhorse model and states the limits of this choice against richer model classes.
Risk-neutral calibration	Loss-function choice, instrument selection, cap/swaption information content, and calibration risk.	A small pricing error is often the main reported result, while the local diagnostic structure of the fitted map receives less attention.	Uses residuals, leverage, influence, EDoF, local uncertainty, and active constraints to assess G2++ calibration beyond the objective value.
Real-world premia and forecasts	Expectation hypothesis tests, yield-curve forecasting, and term-premium decompositions.	The $\mathbb{Q}$-calibration does not determine how the model should move under $\mathbb{P}$ or how much forecast information the premium path can use.	Treats the $\mathbb{Q}$-to-$\mathbb{P}$ step as a separate calibration problem and compares constant, two-level, and step-function risk-premium specifications.
Validation and governance	Intended use, scenario plausibility, output testing, documentation, and model validation in actuarial and supervisory settings.	Local calibration diagnostics and validation of simulated actuarial outputs are often discussed as separate tasks.	Connects $\mathbb{Q}$-diagnostics with validation objects such as rate distributions, inverse curves, benchmark portfolios, product classifications, and solvency-relevant outputs.

Table 1.1: Position of the dissertation relative to the main literature strands.

intended application.

Fourth, calibration is incomplete without validation of the simulated outputs. Expected values, quantiles, tail-sensitive summaries, inverse-yield-curve shares, reference portfolios, Chance-Risk Classes, and solvency-relevant quantities are the objects through which the calibrated model is ultimately judged in actuarial applications. This final layer is important because actuarial users do not observe the calibration objective directly; they observe the consequences of the calibrated model in simulation and reporting.

1.4 Outline

Chapter 2 develops the actuarial and regulatory application setting. Chapter 3 introduces the interest-rate quantities and model families needed later. Chapter 4 defines the G2++ model under $\mathbb{Q}$ and $\mathbb{P}$. Chapter 5 explains the general calibration problem.

The three calibration chapters then form the technical core. Chapter 6 treats the initial term structure as the first modelling decision. Chapter 7 studies risk-neutral calibration, diagnostics, and the empirical comparison layer. Chapter 8 develops the real-world risk-premium specifications and their long-run implications.

Chapter 9 applies the calibration chain to the published *openIRM* benchmark. Chapter 10 formulates the validation layer for the simulated outputs. Chapter 11 combines the scientific conclusion with the resulting calibration-and-validation workflow.

2 Financial Market Modelling in Practice

The later chapters develop the mathematical derivation and calibration of the G2++ model. This chapter first sets out the industrial and regulatory context in which that calibration is used. In the domain of long-term pension provisions, particularly within the German life insurance market, the interest-rate model is not only a pricing device. It enters a dependency chain that affects the regulatory classification of pension products. The chapter introduces this chain and shows how calibration can affect the classification of retirement saving products.

2.1 The German Pension Landscape and Regulatory Environment

The environment for German life insurers has undergone substantial change since the deregulation of the insurance market in 1994. In the course of the European insurance single market, the former product control through supervisory pre-approval of tariffs and policy conditions was largely replaced by a more competition-oriented supervisory regime [83]. While various tax-oriented classification schemes exist, most notably the Three-Layer Model (*Drei-Schichten-Modell*), the institutional structure of German old-age provision is commonly described by the Three-Pillar Model (*Drei-Säulen-Modell*), consisting of statutory pension insurance, occupational pension provision, and private pension provision [34].

Driven by the *Altersvermögensgesetz* (AVmG) and the associated pension reforms, state-subsidised funded old-age provision gained importance from 2002 onwards as a supplement to the statutory pension system [17, 35]. A defining feature of certified *Riester* products is the statutory contribution guarantee. In contrast to the *Basisrente* (Rürup), which does not impose the same statutory contribution guarantee and may also be offered in fund-linked forms without guarantee, certified *Riester* contracts require the provider to make at least the paid-in old-age provision contributions available at the beginning of the payout phase [18, 38]. In product information, this guarantee is commonly expressed as covering the customer's own contributions and the state allowances received.

The transformation itself as well as its regulatory requirements have driven product diversification, moving the market beyond the traditional life insurance model. Providers now offer a broad spectrum of product types, ranging from traditional participating policies (*Klassik*) and index participations (*Indexpartizipation*) to static and dynamic hybrid products (*Hybridprodukte*), fund-linked contracts with guarantee mechanisms, and fund savings plans (*Fondssparpläne*). In particular, fund-based products may appear both without a contribution guarantee and in more involved guaranteed forms, where the guarantee is implemented through dynamic allocation mechanisms between risky and risk-reducing assets. This product heterogeneity has heightened the requirements for a transparent presentation of the inherent risks and opportunities. The

regulatory requirement for standardisation and comparability conflicts with the highly specific product characteristics in the German pension market. Risk and return profiles are influenced by a multitude of factors, including company-specific asset management, product-specific guarantee mechanisms, and customer-specific choices regarding capital allocation or investment strategies. Furthermore, the specific design of the contract, such as guarantee levels, costs and the planned retirement age, influences the performance of the product.

In response to this increased complexity, the legislator introduced the *Altersvorsorge-Produktinformationsblatt-Verordnung* (AltvPIBV) to regulate the transparent presentation of certified pension products [19]. For these certified products, the previous general information requirements (VVG-PIB) were replaced by the much more standardised *Produktinformationsblatt* (PIA-PIB). The centrepiece of this document is the Chance-Risk Class (CRK), which serves as a simplified indicator for the consumer. Linked directly to this classification are standardised model calculations for performance illustration and the disclosure of *Effektivkosten*, which measure the reduction in yield due to costs. To facilitate market-wide comparability, providers are required to publish so-called *Muster-Produktinformationsblätter* (Sample Product Information Sheets) for standardised contract constellations on their websites, thereby enabling an online comparison between the heterogeneous products of different manufacturers.

To ensure comparability and transparency for consumers in this complex setting, the legislator introduced the *Produktinformationsstelle Altersvorsorge gGmbH* (PIA). The PIA acts as a central, neutral authority responsible for the classification of subsidised pension products into Chance-Risk Classes (CRK). In the German subsidised pension-product setting, the PIA basic model and its calibration are described by Fraunhofer ITWM and the DAV-related literature as an industry standard for product calculations and CRK classification [47, 72, 73, 77].

The relevance of this modelling framework is not limited to the CRK setting. The DAV guidance on economic scenario generators treats model choice, calibration, data, and validation as central actuarial requirements for real-world and risk-neutral projections [31]. In the PRIIPs-related product-information context, the DAV also develops a standard procedure for category-4 PRIIPs. This procedure is closely linked to the capital-market model used for the Chance-Risk classification of certified retirement products and refers to the PIA specifications and calibrations as a basis for the corresponding standardised PRIIPs calculations [29, 30]. For the present thesis, this is the relevant point: the calibration and stability of the underlying capital-market model are not merely academic concerns, but prerequisites for comparable product information, PRIIPs-style risk-return figures, and model governance.

2.2 The Chain of Dependency in Stochastic Modelling

The classification of a pension product stands at the end of a modelling chain. Choices made at the term structure and interest-rate layer affect the simulated capital market, the derived portfolios, the reference portfolios, and finally the Chance-Risk classification. This section traces this chain from the calibrated stochastic processes to the final regulatory grouping.

Table 2.1 summarises the main links before they are discussed in more detail.

The table makes explicit why calibration quality matters already at the start of the chain. A problem in the short-rate layer is not confined to this layer. It can propagate into portfolio values, reference portfolios, classification boundaries, and finally into the consumer-facing product information.

Stage	Object	Dependence on calibration	Later consequence
Short-rate layer	G2++ short rate with two Ornstein–Uhlenbeck factors, deterministic shift, and real-world risk-premium specification.	Risk-neutral parameters, the initial term structure, and the Q-to-P risk-premium layer determine the primary interest-rate paths.	Instabilities at this stage affect all later interest-rate-sensitive projections.
Asset-class layer	Bond values, equity process, and basic capital-market indices.	Discount factors, risk-free drift, and interest-sensitive asset returns inherit the calibrated rate dynamics.	Bond and equity projections move with the same calibration choices.
Portfolio layer	Funds, the insurer's general cover pool (*Deckungsstock*), and mapped product portfolios.	These objects are built from the simulated bond and equity layers rather than from independent market models.	Product and benchmark return distributions inherit the earlier calibration quality.
Reference-portfolio layer	*Musterportfolios* and the benchmark Chance-Risk coordinate system.	The same calibrated capital-market model drives the conservative and risky reference portfolios.	Calibration instability can shift the benchmark system itself, not only one individual product.
Decision layer	Chance-Risk Classes and product information.	Final classification depends on the simulated product distribution relative to the reference portfolios.	A good local market fit can still lead to unstable or implausible product classifications if the simulated long-term dynamics are not suitable.

Table 2.1: Dependency chain from interest-rate calibration to actuarial product classification in the PIA setting.

2.2.1 From Ornstein–Uhlenbeck Processes to the Short Rate

The stochastic dynamics of the instantaneous short rate form the first layer of the modelling chain. Chapter 4 describes how the G2++ model constructs the short rate from two correlated Ornstein–Uhlenbeck factors and a deterministic shift function. For risk-neutral pricing, the relevant inputs are the calibrated mean-reversion speeds, volatilities, correlation, and the initial term structure. For real-world simulation, this layer also requires a risk-premium specification, which determines how the risk-neutral dynamics are transferred to the physical measure. Together, these choices determine the main statistical properties of the simulated interest-rate paths. If they are unstable or economically implausible, the resulting paths may show excessive negativity, unrealistically large rate swings, or unsuitable long-run behaviour. Since the short rate is the central interest-rate state variable, such artefacts affect the later asset-class and portfolio layers.

2.2.2 From the Short Rate to Asset Classes

The second layer translates the simulated short-rate paths into basic capital-market quantities. The most direct dependency occurs for fixed-income instruments. Zero-coupon bond prices and discount factors are derived from the short-rate dynamics and therefore inherit the calibrated rate behaviour. The interest-rate model also enters the equity process. In the standard framework used in the PIA setting, the equity model is based on a generalised Black–Scholes specification in which the drift depends on the risk-free rate and a risk-premium component. Thus a change in the interest-rate calibration affects not only bond simulations, but also the projected equity returns through the risk-free-rate channel. The G2++ model is therefore not an isolated pricing module, but one component of the wider capital-market scenario generator.

2.2.3 From Basic Assets to Investment Funds and Portfolios

Pension products usually do not invest directly in the short rate or in one raw market index. They use funds, mixed portfolios, guarantee mechanisms, or the insurer's general cover pool (*Deckungsstock*). To represent these objects in the model, they are mapped to the basic simulated risk factors, namely bond and equity components. For example, a mixed fund can be represented through a combination of simulated bond and equity components. Similarly, a cover-pool proxy depends on the simulated bond and equity market and on the rules used to smooth or distribute returns. As a consequence, the calibrated interest-rate model affects not only bond values, but also the return distributions of the mapped funds and product portfolios.

2.2.4 From Portfolios to Reference Portfolios

To classify heterogeneous products, the PIA framework uses standardised reference portfolios (*Musterportfolios*); see [72, 73, 77]. These portfolios represent different risk profiles, ranging from conservative to risky. For example, a conservative reference portfolio is mainly driven by fixed-income returns, while a risky reference portfolio is mainly driven by equity returns. The important point is that these reference portfolios are simulated with the same capital-market model as the products themselves.

This creates a relative coordinate system. A product is not assessed in isolation, but relative to the simulated reference portfolios. If the interest-rate calibration produces implausible dynamics, this may also affect the benchmark reference portfolio. In that case, the benchmark system itself moves. The effect is therefore not limited to one commercial product; it can also change the coordinate system used for the classification.

2.2.5 From Reference Portfolios to Chance-Risk Classes (CRK)

The final layer is the assignment of a Chance-Risk Class (CRK) to the pension product. The classification is based on simulated product outcomes for standardised sample customers (*Musterkunden*) and prescribed contract durations, typically 12, 20, 30, and 40 years. For each simulated scenario, the terminal value of the contract is calculated. From the empirical distribution of terminal values, two summary measures are derived. The *chance measure* is the annualised return derived from the mean scenario outcome over all scenarios. The *risk measure* is the corresponding annualised return derived from the mean of the worst 20% of scenario outcomes and is therefore an Expected-Shortfall-type quantity. The product is then assigned to a CRK class according to its position relative to the reference portfolios in the Chance-Risk plane; see [72, 73, 77] for the actuarial construction of this classification logic.

This classification logic is not only a formal mapping from simulated outcomes to discrete classes. It also carries an economic interpretation. In broad terms, one expects a natural product ordering from more risk-reducing and less chance-oriented forms such as *Klassik*, through *Indexpartizipation* and static or dynamic hybrids, to more chance-oriented and risk-bearing fund-based products. This ordering is not a strict mathematical rule for every individual product design, but it is an important plausibility check for the classification system. One role of the CRK framework is therefore to ensure that this economically intuitive ordering is not implausibly distorted by unstable benchmark movements or by a capital-market calibration that shifts the reference system itself.

This dependency structure explains why the interest-rate calibration is relevant for regulatory product information. A change in the G2++ volatility or mean-reversion parameters changes the bond price distribution. This can change the conservative reference portfolio and therefore also the boundaries used in the Chance-Risk classification. Hence a calibration that fits current

market prices well may still be unsuitable for the application if it produces unstable or implausible long-term simulation results. In this sense, calibration quality has to be judged not only at the pricing layer, but also through its effect on the downstream classification and product information.

2.3 Implication for the Calibration Chain

The dependency structure above explains why, in this application, calibration cannot be judged only by one pricing error. A close fit to current market prices is necessary, but it is only one part of the modelling problem. The calibrated interest-rate model also has to generate real-world scenarios that are economically plausible and suitable for the downstream product classification.

The subsequent chapters therefore separate the calibration task into several layers. Chapter 6 treats the construction of the initial term structure as the first modelling decision. Chapter 7 studies the risk-neutral calibration under $\mathbb{Q}$, where market instruments such as caps and swaptions are used to fit and diagnose the model. Chapter 8 then addresses the transition to the physical measure $\mathbb{P}$ through risk-premium specifications. Finally, Chapters 9 and 10 consider how the resulting model behaviour enters application and validation settings.

This view distinguishes the actuarial use of interest-rate models from a purely derivative-pricing use. In derivative pricing, the central question is whether the model reproduces market prices sufficiently well. In the pension-product setting, the same calibration also affects long-term simulations, reference portfolios, Chance-Risk classifications, and product information. It is therefore consistent with recent actuarial work in which G2++ is used for guaranteed pension-type products and not only for derivative-pricing exercises [54].

3 Fundamentals in Interest Rate Theory

This chapter introduces the interest-rate quantities, market conventions, and stochastic pricing frameworks used throughout this work. We define the notation for zero-coupon bonds, short rates, forward rates, and swap rates, and fix the arbitrage-free pricing relations used in the calibration chapters.

We analyse standard one-factor short-rate models, such as the Vasicek and Hull-White models, reviewing both their analytical tractability and their structural limitations [16, 27]. While one-factor models offer efficient closed-form pricing formulas, they imply perfect correlation across maturities and restrict the volatility structure. These limitations motivate the selection of the two-factor G2++ short-rate model.

The chapter first fixes the pricing notation and market-rate conventions. It then discusses calibration instruments and benchmark quote formulas before reviewing the model classes that motivate the G2++ specification used in the subsequent chapters.

3.1 Risk-neutral valuation and actuarial scenario generation

The interest-rate framework used in this dissertation combines two valuation perspectives with different roles. Actuarial valuation has traditionally been concerned with the present value of long-term contingent cash flows, such as life annuities, using biometric information and discounting assumptions; an early reference point is Halley's life-table work [56]. Modern interest-rate modelling, as used for derivative pricing, is instead formulated under an arbitrage-free pricing measure $\mathbb{Q}$, where market-consistent values are obtained from discounted risk-neutral expectations [16].

Both perspectives are needed in the calibration chain studied here. The risk-neutral measure $\mathbb{Q}$ is used for market-consistent valuation and for the calibration of the G2++ model to derivative prices. The physical measure $\mathbb{P}$ is used for real-world scenario generation, long-horizon simulations, and actuarial output validation. In application settings such as openIRM, these two layers appear together: market-consistent valuation requires a risk-neutral model, while solvency and risk assessments are evaluated through real-world scenarios [90]. The distinction between $\mathbb{Q}$ and $\mathbb{P}$ is therefore part of the modelling structure used throughout the dissertation.

3.2 The Arbitrage-Free Pricing Framework

We assume a continuous-time financial market defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{Q})$, where the filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\geq 0}$ satisfies the usual conditions and represents the market information available up to time t. We assume the existence of a risk-neutral measure $\mathbb{Q}$ (also known

as the Equivalent Martingale Measure, EMM), under which the discounted price processes of all tradeable assets are martingales.

The Bank Account and the Stochastic Discount Factor

Following the framework established in [16], we define the bank account (or money market account) $B(t)$ as the value of a risk-free investment starting with one unit of currency at time $t = 0$ and accruing interest at the instantaneous short rate r_t:

$$B(t) = \exp\left(\int_0^t r_u \, \mathrm{d}u\right). \tag{3.1}$$

We assume the standard arbitrage-free pricing framework, where $B(t)$ serves as the canonical numeraire associated with the risk-neutral measure $\mathbb{Q}$. Under this measure, the discounted price process of any tradeable asset $V(t)$ is a martingale:

$$\frac{V(t)}{B(t)} = \mathbb{E}^{\mathbb{Q}}\left[\frac{V(T)}{B(T)} \,\middle|\, \mathcal{F}_t\right], \quad \text{for } t \leq T. \tag{3.2}$$

This yields the general risk-neutral pricing formula:

$$V(t) = \mathbb{E}^{\mathbb{Q}}\left[D(t,T) \cdot V(T) \,\middle|\, \mathcal{F}_t\right], \tag{3.3}$$

where

$$D(t,T) = \frac{B(t)}{B(T)} = \exp\left(-\int_t^T r_u \, \mathrm{d}u\right) \tag{3.4}$$

is the stochastic discount factor.

Jensen's Inequality and Convexity Adjustment

Under stochastic interest rates, the instantaneous forward rate $f(t,T)$ does not in general coincide with the expected future short rate $\mathbb{E}^{\mathbb{Q}}[r_T \mid \mathcal{F}_t]$. This effect already appears under the risk-neutral measure and does not require a real-world risk premium. Bond prices are obtained as risk-neutral expectations of stochastic discount factors:

$$P(t,T) = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^T r_u \, \mathrm{d}u\right) \,\middle|\, \mathcal{F}_t\right], \tag{3.5}$$

whereas instantaneous forward rates are obtained from

$$f(t,T) = -\frac{\partial \ln P(t,T)}{\partial T}. \tag{3.6}$$

Jensen's inequality enters when $-\log P(t,T)$ is compared with the conditional expectation, under $\mathbb{Q}$, of the integrated short rate: in general, the negative logarithm of the conditional expectation of the stochastic discount factor is not equal to the conditional expectation of the integral. In the corresponding risk-neutral comparison, the difference between forward rates and expected future short rates is commonly referred to as a *convexity adjustment*.

In exchange-traded futures contracts, such convexity adjustments are often visible explicitly. For standard over-the-counter forward rates, the corresponding effect is embedded in the arbitrage-free drift condition of the term-structure model, for example in the HJM framework. Consequently, expected future short rates, forward rates, and model-implied discount factors must not be treated as interchangeable in stochastic interest-rate modelling.

3.2.1 Zero-Coupon Bonds

The fundamental instrument in interest rate modelling is the *zero-coupon bond (ZCB)*. A ZCB with maturity T guarantees the payment of one unit of currency at time T. Let $P(t,T)$ denote the price of such a bond at time $t \leq T$. By definition, $P(T,T) = 1$. The price of a zero-coupon bond is given by the risk-neutral expectation of the stochastic discount factor:

$$P(t,T) = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^T r_u \,\mathrm{d}u\right) \mid \mathcal{F}_t\right], \tag{3.7}$$

where r_t is the *instantaneous short rate*, the rate at which an investment grows over an infinitesimally short period $\mathrm{d}t$. This short rate r_t is the primary state variable we aim to model in Chapter 4.

3.3 Measure Theory and Numeraire Change

The change-of-numeraire technique provides the mathematical foundation for deriving analytical pricing formulas [49]. Let $N(t)$ and $U(t)$ be strictly positive traded numeraires with associated pricing measures $\mathbb{Q}^N$ and $\mathbb{Q}^U$. The Radon-Nikodym density process is given by

$$Z_t^{N|U} = \left.\frac{\mathrm{d}\mathbb{Q}^N}{\mathrm{d}\mathbb{Q}^U}\right|_{\mathcal{F}_t} = \frac{N(t)/U(t)}{N(0)/U(0)}. \tag{3.8}$$

For an integrable payoff X at time T, Bayes' formula gives

$$\mathbb{E}^{\mathbb{Q}^N}\left[\frac{X}{N(T)} \mid \mathcal{F}_t\right] = \mathbb{E}^{\mathbb{Q}^U}\left[\frac{X}{N(T)}\frac{Z_T^{N|U}}{Z_t^{N|U}} \mid \mathcal{F}_t\right]. \tag{3.9}$$

Equivalently, the same arbitrage-free value can be represented under either numeraire:

$$V(t) = N(t)\mathbb{E}^{\mathbb{Q}^N}\left[\frac{X}{N(T)} \mid \mathcal{F}_t\right] = U(t)\mathbb{E}^{\mathbb{Q}^U}\left[\frac{X}{U(T)} \mid \mathcal{F}_t\right]. \tag{3.10}$$

3.3.1 The T-Forward Measure $\mathbb{Q}^T$

Choosing the zero-coupon bond $P(t,T)$ as numeraire yields the *T-forward measure* $\mathbb{Q}^T$. For a payoff X paid at time T, the pricing formula becomes

$$V(t) = P(t,T)\mathbb{E}^{\mathbb{Q}^T}\left[X \mid \mathcal{F}_t\right]. \tag{3.11}$$

Under this measure, prices of traded assets expressed in units of the numeraire $P(t,T)$ are martingales. In particular, bond-price ratios of the form $P(t,S)/P(t,T)$ are martingales under $\mathbb{Q}^T$ for times $t \leq S \wedge T$. This property is used later for caplet pricing, where the T_i-forward measure is associated with the payoff date T_i.

The same measure-change logic also appears in swaption pricing. In one-factor short-rate models, Jamshidian-type decompositions combine arbitrage-free bond pricing with a one-dimensional monotonicity property of coupon-bond values [66]. In the two-factor G2++ setting used here, the corresponding pricing problem is treated through the semi-analytical swaption representation discussed in Chapter 7, where the exercise boundary is handled implicitly rather than through a one-dimensional monotonicity reduction.

3.4 Interest Rate Definitions and the HJM Framework

While the short rate r_t is the theoretical driver of the model, it is not directly observable in the market. In practice, we observe rates associated with specific time horizons.

3.4.1 Continuously Compounded Spot and Forward Rates

The *continuously compounded spot rate* $i(t,T)$ is the constant rate that equates the bond price to its discounted face value:

$$P(t,T) = e^{-i(t,T)\cdot(T-t)} \quad \Longleftrightarrow \quad i(t,T) = -\frac{\ln P(t,T)}{T-t}.$$

The collection of spot rates for all maturities $T > t$ constitutes the *term structure of interest rates* (or yield curve) at time t.

Closely related is the *instantaneous forward rate* $f(t,T)$, which represents the interest rate applicable to an infinitesimal period $[T, T+\mathrm{d}T]$, as seen from time t. As described in [27], it is related to the bond price via the partial derivative with respect to maturity:

$$f(t,T) = -\frac{\partial \ln P(t,T)}{\partial T}. \tag{3.12}$$

Conversely, the bond price can be expressed as the integral of the forward rates:

$$P(t,T) = \exp\left(-\int_t^T f(t,u)\,\mathrm{d}u\right).$$

3.4.2 Connection to the Heath-Jarrow-Morton (HJM) Framework

A short-rate model specifies the short-rate or factor dynamics first and derives bond prices and forward rates from risk-neutral discounting. The identity $r_t = f(t,t)$ links the short rate to the short end of the forward curve. The HJM framework instead specifies the dynamics of the full forward-rate curve directly under $\mathbb{Q}$ [58]. A short-rate model therefore induces a particular HJM-consistent forward-rate dynamics, whereas the general HJM framework starts from the forward curve itself:

$$\mathrm{d}f(t,T) = \alpha(t,T)\,\mathrm{d}t + \sigma(t,T)\,\mathrm{d}W_t. \tag{3.13}$$

For notational simplicity, we present the one-dimensional case. The multi-factor extension uses vector-valued volatilities and the corresponding covariance terms in the drift condition. Under the risk-neutral measure, the HJM drift $\alpha(t,T)$ is determined by the volatility structure $\sigma(t,T)$:

$$\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)\,\mathrm{d}u. \tag{3.14}$$

The Hull–White and G2++ models used later can be interpreted as finite-dimensional factor specifications that induce particular HJM volatility structures; see [16].

3.4.3 Expectations Hypothesis and Term Premia

A classical interpretation of the term structure is given by the *expectations hypothesis*, which states that long-term interest rates can be interpreted as averages of expected future short-term rates. In its pure form, this hypothesis implies that forward rates are unbiased predictors of future spot rates under the physical measure $\mathbb{P}$.

In this form, the hypothesis is too restrictive as a general empirical principle. Yield spreads and forward rates contain information about future rate movements, but the connection is mediated by time-varying premia, horizon effects, and sample dependence. The classic evidence of Campbell and Shiller, the contrasting evidence in Froot, and euro-rate evidence by Gerlach and Smets all support a more careful conclusion: forward rates should not be used as real-world forecasts without a term-premium adjustment, although they still contain an expectations

component [20, 48, 50]. Consequently, forward rates embed both expectations about future short rates and risk premia. This decomposition is central in the term-premium and yield-curve-forecasting literature [71, 2, 22, 10].

Formally, this distinction is captured by the difference between the risk-neutral and real-world measures. Under $\mathbb{Q}$, forward rates represent arbitrage-free pricing objects, while under $\mathbb{P}$ they incorporate excess returns required by investors. This distinction is fundamental to our calibration workflow, where risk-neutral parameters are fitted under $\mathbb{Q}$ to replicate market prices, while long-term simulations and capital requirements are evaluated under the real-world dynamics of $\mathbb{P}$.

In later chapters, we represent the term-premium component through time-dependent drift adjustments in the real-world dynamics. This separates the dynamics used for market-consistent pricing from those used for physical-measure projections.

Formally, the change from the risk-neutral measure $\mathbb{Q}$ to the physical measure $\mathbb{P}$ is governed by a *market price of risk*, denoted here by ϑ_t in a schematic one-dimensional notation. Let $W_t^{\mathbb{P}}$ be a Brownian motion under the real-world measure. If the dynamics of a short-rate factor are written as

$$\mathrm{d}r_t = \mu^{\mathbb{P}}(t, r_t)\,\mathrm{d}t + \sigma(t, r_t)\,\mathrm{d}W_t^{\mathbb{P}}.$$

then Girsanov's theorem implies, under the usual equivalence assumptions, the existence of a process ϑ_t such that $\mathrm{d}W_t^{\mathbb{Q}} = \mathrm{d}W_t^{\mathbb{P}} + \vartheta_t\,\mathrm{d}t$ is a Brownian motion under $\mathbb{Q}$. The drift under the risk-neutral measure becomes

$$\mu^{\mathbb{Q}}(t, r_t) = \mu^{\mathbb{P}}(t, r_t) - \vartheta_t \sigma(t, r_t).$$

The term $\vartheta_t \sigma(t, r_t)$ represents the risk premium component in the drift. The pure expectations hypothesis effectively suppresses this component. In the two-factor G2++ model used later, this measure change is implemented in a factor-specific form. Instead of formulating the market prices of risk directly, the real-world dynamics are parametrised through physical-measure mean-reversion levels λ_t^x and λ_t^y. In Chapter 8, the resulting drift adjustments are calibrated directly to connect market-consistent pricing with physical-measure projections. They are therefore part of the real-world calibration layer and not a simple post-hoc adjustment to the risk-neutral parameters.

3.5 Modern Market Context: Multi-Curve and Standard Formulas

Stochastic modelling of interest-rate derivatives must reflect the post-crisis market conventions governing discounting and forwarding.

3.5.1 Multi-Curve Discounting and Tenor Forwarding

Following the 2008 financial crisis and subsequent regulatory shifts, the classical single-curve assumption in which the same term structure governs both discounting and rate projection no longer holds for the general market [59]. In modern collateralised derivative markets, cash flows are discounted using the Overnight Index Swap (OIS) curve (e.g., €STR in Europe), which approximates the risk-free rate, whereas floating payments (e.g., EURIBOR) are projected using distinct tenor forwarding curves that reflect credit and liquidity risk premiums. This multi-curve framework is enforced by Credit Support Annexes (CSAs), which require daily collateral posting against the mark-to-market value of derivatives. Because this collateral earns the overnight rate, the OIS curve represents the standard collateral-consistent discounting curve, separate from the tenor curves used to forecast floating payments.

3.5.2 Single-Curve Core Setup and Its Limits

The multi-curve framework is the standard benchmark description of modern collateralised derivative markets. In a trading environment, discounting and forwarding must be separated; under this multi-curve framework, basis spreads represent distinct market objects rather than numerical anomalies.

For the calibration framework examined in this dissertation, the focus is narrower. The objective is to build a coherent calibration chain for long-horizon actuarial use: from an application-adequate initial term structure, through calibration under $\mathbb{Q}$, to calibration under $\mathbb{P}$, and finally to simulation, validation, and governance in a PIA- or Solvency II-related setting. In this narrower context, a single-curve core setup is adequate because it is used as a controlled calibration convention rather than as a full description of collateralised multi-curve valuation.

Single-curve and multi-curve modelling are not equally rich. While a single-curve representation omits tenor-basis dynamics and cannot replicate all structural spreads of collateralised pricing, it provides a stable and consistent risk-free reference convention across the calibration and simulation steps.

For this reason, the OIS-based curve is selected as the primary risk-free term structure and the analysis works in a single-curve representation built around that reference. Tenor-specific forwarding curves are therefore not modelled explicitly. This isolates the general level and shape dynamics of the risk-free term structure required for asset-liability management.

This simplification requires three clarifications. First, multi-curve spreads are omitted. Second, discount factors, ATM strikes, and target prices are reconstructed under the same single-curve convention. Third, calibration residuals and parameter instabilities are interpreted conditional on this convention, since part of the discrepancy may reflect the omitted basis structure.

The detailed justification of this choice, and its consequences for the initial curve construction and calibration targets, is taken up again in Chapter 6.

3.5.3 Discrete Market Rates: LIBOR and Swaps

The Simple Forward Rate (LIBOR)

The simple forward rate $L(t; T_1, T_2)$ for the period $[T_1, T_2]$ is defined as:

$$L(t; T_1, T_2) = \frac{1}{\tau(T_1, T_2)} \left(\frac{P(t, T_1)}{P(t, T_2)} - 1 \right), \tag{3.15}$$

where $\tau(T_1, T_2)$ is the year fraction between T_1 and T_2. This is the idealised no-arbitrage rate in a single-curve setting. This rate is the underlying quantity for *interest rate caps*. For the coupon, caplet, floorlet, and swap schedules below, we write $\delta_i := t_i - t_{i-1}$ for the accrual period on the year-fraction time scale used below. If a day-count convention is explicitly considered, δ_i denotes the corresponding year fraction.

The shift to backward-looking rates

The transition from interbank offered rates (IBORs) to risk-free rates (RFRs), such as the euro short-term rate (€STR) and the Secured Overnight Financing Rate (SOFR), changed the timing of many floating-rate coupons. Legacy term rates are fixed at the beginning of the accrual period. Backward-looking RFR coupons are instead based on overnight rates realised during the accrual period, usually through compounded-in-arrears conventions.

The empirical cap and swaption calibration in this dissertation follows the market-quote conventions available in the data and uses the corresponding European-style pricing maps. Products based on fully backward-looking compounded rates require adjusted payoff definitions and pricing formulas. The diagnostic methodology can be applied to such products once the corresponding pricing functions are available. Their derivation is not part of the empirical design considered here.

Swap Rates and the Annuity Measure

Consider a standard payer swap in which the fixed rate K is paid and the floating rate is received at payment dates $T_{\alpha+1}, \ldots, T_\beta$. The value of this swap at time $t < T_\alpha$ is:

$$V_{\text{swap}}(t) = \sum_{j=\alpha+1}^{\beta} P(t, T_j)\delta_j(L(t; T_{j-1}, T_j) - K), \tag{3.16}$$

where $\delta_j = \tau(T_{j-1}, T_j)$. The *forward swap rate* $S_{\alpha,\beta}(t)$ is the rate K that makes the present value of the swap zero. Following [16], we define the *annuity (PVBP)* as:

$$A_{\alpha,\beta}(t) = \sum_{j=\alpha+1}^{\beta} \delta_j P(t, T_j). \tag{3.17}$$

The swap rate is then the ratio of the floating leg value to the annuity:

$$S_{\alpha,\beta}(t) = \frac{P(t, T_\alpha) - P(t, T_\beta)}{A_{\alpha,\beta}(t)}. \tag{3.18}$$

The forward swap rate $S_{\alpha,\beta}(t)$ is a martingale under the corresponding annuity measure $\mathbb{Q}^{A_{\alpha,\beta}}$, which uses the annuity $A_{\alpha,\beta}(t)$ as numeraire.

3.6 Calibration Instruments and Market Quoting Conventions

The calibration of stochastic models relies on liquid derivative quotes rather than raw short-rate observations. This section establishes the market-side pricing conventions, while Chapter 6 details the curve reconstruction and quote-to-price translation used in our empirical study.

3.6.1 Caps and Floors

Caps and floors are among the standard liquid instruments used to infer information about future interest-rate uncertainty. A cap is a strip of European call options, called *caplets*, written on a floating reference rate such as EURIBOR. A floor is the analogous strip of European put options, called *floorlets*. For the tenor structure

$$0 = t_0 < t_1 < \cdots < t_n,$$

the payoff of the i-th caplet at the payment date t_i is

$$\Pi_i^{\text{caplet}}(t_i) = N\delta_i\big(L(t_{i-1}; t_{i-1}, t_i) - K\big)^+,$$

where N is the notional and K is the strike. The corresponding floorlet payoff is

$$\Pi_i^{\text{floorlet}}(t_i) = N\delta_i\big(K - L(t_{i-1}; t_{i-1}, t_i)\big)^+.$$

The pricing of these options depends on specific economic timing: the floating rate is fixed at the beginning of the accrual period, namely at time t_{i-1}, whereas payment occurs only at the

end of the period, at time t_i. A cap price is therefore the sum of discounted caplet values, and analogously for floors.

An additional practical distinction concerns the strike convention. A cap is a portfolio of caplets based on forward rates, but an at-the-money cap quote usually uses one common strike for the whole cap. In the market convention used later, this ATM strike is chosen as the fair fixed rate of a newly entered plain-vanilla swap with the same maturity structure. Thus, the caplet payoff is based on a forward rate, whereas the ATM strike of the full cap is determined by the corresponding swap-rate convention.

During positive-rate periods, market quotes are often interpreted through Black-type log-normal formulas. In negative-rate environments, the unshifted log-normal convention is no longer suitable for rates close to or below zero. Market quotes may then be represented by shifted log-normal or normal, Bachelier-type, conventions. This distinction is used later when quoted volatilities or convention-based prices are translated into monetary calibration targets.

3.6.2 Swaptions

A swaption is an option to enter an interest-rate swap at a future exercise date. One distinguishes European, Bermudan, and American exercise rights. The chosen empirical setup uses European swaptions, because both the available market data and the model pricing formulas considered later are of that type.

The underlying swap may itself be entered in two directions. A *payer swaption* grants the right to pay the fixed rate and receive the floating rate. A *receiver swaption* grants the right to receive the fixed rate and pay the floating rate. Economically, payer swaptions hedge against rising interest rates, whereas receiver swaptions hedge against falling interest rates.

If the underlying forward-starting swap begins at time T_m and ends at time T_n, then the relevant strike convention is tied to the forward swap rate introduced above. An ATM swaption satisfies

$$K = S_{m,n}(0).$$

For the market data used later, this convention is central: observed swaption quotes are indexed by option expiry and swap tenor, but their interpretation always presupposes a precise reconstruction of the corresponding forward swap rate and annuity.

For ATM instruments, payer and receiver prices coincide by swaption put–call parity. In the notation above,

$$V^{\text{payer}}(0) - V^{\text{receiver}}(0) = NA_{m,n}(0)\big(S_{m,n}(0) - K\big),$$

which vanishes when $K = S_{m,n}(0)$. This is why ATM swaption panels can be discussed without separately carrying a payer and receiver branch in the empirical comparison.

3.6.3 ATM Strikes, Swap Annuities, and Forward Swap Rates

The definition of the annuity $A_{\alpha,\beta}(t)$ and the forward swap rate $S_{\alpha,\beta}(t)$ is the bridge between instrument payoffs and market quotes. For caps, each caplet is written on a simply compounded forward rate, whereas the ATM strike of the whole cap is commonly expressed through the corresponding swap-rate convention. For swaptions, the annuity enters both as the numeraire of the swap measure and as the scale factor in the benchmark price.

This has a direct interpretative consequence for calibration work. Discount factors, forward rates, annuities, and ATM strikes are not independent inputs. They form one convention-dependent quoting system. Hence, an implied volatility quote cannot be translated into a

monetary benchmark price unless the underlying curve construction and strike convention have been fixed consistently.

3.6.4 Black-Type Market Quotes

Market quotes for caps and swaptions are usually observed as implied volatilities relative to a benchmark pricing formula. The benchmark formula is often Black-76 in positive-rate settings and a shifted log-normal or normal variant in negative-rate environments. In each case, the implied volatility is not yet the object used for model calibration. It first has to be translated into a monetary price.

The Black-76 framework acts as the interest-rate analogue of the Black-Scholes model. Given an appropriate numeraire and pricing measure, the relevant forward rate or swap rate is modelled as log-normal under that measure, which leads to a closed-form option-pricing formula.

For caplets, the corresponding measure is the t_i-forward measure $\mathbb{Q}^{t_i}$. If the forward rate $L_i(t) := L(t; t_{i-1}, t_i)$ is assumed to follow a log-normal diffusion under that measure, then the time-zero caplet price becomes

$$V_i^{\text{Black caplet}}(0) = N\delta_i P(0, t_i)\Big[L_i(0)\Phi(d_{1,i}) - K\Phi(d_{2,i})\Big], \tag{3.19}$$

with

$$d_{1,i} = \frac{\log(L_i(0)/K) + \frac{1}{2}(\sigma_i^L)^2 t_{i-1}}{\sigma_i^L \sqrt{t_{i-1}}}, \qquad d_{2,i} = d_{1,i} - \sigma_i^L \sqrt{t_{i-1}}.$$

For European swaptions, the corresponding measure is the annuity measure associated with the annuity $A_{m,n}(t)$. If the forward swap rate is assumed to be log-normal under that measure, then the payer swaption price is

$$V^{\text{Black payer swpt}}(0) = NA_{m,n}(0)\Big[S_{m,n}(0)\Phi(d_1) - K\Phi(d_2)\Big], \tag{3.20}$$

with

$$d_1 = \frac{\log(S_{m,n}(0)/K) + \frac{1}{2}(\sigma_{m,n}^{\text{swpt}})^2 T_m}{\sigma_{m,n}^{\text{swpt}} \sqrt{T_m}}, \qquad d_2 = d_1 - \sigma_{m,n}^{\text{swpt}} \sqrt{T_m}.$$

The receiver formula follows analogously and by put–call parity.

3.6.5 Black volatilities as quoting devices

The Black caplet and swaption quoting conventions are benchmark conventions, not structural term-structure models. A caplet formula assumes log-normal dynamics for one forward rate under the relevant forward measure. A swaption formula assumes log-normal dynamics for one forward swap rate under the annuity measure. These assumptions are useful instrument-wise quote conventions, but they do not define one globally coherent log-normal model for all forward and swap rates. In particular, a swap rate is a weighted sum of forward rates, and weighted sums of log-normal variables are not log-normal in general.

The calibration objective is formulated in monetary prices. The market input is therefore not used as a structural assumption on the G2++ dynamics. If the available quote is an implied Black-type volatility, it is first translated into a monetary benchmark price under the relevant curve, strike, annuity, and volatility convention. If the data already contain convention-based prices, these prices play the same role. Chapter 6 describes the reconstruction of these target prices.

The structural calibration model then tries to reproduce the resulting monetary targets. The logical chain used later is therefore

$$\text{market quote} \quad \longrightarrow \quad \text{monetary calibration target} \quad \longrightarrow \quad \text{G2++ model price.}$$

The benchmark quoting convention and the structural calibration model need not coincide distributionally. The relevant requirement is that the target prices are reconstructed consistently and that the structural model reproduces them with sufficient quality.

This separation is used in the later empirical chapters. Chapter 6 fixes the curve and quote-conversion convention used to obtain the monetary targets. Chapter 7 then studies the G2++ calibration to those prices and interprets the resulting residuals, leverage, influence, and parameter uncertainty.

3.7 Affine Term Structure Models (ATSM)

Before analysing specific short-rate models, it is useful to recall the class of *affine term structure models*. In an affine term-structure model, zero-coupon bond prices are exponential-affine functions of a finite-dimensional state vector X_t:

$$P(t,T) = \exp\left(A(t,T) - B(t,T)^\top X_t\right), \tag{3.21}$$

where $A(t,T)$ and $B(t,T)$ are deterministic functions, typically obtained from Riccati-type ordinary differential equations; see [39]. In a one-factor short-rate model, the state vector may be the short rate itself. In a multi-factor Gaussian model such as G2++, the state vector consists of the Gaussian factors driving the short rate.

The relevance of this structure is computational. Bond prices are determined by the current factor values and time, rather than by the full past evolution of the yield curve. In this factor sense, the models considered below have a finite-dimensional Markov representation: conditional future distributions and prices depend on the current state vector, not on the entire history. This property supports repeated valuation, calibration, and sensitivity calculations.

This finite-dimensional structure also distinguishes short-rate factor models from the general HJM framework, where the full forward-rate curve is the state variable. The HJM formulation is more general, but its direct implementation is typically infinite-dimensional. The Vasicek, Hull–White, and G2++ models used below keep the state representation low-dimensional while preserving analytical or semi-analytical bond and derivative pricing formulas. This is the main reason why they are suitable for the repeated calibration tasks considered in this dissertation.

3.8 One-Factor Short-Rate Models

While one-factor models make the underlying pricing mechanisms transparent, their simplicity introduces restrictions once the model is used for calibration across several maturities.

3.8.1 The Ho-Lee Model

Ho and Lee (1986) pioneered the class of "exogenous" no-arbitrage models. Unlike Vasicek, which generates an endogenous yield curve determined by model parameters, Ho-Lee takes the observed market yield curve as an input. The dynamics are given by:

$$\mathrm{d}r_t = \theta(t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t. \tag{3.22}$$

By making the drift term $\theta(t)$ time-dependent, the model can fit the initial term structure of interest rates exactly. However, it lacks mean reversion, implying that interest rates can diverge to positive or negative infinity over long horizons.

3.8.2 The Vasicek Model

Vasicek (1977) introduced mean reversion via the Ornstein-Uhlenbeck process:

$$\mathrm{d}r_t = k(\theta - r_t)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t. \tag{3.23}$$

The model is analytically tractable and Gaussian. Its constant long-term mean θ, however, makes it impossible to fit an arbitrary initial yield curve $P^M(0,T)$ without modification.

3.8.3 The Hull-White Model (Extended Vasicek)

Hull and White (1990) extended Vasicek with a time-dependent mean-reversion level:

$$\mathrm{d}r_t = [\Psi(t) - ar_t]\,\mathrm{d}t + \sigma\,\mathrm{d}W_t. \tag{3.24}$$

The deterministic function $\Psi(t)$ is calibrated to fit the initial term structure exactly at $t = 0$.

This model serves as our primary baseline, demonstrating the capabilities of a one-factor Gaussian short-rate model with an exact fit to the initial curve, while illustrating why exact curve fitting alone is insufficient.

3.8.4 The Cox-Ingersoll-Ross (CIR) Model

The CIR model prevents negative interest rates by scaling volatility:

$$\mathrm{d}r_t = k(\theta - r_t)\,\mathrm{d}t + \sigma\sqrt{r_t}\,\mathrm{d}W_t. \tag{3.25}$$

While non-negativity was historically desirable, the Eurozone experience with negative rates between 2014 and 2022 required extensions such as the *shifted CIR* or displaced-diffusion model ($r_t = \tilde{r}_t - \delta$, with $\tilde{r}_t$ following CIR). Gaussian models, including G2++, allow negative rates without such a shift and therefore remain tractable in negative-rate environments.

3.9 Alternative Framework: The LIBOR Market Model

A major alternative to short-rate models is the *LIBOR Market Model (LMM)*; see [16, 15, 67]. Unlike short-rate models, which model r_t and derive other rates, the LMM models the dynamics of discrete forward rates $L(t; T_j, T_{j+1})$ directly, typically assuming log-normal dynamics. The LMM offers a richer market-rate representation, but its dimensionality and numerical cost make it less suitable for the repeated daily calibration design used here.

3.10 Motivation for the Two-Factor G2++ Model

The preceding model classes show the trade-off that motivates the chosen short-rate model. One-factor Gaussian models are analytically tractable and can fit the initial term structure exactly after a deterministic shift. Their limitation is the restricted movement of the yield curve across maturities and, in particular, the restricted range of curve shapes they can generate, such as pronounced dipped or humped curves. Market-rate models such as the LIBOR Market Model describe the term structure more directly, but their dimensionality and numerical cost are less suitable for the repeated daily calibration design used here.

3.10.1 The correlation problem

In one-factor models such as Vasicek and Hull–White, all yield-curve movements are driven by a single Brownian motion. The yield curve can shift, but independent twists or changes in slope cannot be generated by separate stochastic drivers. This is restrictive for calibration and risk management, since short, medium, and long maturities are not perfectly correlated in their movements. A second Gaussian factor relaxes this restriction and allows the curve to steepen or flatten through the interaction of two stochastic drivers.

3.10.2 Volatility structures

One-factor Gaussian short-rate models also restrict the term structure of volatility. In the Hull–White case, the volatility pattern is controlled by one mean-reversion speed and one volatility parameter. Market data often show richer shapes, including medium-maturity humps where rates around, for example, 2–5 years are more volatile than short or long rates.

Figure 3.1 illustrates this distinction using representative spot-rate volatility structures. In the

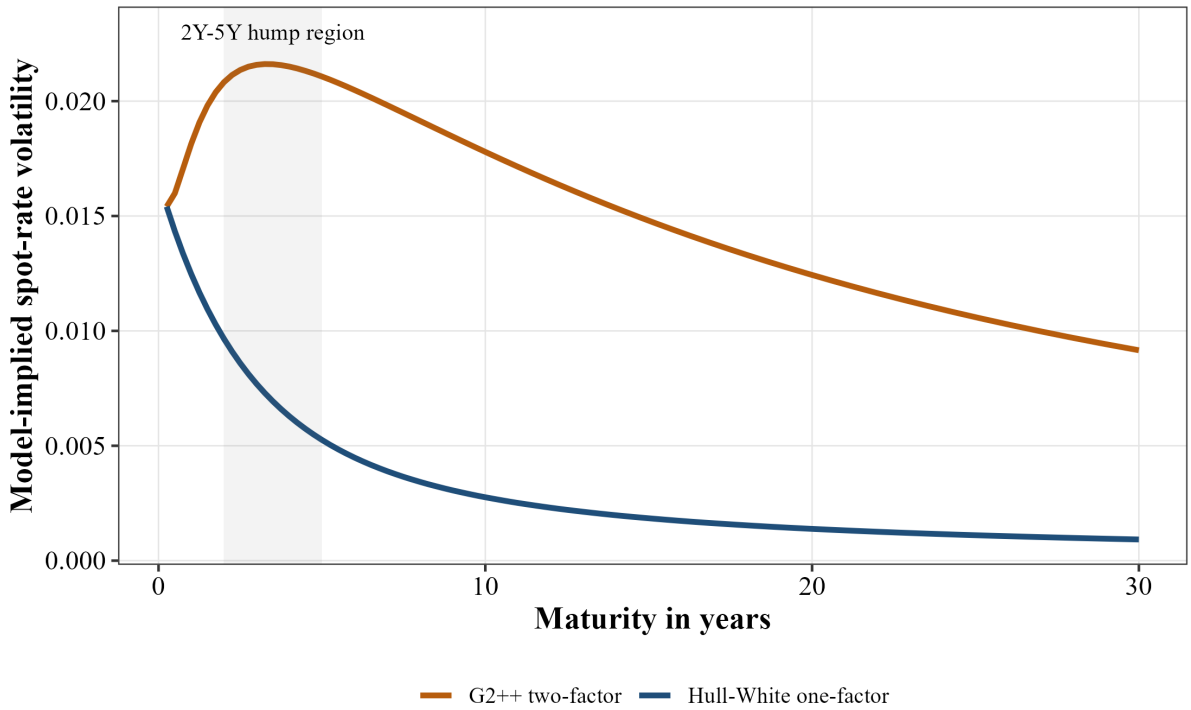


Figure 3.1: Illustrative model-based volatility-term-structure comparison. The curves plot a representative proxy for $\sqrt{\mathsf{Var}(i_t^S)}$ at the fixed horizon $t = 1$ under a one-factor Hull–White model and a two-factor G2++ model. The Hull–White curve is monotone, whereas the G2++ structure can generate a medium-maturity hump. The figure is illustrative and does not claim an empirical estimation result.

two-factor model, the volatility structure emerges from the interaction of the two factors rather than being controlled by a single scalar volatility parameter.

3.10.3 The G2++ specification

The G2++ model provides a two-factor Gaussian short-rate specification with higher flexibility than one-factor models while preserving affine tractability. The short rate is written as

$$r_t = x_t + y_t + \psi(t),$$

where x_t and y_t are two correlated Gaussian factors. The deterministic shift function $\psi(t)$ is chosen so that the model fits the initial market term structure exactly at the valuation date. The full risk-neutral and real-world dynamics of the model are specified in Chapter 4.

This model provides the level of flexibility and tractability required for the calibration framework used in the later chapters.

3.10.4 Analytical tractability and model scope

The chosen model has to be calibrated repeatedly on a long daily market-data panel. Analytical tractability is therefore part of the model choice. Monte Carlo methods introduce stochastic noise into the objective function, and PDE-based methods are computationally intensive. In the cap setting, the G2++ model provides analytical pricing formulas and analytical derivatives. This supports the local diagnostic analysis in Chapter 7, including leverage, influence, and Hessian-based checks.

More complex interest-rate models with stochastic volatility, jumps, regime switching, or explicit credit and liquidity components can represent additional market features. They also reduce analytical tractability and significantly increase the computational burden. For this reason, such models are treated as extensions outside the chosen empirical design.

4 The Interest Rate Model: Risk-Neutral and Real-World Dynamics

4.1 Introduction to capital market modelling

Chapter 2 explained why the interest-rate model is an early component of the actuarial modelling chain. It affects discount factors, bond and equity returns, portfolio values, reference portfolios, and ultimately product information. We now turn to the model used for this layer.

In contrast to the classical Black–Scholes framework with a deterministic short rate, we model the short rate stochastically. This allows the term structure to evolve over time and makes the interest-rate model relevant for asset-liability valuation. The short-rate model is therefore an important component of asset-liability management (ALM) in insurance and pension applications.

Short-rate models provide a tractable way to describe this evolution by specifying the stochastic dynamics of the instantaneous short rate r_t. Among the various short-rate models, the two-factor Gaussian model, commonly known as the G2++ model, has become one of the most widely used tractable frameworks in practice. Recent actuarial applications also use this modelling family for simulation and valuation problems close to the present setting [54, 90].

While the possibility of negative interest rates was long seen as a drawback of the G2++ model, today, after a period of observable negative interest rates in European markets between roughly 2014 and 2022, this property is one of the main reasons for its popularity. Additionally, its ability to fit the observable term structure exactly at the beginning of a simulation makes the G2++ model particularly attractive for actuarial applications, such as market-consistent valuation and long-term ALM projections.

The following sections introduce the G2++ model formally. We first state the model under the risk-neutral measure $\mathbb{Q}$, which serves as the basis for market-consistent calibration and valuation. We then describe the corresponding real-world dynamics under $\mathbb{P}$, which are needed for capital market simulations relevant to pension-product evaluation and actuarial risk assessment.

4.2 The G2++ model under the risk-neutral measure $\mathbb{Q}$

The short rate under the risk-neutral measure $\mathbb{Q}$ is defined as follows (see [16, 27, 11]):

$$r_t^* = x_t + y_t + \psi_t, \tag{4.1}$$

where the dynamics of x_t and y_t are given by

$$\begin{aligned} \mathrm{d}x_t &= -a_x x_t \mathrm{d}t + \sigma_x \mathrm{d}W_t^{x,\mathbb{Q}}, \quad x_0 = 0\ , \\ \mathrm{d}y_t &= -a_y y_t \mathrm{d}t + \sigma_y \mathrm{d}W_t^{y,\mathbb{Q}}, \quad y_0 = 0\ . \end{aligned}$$

Here, $W_t^{x,\mathbb{Q}}$ and $W_t^{y,\mathbb{Q}}$ are standard Brownian motions under $\mathbb{Q}$ with the instantaneous correlation

$$\mathrm{d}\langle W_t^{x,\mathbb{Q}}, W_t^{y,\mathbb{Q}}\rangle_t = \rho_{xy}\mathrm{d}t,$$

which implies that for any time step Δt, the increments are correlated with

$$\mathrm{Corr}(\Delta W^{x,\mathbb{Q}}, \Delta W^{y,\mathbb{Q}}) = \rho_{xy}.$$

ψ_t is a deterministic process, ensuring the criterion of perfectly fitting the observed bond prices to the model prices at time $t = 0$:

$$\psi_t = f^M(0,t) + \tfrac{{\sigma_x}^2}{2{a_x}^2}\left(1-e^{-a_x t}\right)^2 + \tfrac{{\sigma_y}^2}{2{a_y}^2}\left(1-e^{-a_y t}\right)^2 + \rho_{xy}\tfrac{\sigma_x\sigma_y}{a_x a_y}\left(1-e^{-a_x t}\right)\left(1-e^{-a_y t}\right), \quad (4.2)$$

where $f^M(0,t)$ is the instantaneous market forward rate for maturity t, which can be connected to the observed bond prices as follows:

$$P(0,T) = e^{-\int_0^T f^M(0,t)\mathrm{d}t}.$$

The explicit solutions for x_t and y_t are

$$x_t = \sigma_x \int_0^t e^{-a_x(t-u)}\,\mathrm{d}W_u^{x,\mathbb{Q}}, \quad y_t = \sigma_y \int_0^t e^{-a_y(t-u)}\,\mathrm{d}W_u^{y,\mathbb{Q}},$$

and the two processes are normally distributed with

$$x_t \sim \mathcal{N}\left(0, {\sigma_x}^2\frac{1-e^{-2a_x t}}{2a_x}\right), \quad y_t \sim \mathcal{N}\left(0, {\sigma_y}^2\frac{1-e^{-2a_y t}}{2a_y}\right). \quad (4.3)$$

Hence, under the risk-neutral measure $\mathbb{Q}$, the expected value of the short rate is given by

$$\mathbb{E}_{\mathbb{Q}}\left[r_t^*\right] = \psi_t,$$

whereas the variance is given by

$$\mathbb{V}\left[r_t^*\right] = \frac{{\sigma_x}^2}{2a_x}\left(1-e^{-2a_x t}\right) + \frac{{\sigma_y}^2}{2a_y}\left(1-e^{-2a_y t}\right) + 2\rho_{xy}\frac{\sigma_x\sigma_y}{a_x+a_y}\left(1-e^{-(a_x+a_y)t}\right).$$

An important tractability property of the model is that bond prices under the risk-neutral measure can be expressed explicitly (see [16])

$$P(t,T) = e^{-\int_t^T \psi_u \mathrm{d}u - \frac{1-e^{-a_x(T-t)}}{a_x}x_t - \frac{1-e^{-a_y(T-t)}}{a_y}y_t + \frac{1}{2}V(t,T)} \quad (4.4)$$

with $V(t,T)$ as follows:

$$\begin{aligned} V(t,T) &= \frac{{\sigma_x}^2}{{a_x}^2}\left[T-t+\frac{2}{a_x}e^{-a_x(T-t)} - \frac{1}{2a_x}e^{-2a_x(T-t)} - \frac{3}{2a_x}\right] \\ &+ \frac{{\sigma_y}^2}{{a_y}^2}\left[T-t+\frac{2}{a_y}e^{-a_y(T-t)} - \frac{1}{2a_y}e^{-2a_y(T-t)} - \frac{3}{2a_y}\right] \\ &+ 2\rho_{xy}\frac{\sigma_x\sigma_y}{a_x a_y}\left[T-t+\frac{e^{-a_x(T-t)}-1}{a_x} + \frac{e^{-a_y(T-t)}-1}{a_y} - \frac{e^{-(a_x+a_y)(T-t)}-1}{a_x+a_y}\right]. \end{aligned}$$

4.3 Model-Implied Pricing Objects under $\mathbb{Q}$

Equation (4.4) is the central analytical object of the risk-neutral G2++ model. Once zero-coupon bond prices are available in explicit form, the model can generate the derivative-pricing functions used for the calibration under $\mathbb{Q}$ and the spot rates used later under $\mathbb{P}$. Table 4.1 summarises the main objects that reappear throughout the calibration chain.

Object	Role in the G2++ model	Used later for
Zero-coupon bond price $P(t,T)$	Exponential-affine model-implied discount factor generated by the two Gaussian factors and the deterministic shift.	Initial curve fit, derivative pricing, and spot-rate construction.
Finite-maturity spot rate i_t^S	Yield implied by the model bond price over the maturity interval $[t, t+S]$.	$\mathbb{P}$-measure calibration and ex-post validation.
Caplet and cap price	Option on a zero-coupon bond and sum of caplets with closed-form integrated variance.	Risk-neutral cap calibration and analytical diagnostics.
Swaption price	Option on a coupon-bearing bond evaluated by a semi-analytical one-dimensional integral.	Cap-versus-swaption robustness comparison.
Jacobian of the pricing map	Sensitivity of the model prices with respect to the G2++ parameter vector.	Leverage, influence, effective degrees of freedom, and uncertainty diagnostics.

Table 4.1: Model-implied G2++ pricing objects and their role in the calibration chain.

The pricing objects themselves will follow in this chapter. The zero-coupon bond formula (4.4) is the common analytical basis. It gives the model-implied discount factor for every maturity pair (t,T), generates the derivative-pricing functions used under $\mathbb{Q}$, and determines the spot rates for fixed maturities used later under $\mathbb{P}$. The local differentiability arguments and Jacobian-governance details needed for the later diagnostics are deferred to Appendix C.

4.3.1 Bridging the Gap: From Market Formulas to G2++ Pricing

Chapter 3 introduced Black-type benchmark formulas as market quoting devices. The present chapter now provides the structural G2++ pricing functions to which those benchmark prices are compared. The two roles must remain distinct: Black-type formulas reconstruct benchmark prices from market quotes, whereas the G2++ formulas below generate the structural model prices fitted later. The key observation is that the relevant log-normal structure in G2++ does not arise at the level of the market rate itself, but at the level of bond prices. Since G2++ is an affine Gaussian short-rate model, zero-coupon bond prices are exponential-affine functions of the Gaussian state variables. Under the relevant forward measures, a single bond price is therefore log-normal. This is the mechanism behind the exact cap formula. The swaption case is more involved because the underlying object is a coupon-bearing bond, that is, a weighted sum of several zero-coupon bonds.

Caps as Options on Zero-Coupon Bonds

Consider the i-th caplet for the accrual period $[t_{i-1}, t_i]$ with strike K, accrual factor $\delta_i = t_i - t_{i-1}$, and notional N. Using the identity

$$L(t_{i-1}; t_{i-1}, t_i) = \frac{1}{\delta_i}\left(\frac{1}{P(t_{i-1}, t_i)} - 1\right),$$

the caplet value at the fixing date can be rewritten as

$$V_i^{\text{caplet}}(t_{i-1}) = N\big(1 - (1+\delta_i K)P(t_{i-1}, t_i)\big)^+.$$

Hence a caplet is exactly a European put option on the single zero-coupon bond $P(t_{i-1}, t_i)$ with modified strike $K_i^* = \frac{1}{(1+\delta_i K)}$.

Cap Pricing Formula

Under the t_{i-1}-forward measure, the bond price $P(t_{i-1}, t_i)$ is log-normal in G2++. Writing

$$F_i := \frac{P(0, t_i)}{P(0, t_{i-1})}$$

for the corresponding bond ratio and

$$\begin{aligned}\Sigma_i^2 = {} & \sigma_x^2 C(a_x, t_{i-1}, t_i)^2 C(2a_x, 0, t_{i-1}) \\ & + \sigma_y^2 C(a_y, t_{i-1}, t_i)^2 C(2a_y, 0, t_{i-1}) \\ & + 2\rho_{xy}\sigma_x\sigma_y C(a_x, t_{i-1}, t_i) C(a_y, t_{i-1}, t_i) C(a_x + a_y, 0, t_{i-1}),\end{aligned}$$

with

$$C(a, t_1, t_2) := \frac{1 - e^{-a(t_2 - t_1)}}{a},$$

the exact G2++ caplet price is

$$\text{Caplet}_i(0) = N\Big[P(0, t_{i-1})\Phi(h_{i,+}) - (1+\delta_i K)P(0, t_i)\Phi(h_{i,-})\Big], \tag{4.5}$$

where

$$h_{i,\pm} = \frac{1}{\Sigma_i}\left[\log\left(\frac{P(0, t_{i-1})}{(1+\delta_i K)P(0, t_i)}\right) \pm \frac{1}{2}\Sigma_i^2\right]. \tag{4.6}$$

Summing over all caplets of the cap yields

$$\text{Cap}(0, \mathcal{T}, N, K, \Pi) = \sum_{i=m+1}^{n} N\Big[P(0, t_{i-1})\Phi(h_{i,+}) - (1+\delta_i K)P(0, t_i)\Phi(h_{i,-})\Big]. \tag{4.7}$$

This formula is exact. Its structural resemblance to the Black caplet formula is therefore not accidental. In the market benchmark formula, the log-normal object is the forward rate itself. In G2++, the log-normal object is the corresponding bond ratio under the forward measure.

Swaptions as Options on Coupon-Bearing Bonds

For swaptions, the same general principle applies, but the underlying object is more complicated. At the exercise date T_m, the value of a payer swap can be written, with $\delta_j = T_j - T_{j-1}$, as

$$V^{\text{payer swap}}(T_m) = 1 - P(T_m, T_n) - K\sum_{j=m+1}^{n} \delta_j P(T_m, T_j).$$

Hence the payer swaption payoff is an option on a coupon-bearing bond, that is, on a sum of zero-coupon bond prices. Each individual bond price is log-normal in the Gaussian model, but a sum of log-normal random variables is not log-normal in general. This is the essential reason why the swaption formula is no longer fully explicit in the same way as the cap formula.

Swaption Pricing Formula

The two-factor Gaussian structure nevertheless remains tractable enough to yield the standard Brigo–Mercurio semi-analytical representation. Consider a swaption with expiry T, notional N, strike K, option type $\chi \in \{1, -1\}$, and swap payment dates $\mathcal{T} = \{t_1, \dots, t_n\}$, where $T < t_1 < \cdots < t_n$. We use the convention $\chi = 1$ for a payer swaption and $\chi = -1$ for a receiver swaption. Let $\delta_i = t_i - t_{i-1}$ denote the accrual periods and define the cash-flow coefficients

$$c_i = \begin{cases} K\delta_i & \text{for } i \in \{1, \dots, n-1\}, \\ 1 + K\delta_n & \text{for } i = n. \end{cases}$$

For each payment date t_i, write

$$B_x^i = C(a_x, T, t_i), \qquad B_y^i = C(a_y, T, t_i),$$

and define the deterministic bond-price coefficient

$$A(T, t_i) = \exp\left(-\int_T^{t_i} \psi_u \, \mathrm{d}u + \frac{1}{2} V(T, t_i) \right).$$

Then the G2++ zero-coupon bond price at the exercise date has the form

$$P(T, t_i) = A(T, t_i) \exp\left(-B_x^i x_T - B_y^i y_T \right).$$

Under the T-forward measure, the vector (x_T, y_T) is bivariate normal. Its variances and correlation are

$$\Sigma_x^2 = \frac{{\sigma_x}^2}{2a_x}(1 - e^{-2a_x T}), \qquad \Sigma_y^2 = \frac{{\sigma_y}^2}{2a_y}(1 - e^{-2a_y T}), \qquad \rho_T = \rho_{xy} \frac{\sigma_x \sigma_y}{\Sigma_x \Sigma_y} \frac{1 - e^{-(a_x + a_y)T}}{a_x + a_y}.$$

The corresponding means are

$$\mu_x = -\frac{{\sigma_x}^2}{2{a_x}^2} \left(1 - e^{-a_x T}\right)^2 - \rho_{xy} \frac{\sigma_x \sigma_y}{a_y} \left[\frac{1 - e^{-a_x T}}{a_x} - \frac{1 - e^{-(a_x + a_y)T}}{a_x + a_y} \right],$$

$$\mu_y = -\frac{{\sigma_y}^2}{2{a_y}^2} \left(1 - e^{-a_y T}\right)^2 - \rho_{xy} \frac{\sigma_x \sigma_y}{a_x} \left[\frac{1 - e^{-a_y T}}{a_y} - \frac{1 - e^{-(a_x + a_y)T}}{a_x + a_y} \right].$$

For a fixed value x of the first factor, the conditional distribution of y_T given $x_T = x$ is normal with mean

$$m_y(x) = \mu_y + \rho_T \frac{\Sigma_y}{\Sigma_x}(x - \mu_x)$$

and standard deviation

$$s_y = \Sigma_y \sqrt{1 - \rho_T^2}.$$

The exercise boundary $\bar{y}(x)$ is defined as the unique solution of

$$\sum_{i=1}^{n} c_i A(T, t_i) \exp\left(-B_x^i x - B_y^i \bar{y}(x) \right) = 1.$$

It separates the exercise and non-exercise regions conditional on $x_T = x$.

The functions entering the one-dimensional pricing integral are

$$h_1(x) = \frac{\bar{y}(x) - m_y(x)}{s_y},$$

$$h_2(x, t_i) = \frac{\bar{y}(x) - m_y(x) + B_y^i s_y^2}{s_y} = h_1(x) + B_y^i s_y,$$
$$\xi_i(x) = c_i A(T, t_i) \exp(-B_x^i x),$$

and

$$\kappa_i(x) = -B_y^i m_y(x) + \frac{1}{2}(B_y^i)^2 s_y^2.$$

With these definitions, the time-zero swaption price is

$$\begin{aligned}\text{Swaption}(0, T, \mathcal{T}, N, K, \chi, \Pi) = \chi N P(0, T) \int_{-\infty}^{\infty} & \frac{\exp\left[-\frac{1}{2}\left(\frac{x-\mu_x}{\Sigma_x}\right)^2\right]}{\Sigma_x\sqrt{2\pi}} \\ & \times \left[\Phi(-\chi h_1(x)) - \sum_{i=1}^{n} \xi_i(x) e^{\kappa_i(x)} \Phi(-\chi h_2(x, t_i))\right] \mathrm{d}x.\end{aligned} \tag{4.8}$$

This is the Brigo–Mercurio semi-analytical formula for the G2++ swaption price [16].

Why Swaption Prices Are Semi-Analytical in G2++

The cap formula is fully explicit because a caplet can be written as an option on one zero-coupon bond. The swaption case is different. The underlying object is a coupon-bearing bond, hence a weighted sum of zero-coupon bond prices. In G2++, each individual zero-coupon bond price is exponential-affine in the Gaussian factors, but the sum of such terms does not lead to a single log-normal object.

The remaining analytical difficulty is therefore the exercise boundary $\bar{y}(x)$. For each fixed value of x, this boundary has to be solved from the one-dimensional equation stated above. The pricing formula keeps the Gaussian conditioning structure, but its evaluation still requires the numerical solution of $\bar{y}(x)$ and a one-dimensional numerical integration. This is why the formula is semi-analytical rather than fully explicit. This level of structure is sufficient for the swaption robustness comparison in Chapter 7. The local differentiability and Jacobian-governance details used in the diagnostics are collected separately in Appendix C.

4.4 Switching to the real-world measure $\mathbb{P}$

When simulating retirement provision products and the interest-rate products embedded in them, the risk-neutral short rate is not sufficient. The risk-neutral measure is designed for market-consistent valuation, not for describing the expected real-world evolution of asset returns. Investors require compensation for bearing uncertainty about future asset prices, and this compensation is represented by a risk premium. In the present G2++ specification, this produces a real-world short-rate representation of the form

$$r_t := r_t^* + \lambda_t,$$

under the real-world measure $\mathbb{P}$; see [16]. More precisely, the two Ornstein-Uhlenbeck factors x_t and y_t each receive an additional drift component, denoted by λ_t^x and λ_t^y. The adjusted dynamics are

$$\begin{aligned}\mathrm{d}x_t &= a_x\left(\lambda_t^x - x_t\right)\mathrm{d}t + \sigma_x \mathrm{d}W_t^{x,\mathbb{P}}, \quad x_0 = 0\ , \\ \mathrm{d}y_t &= a_y\left(\lambda_t^y - y_t\right)\mathrm{d}t + \sigma_y \mathrm{d}W_t^{y,\mathbb{P}}, \quad y_0 = 0\ .\end{aligned}$$

The Brownian motions $\mathrm{d}W_t^{x,\mathbb{P}}$ and $\mathrm{d}W_t^{y,\mathbb{P}}$ are obtained from the risk-neutral Brownian motions by a Girsanov drift shift; see [11] and Appendix B.1 for the corresponding calculation in the

present notation. This is not the only possible real-world specification. Its advantage for the present calibration framework is that it keeps the model inside the Gaussian G2++ framework while making the two risk-premium components explicit.

While the expected value of both processes x_t and y_t under the risk-neutral measure is zero, under the real-world measure, x_t and y_t are normally distributed with

$$x_t \sim \mathcal{N}\left(\int_0^t e^{-a_x(t-u)} a_x \lambda_u^x \mathrm{d}u, {\sigma_x}^2 \frac{1-e^{-2a_x t}}{2a_x}\right),$$
$$y_t \sim \mathcal{N}\left(\int_0^t e^{-a_y(t-u)} a_y \lambda_u^y \mathrm{d}u, {\sigma_y}^2 \frac{1-e^{-2a_y t}}{2a_y}\right),$$

with

$$\mathbb{E}_{\mathbb{P}}[r_t] = \psi_t + \int_0^t e^{-a_x(t-u)} a_x \lambda_u^x \mathrm{d}u + \int_0^t e^{-a_y(t-u)} a_y \lambda_u^y \mathrm{d}u.$$

The variance stays the same, since the risk-premium specification changes the deterministic drift levels of the factors but not their diffusion coefficients or instantaneous correlation. This invariance is shown explicitly in Appendix B. Hence,

$$\mathbb{V}[r_t] = \mathbb{V}[r_t^*].$$

4.5 Spot Rates under the Real-World Measure

The later calibration and validation steps require not only the instantaneous short rate itself, but also spot rates for fixed maturities. These rates enter observed yield curves, forecast targets, actuarial valuation formulas, and several validation quantities. We therefore record the corresponding representation under $\mathbb{P}$ before turning to the calibration chapters.

4.5.1 Representation of Maturity-Specific Spot Rates

The continuously compounded spot rate observed at time t for maturity $t+S$, with $S>0$, is given by

$$i(t, t+S) = i_t^S = -\frac{\log(P(t, t+S))}{S}.$$

Using the G2++ bond price formula (4.4), it follows that

$$i_t^S = \frac{1}{S}\int_t^{t+S} \psi_u \,\mathrm{d}u + \frac{1-e^{-a_x S}}{a_x S} x_t + \frac{1-e^{-a_y S}}{a_y S} y_t - \frac{V(S)}{2S},$$

where $V(T)$ is given by (for $T>0$)

$$\begin{aligned} V(T) =& \frac{{\sigma_x}^2}{{a_x}^2}\left(T + \frac{2}{a_x}e^{-a_x T} - \frac{1}{2a_x}e^{-2a_x T} - \frac{3}{2a_x}\right) \\ &+ \frac{{\sigma_y}^2}{{a_y}^2}\left(T + \frac{2}{a_y}e^{-a_y T} - \frac{1}{2a_y}e^{-2a_y T} - \frac{3}{2a_y}\right) \\ &+ 2\rho_{xy}\frac{\sigma_x \sigma_y}{a_x a_y}\left(T + \frac{e^{-a_x T}-1}{a_x} + \frac{e^{-a_y T}-1}{a_y} - \frac{e^{-(a_x+a_y)T}-1}{a_x+a_y}\right). \end{aligned}$$

Hence, the expected value of the continuously-compounded spot rate is given as

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[i_t^S] = \frac{1}{S}\int_t^{t+S} \psi_u \,\mathrm{d}u + \frac{1-e^{-a_x S}}{a_x S}\int_0^t a_x e^{-a_x(t-u)}\lambda_u^x \,\mathrm{d}u \\ + \frac{1-e^{-a_y S}}{a_y S}\int_0^t a_y e^{-a_y(t-u)}\lambda_u^y \,\mathrm{d}u - \frac{V(S)}{2S}.\end{aligned}$$

The integral over ψ can be calculated as (see [16]):

$$\int_t^{t+S} \psi_u \,\mathrm{d}u = \log\left(\frac{P^M(0,t)}{P^M(0,t+S)}\right) + \frac{V(t+S)-V(t)}{2},$$

where $P^M(0,T)$ denotes the observed market zero-coupon bond price at time zero. Since the risk-premium specification changes the drift but not the diffusion coefficients, the variance of i_t^S is the same for the different premium functions considered later:

$$\begin{aligned}\mathbb{V}\left[i_t^S\right] = &\left(\frac{1-e^{-a_x S}}{a_x S}\right)^2 \frac{{\sigma_x}^2}{2a_x}\left(1-e^{-2a_x t}\right) \\ &+ \left(\frac{1-e^{-a_y S}}{a_y S}\right)^2 \frac{{\sigma_y}^2}{2a_y}\left(1-e^{-2a_y t}\right) \\ &+ 2\rho_{xy}\frac{\sigma_x\sigma_y}{a_x+a_y}\left(\frac{1-e^{-a_x S}}{a_x S}\right)\left(\frac{1-e^{-a_y S}}{a_y S}\right)\left(1-e^{-(a_x+a_y)t}\right).\end{aligned}$$

4.5.2 Interpretation of the Two Gaussian Factors

A tempting interpretation of the two-factor structure is to associate one Ornstein–Uhlenbeck factor with the short end of the curve and the other one with the long end. In the present notation, one might try to read x_t as a short-rate factor and y_t as a long-rate factor. This interpretation is useful at most as a heuristic. It is not an intrinsic property of the G2++ model; see also the general model discussion in [16, 11].

A related interpretation is available from the classical two-factor Hull–White parametrisation. In the notation of [16, Sec. 4.2.5], one can write

$$\begin{aligned}\mathrm{d}r(t) &= \big(\theta(t) + u(t) - \bar{a}r(t)\big)\,\mathrm{d}t + \sigma_1\,\mathrm{d}Z_1(t), \\ \mathrm{d}u(t) &= -\bar{b}\,u(t)\,\mathrm{d}t + \sigma_2\,\mathrm{d}Z_2(t),\end{aligned}$$

with correlated Brownian motions Z_1, Z_2. The first equation can be rewritten as

$$\mathrm{d}r(t) = \bar{a}\left(\frac{\theta(t)+u(t)}{\bar{a}} - r(t)\right)\mathrm{d}t + \sigma_1\,\mathrm{d}Z_1(t).$$

Thus the instantaneous reversion level of the short rate is

$$m(t) = \frac{\theta(t)+u(t)}{\bar{a}}.$$

In this parametrisation, the second factor $u(t)$ shifts the reversion level of the short-rate equation and can therefore be read as a component affecting the longer-run development of the short rate. In the additive G2++ representation used here, the corresponding factor interpretation is based on the maturity-dependent coefficients in the spot-rate representation. Before taking expectations, the state-dependent part of i_t^S is

$$\frac{1-e^{-a_x S}}{a_x S}x_t + \frac{1-e^{-a_y S}}{a_y S}y_t.$$

Both factors therefore enter every spot rate i_t^S, with coefficients that depend on the maturity S. In particular, for short maturities both coefficients converge to one. Hence the short end does not isolate one of the two factors. For large maturities, both coefficients tend to zero. If the mean-reversion speeds are strongly separated, the factor with the smaller mean-reversion speed may remain relatively more important for longer maturities. This is a relative statement about the estimated coefficients, not an intrinsic assignment of one factor to the long end of the curve.

This distinction is relevant for calibration. Cap and swaption prices are nonlinear functions of the full parameter vector and therefore of the joint pricing map. In the cap case, longer caps are portfolios of caplets and still contain shorter accrual periods. Moreover, the G2++ caplet variance contains both factor-variance terms and the correlation term, so mean reversion, volatility, and correlation can compensate each other in the price map. The same issue persists for swaptions, where both Gaussian factors enter the semi-analytical pricing function. Consequently, calibration approaches that assign the two factors to separate maturity segments should be interpreted with caution, since this separation is not implied by the model structure. Even the labels x_t and y_t carry no intrinsic economic content, because the two factors can be relabelled together with their parameter blocks without changing the structure of the model.

5 Calibration: What Does It Mean?

The preceding chapters defined the G2++ model and derived the bond and option pricing formulas used later. Calibration links this model to observable financial market data by determining the parameter values.

One important application in practice is the simulation of long-term pension products as discussed in Chapter 2. Another example where calibration is necessary is the fulfilment of stringent regulatory requirements, such as market-consistent valuation under Solvency II, which is analysed in the dedicated case study in Chapter 9 based on the published *openIRM* paper [90].

At the pricing layer, calibration can be viewed as a statistical data-fitting problem rather than an equation-solving exercise: for a given model, instrument set, and objective function, we seek a parameter set that brings model prices close to observed market prices. In the present application, however, this pricing fit is only one part of the calibration chain; stability, interpretability, and suitability for later real-world simulations also matter. In this dissertation, the pricing-layer problem is treated within the framework of *nonlinear regression*, which is analysed in depth in Chapter 7 (specifically Section 7.2). This chapter first defines the scope of calibration, distinguishes between historical estimation and implied calibration, and then discusses the mathematical challenges created by the nonlinear structure of interest-rate models.

5.1 Historical Estimation vs. Implied Calibration

An important distinction in financial mathematics, as highlighted for instance by [27], lies in the source of information used to determine model parameters. This choice depends entirely on the intended application of the model and the probability measure under consideration. For a more in-depth theoretical treatment of these statistical foundations, we refer the reader to the comprehensive discussions in [27].

5.1.1 Historical Estimation

Historical estimation, or statistical estimation, uses a time series of past observations to infer the parameters of the data-generating process. For example, one might use the history of 3-month EURIBOR rates over the last decade to estimate the mean reversion speed and volatility of the interest rate using Maximum Likelihood Estimation (MLE) or Time Series Analysis. The primary goal here is to capture the statistical properties of the market evolution based on realised past behaviour. This perspective is traditionally employed in risk management applications, where the focus lies on the distribution of future returns based on historical patterns. However, relying solely on historical data can be problematic for pricing derivatives, as it may not reflect the current market consensus on future volatility.

5.1.2 Implied Calibration

In contrast, implied calibration largely ignores the history of the asset. Instead, it focuses on the cross-section of currently traded liquid derivatives, such as the prices of caps and swaptions observed at the valuation date $t = 0$. The objective is to find a parameter set Π such that the model reproduces the observed market prices of these liquid instruments as closely as possible.

This difference also changes the statistical interpretation of the optimisation problem. In historical estimation, one usually starts from an explicit stochastic law for observed time series, so likelihood-based methods such as MLE have a natural role. In implied calibration, the data are a market cross-section with quote conventions, bid–ask effects, and instrument-selection choices. Unless an explicit probability model for these quote errors is specified, the calibration objective should not be read as a full likelihood. In Chapter 7, the RMSRE criterion is therefore interpreted more modestly and more usefully as a local weighted least-squares problem: relative price errors are compared in a weighted price space, and the local Jacobian describes how stable the fitted parameters are in that space.

This approach operates strictly under the risk-neutral measure $\mathbb{Q}$. The philosophy behind implied calibration rests on the assumption that the current prices contain all relevant information about the future distribution of the underlying asset required for pricing. As noted by [27], this presents a specific challenge for simulation purposes. While implied calibration efficiently determines the parameters consistent with current market prices, it determines the drift of the process solely based on no-arbitrage arguments. Consequently, under $\mathbb{Q}$, the expected return is the risk-free rate, which does not reflect the actual growth of assets over time. Therefore, a model calibrated solely to market prices is insufficient for the simulation of real-world scenarios, such as pension products. To bridge this gap, the calibration chain explicitly changes from the risk-neutral dynamics to the real-world measure $\mathbb{P}$. Chapter 8 implements this transition through a risk-premium layer. This layer modifies the real-world drift specification of the G2++ factors while the risk-neutral calibration remains the market-consistent pricing input.

5.2 Validation and Criteria for a Robust Calibration

While a strong fit to market prices is the immediate goal of the optimisation described in the previous section, [27] indicate that this addresses the residual size, not the local parameter geometry. The fuller distinction developed in Chapter 7 asks whether the fitted parameter vector is stable and interpretable, and Chapter 10 adds the downstream actuarial output validation.

It is often reasonable to use all available market prices when calibrating the parameters, since this uses the full cross-sectional information in the market data. At the same time, this can increase sensitivity to current-data noise and therefore to *overfitting*, i.e., to adapting the model too strongly to the specific data set at hand. This issue is particularly relevant when models of different dimension are compared: a higher-dimensional model can attain a lower in-sample error than a nested or otherwise more restrictive specification without necessarily describing the underlying market dynamics more accurately.

To address this from a statistical perspective, [27] suggest a validation step where the available data are split into a calibration set (used to determine the parameters) and a validation set (used to test the model on "unseen" instruments). Chapter 7 gives the instrument selection, robustness checks, and cross-instrument transfer interpretation used in this dissertation.

In the context of pension product simulation and asset-liability management, this broader view is essential. A calibration might fit the initial volatility surface well and still produce implausible

long-term simulation paths. Therefore, we extend the validation discussion to include a broader set of economic and regulatory criteria:

1. **Plausibility of Interest Rate Paths:** We analyse the expected value of the short rate over the entire simulation horizon (e.g., 40 years). The simulated rates must align with long-term economic forecasts and not diverge to unrealistic levels. Closely related to this is the shape of the yield curve. A common artefact in multi-factor models is the excessive occurrence of inverted yield curves (where short-term rates exceed long-term rates) in the long run. A robust calibration must produce a distribution of yield curve shapes that reflects realistic market behaviour.

2. **Behaviour of Dependent Processes:** Since the interest rate model serves as a foundation for other asset classes, we must verify the behaviour of dependent processes. Specifically, the simulation of stock returns often depends directly on the stochastic short rate. If the interest-rate calibration leads to excessive volatility, implausible long-run rate levels, or unsuitable rate dynamics, this can distort the drift of the equity model and yield implausible stock-return projections.

3. **Regulatory Stability:** Finally, a further test of validity is the impact on the regulatory classification of financial products. In the German pension-product setting discussed in Chapter 2, products are categorised into Chance-Risk Classes (CRK) based on their simulated payoff profiles. A stable calibration implies that slight changes in market data or the calibration date should not lead to erratic jumps in a product's CRK. Hence, a robust model must ensure that the calculated chance and risk measures and the resulting categorisation remain stable and explainable over time.

5.3 The Challenge of Nonlinearity

Calibration is often described informally as the solution of an optimisation problem. While this description is formally correct, it can be misleading if it suggests that calibration is a purely mechanical or numerically trivial task. In interest rate models such as G2++, the relationship between model parameters and observable market prices is inherently nonlinear, and this nonlinearity poses substantial conceptual and numerical challenges.

To illustrate this point, it is useful to contrast the calibration problem with standard linear regression. In a linear setting, model outputs depend linearly on the parameters, and the resulting least-squares problem admits a unique global minimum with well-understood statistical properties. Small changes in the data lead to small and predictable changes in the estimated parameters.

In contrast, the pricing formulas (see Chapter 4) of the G2++ model depend nonlinearly on the parameter vector $\Pi = \{a_x, a_y, \sigma_x, \sigma_y, \rho_{xy}\}$. As a consequence, the objective function measuring the deviation between model prices and observed market prices exhibits a much more complex structure. This complexity is present in several closely related ways:

1. **Nonlinear Parameter–Price Mapping:** The G2++ model cannot reproduce every arbitrary combination of prices observed in the market. The set of all possible price vectors that the model *can* generate forms a curved surface embedded within the space of market prices.

   This concept is visualized in Figure 5.1 for a calibration against three observed cap prices. The blue surface represents the prices the model is capable of producing. The black

dot represents the observed market prices, the combination of all three cap prices, which typically lies slightly off this surface due to market noise or model limitations.

Calibration is geometrically equivalent to finding the point on the blue surface (the red dot, "Fitted Value") that is closest to the black market data point. Because the surface is curved, a simple linear projection (represented by the orange "Tangent Plane") is only accurate locally.

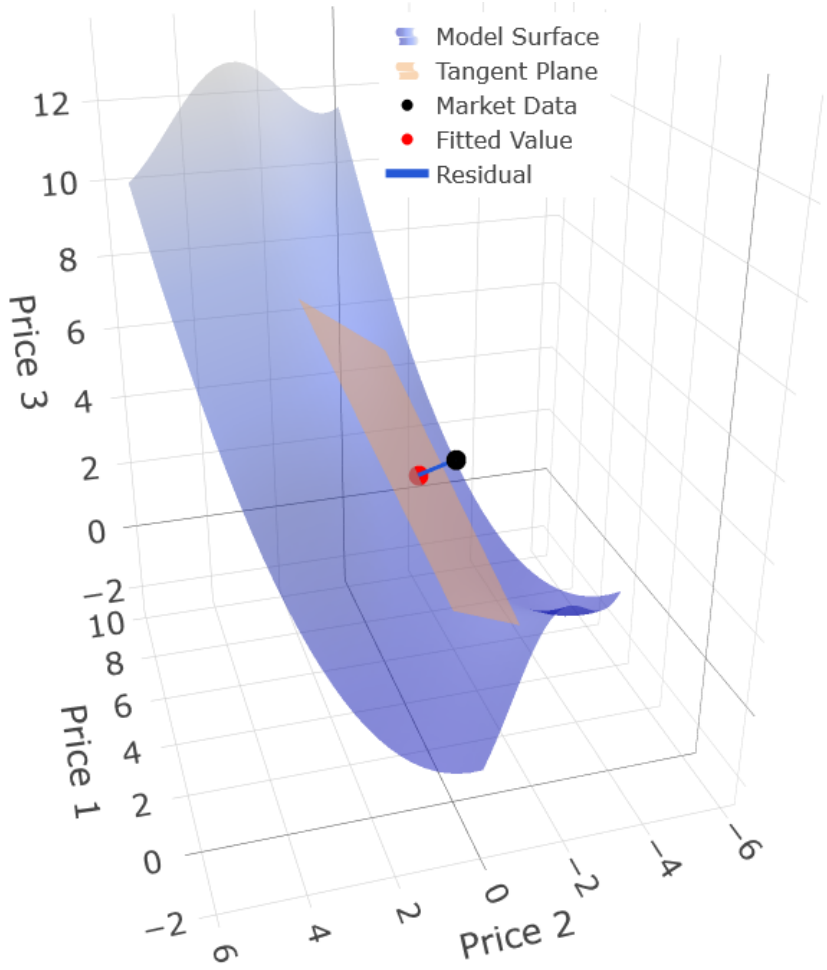


Figure 5.1: Visualising Nonlinear Calibration. The observed market-data point (black) lies off the model surface (blue). Calibration corresponds to an orthogonal projection onto this surface (red), locally approximated by the tangent plane (orange).

2. **Multiple Local Optima:** Due to the nonlinear structure of the pricing functions, the objective function may possess several local minima. Standard gradient-based optimisation algorithms can converge to different parameter sets with almost equal objective values. The optimisation algorithm and its settings, including parameter bounds, are therefore part of the calibration design, not merely implementation details.

3. **Parameter Interdependence and Flat Directions:** In many cases, different parameters influence market prices in a similar way. E.g., an increase in the mean reversion speed of the first factor (a_x) can often be compensated by a corresponding adjustment in the second factor (a_y) or the volatilities to produce nearly identical derivative prices.

   This phenomenon is clearly visible in Figure 5.2, which plots the RMSRE objective function

against the mean reversion parameters. Instead of a distinct minimum, the objective function forms a flat valley along the diagonal. Geometrically, this means we can vary the parameters along the floor of this valley without significantly changing the calibration error. In such regions, the optimisation problem becomes numerically unstable, as small changes in market data can result in large jumps in the calibrated parameters. Chapter 7 quantifies this instability through the local diagnostic geometry.

The later diagnostics are therefore read conditionally on the selected daily calibration solution. They do not prove global optimality of the objective. This is a scope decision: the dissertation combines local diagnostics with the solver, objective, and subset comparisons documented in Chapter 7, rather than turning the calibration problem into a separate global-optimisation study.

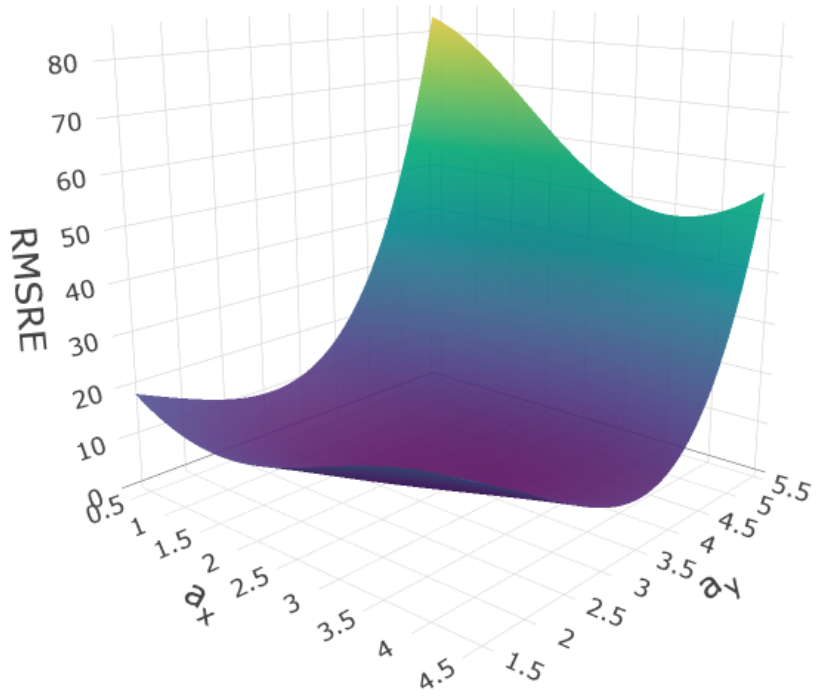


Figure 5.2: Visualising Parameter Interdependence. The objective function (RMSRE) plotted against a_x and a_y forms a flat valley rather than a distinct minimum.

Calibration in nonlinear interest-rate models therefore requires analysis of parameter dependencies, numerical stability, and the economic plausibility of the resulting estimates in addition to minimising an error function. In Chapter 7, we develop diagnostic tools that make these issues explicit. They allow us to assess the robustness and interpretability of a given calibration beyond its residual fit by combining numerical optimisation with statistical regression diagnostics.

6 Calibration Step I: The Term Structure

6.1 Why the initial term structure is part of calibration

Interest rate calibration begins with the construction of the initial term structure before stochastic model parameters are optimised. This point is sometimes hidden by the notation, because the initial curve is often treated as a known market input. In a practical calibration, however, it is already the result of several modelling decisions. The available market quotes have to be selected, cleaned, transformed into discount factors or zero rates, interpolated between observed maturities, and extrapolated where the later application requires maturities beyond the liquid part of the market.

For the G2++ model this is especially important. The model is built in such a way that it can reproduce the observed initial bond prices exactly at time zero. Under the risk-neutral measure $\mathbb{Q}$, the short rate is written as

$$r_t^* = x_t + y_t + \psi_t,$$

where the stochastic factors x_t and y_t describe deviations around the deterministic shift function ψ_t. This deterministic component is chosen so that the model is consistent with the market term structure at the valuation date. In the notation used in Chapter 4, it contains the instantaneous market forward rate $f^M(0,t)$. Hence, the observed curve enters the model before the parameters a_x, a_y, σ_x, σ_y, and ρ_{xy} are calibrated.

This has a direct consequence for the interpretation of calibration results. Two calibrations with identical stochastic parameters but different initial curves are not the same model for an actuarial application. They imply different discount factors, different forward rates, different derivative target prices, and different long-term expected levels after the transition to the real-world measure. Therefore, the initial curve must be treated as the first modelling decision in the calibration chain. It is not a neutral preprocessing step.

This chapter makes this first choice explicit. It first recalls the market objects that are needed for a short-rate calibration. It then discusses how an initial term structure can be constructed from market data and how interpolation and extrapolation enter the procedure. Finally, it explains why the choice between published curves and reconstructed market curves is not neutral for the later risk-neutral calibration, the risk-premium calibration, and the validation of long-horizon simulations. The relevant literature therefore includes curve-construction work, supervisory long-end methodology, and the post-crisis multi-curve literature [55, 86, 59, 42, 43].

6.2 Market objects and conventions

The term structure can be represented in several equivalent ways. For modelling purposes, the most convenient object is often the zero-coupon bond price $P(0,T)$, because the G2++ pricing formulas for caps and swaptions are written in terms of discount factors. For interpretation, however, zero rates and forward rates are usually more transparent.

The continuously compounded zero rate (or zero-coupon yield) for maturity T is defined by

$$i(0,T) = -\frac{1}{T}\log\left(P(0,T)\right), \tag{6.1}$$

or equivalently,

$$P(0,T) = \exp\left(-T\,i(0,T)\right). \tag{6.2}$$

The instantaneous forward rate is obtained from the maturity derivative of the logarithmic bond price,

$$f(0,T) = -\frac{\partial}{\partial T}\log\left(P(0,T)\right). \tag{6.3}$$

Conversely, bond prices can be recovered from the forward curve by

$$P(0,T) = \exp\left(-\int_0^T f(0,u)\,\mathrm{d}u\right). \tag{6.4}$$

These formulas show why the representation of the initial curve matters. If the model requires $f^M(0,t)$, then a curve available only at a finite set of zero-rate maturities must be transformed into a sufficiently smooth function. If the curve is noisy, not differentiable, or extrapolated in an economically implausible way, the deterministic shift function inherits these features. This can later affect the simulated short rate even if the stochastic calibration itself appears numerically successful.

In addition to the mathematical representation, one must specify market conventions. A quoted interest rate is not yet a model input. It comes with a day-count convention, a compounding convention, a payment frequency, and a definition of the underlying tenor. Derivative quotes add a second layer of convention. Caps and swaptions are frequently quoted by implied volatilities under market-standard formulas, while the G2++ calibration uses prices as targets. The observed volatility quote therefore has to be translated into a price before it can be compared with the model price; see Chapter 3 for the market-side quote conventions and Chapter 4 for the corresponding G2++ pricing functions. This translation is only meaningful relative to the discount and forward curves used in the market formula. Curve construction and interpolation are therefore part of the calibration design, as emphasised by [55].

For this reason, the initial term structure is understood here as a collection of consistent market objects: discount factors, zero rates, forward rates, and the conventions that connect quoted instruments to those quantities. The curve defines the deterministic environment in which the stochastic interest-rate model is embedded.

6.2.1 A numerical convention check: yields, discount factors, and dates

The practical meaning of these conventions becomes visible already in a very small reconstruction exercise. A simple convention check inspects the same curve data both as quoted yields and as discount factors. A representative OIS curve snapshot contained, for example, an overnight yield of 0.0358064 and an overnight discount factor of 0.99990362. If the maturity is interpreted

as one actual day and the quote is compounded annually with a 365-day basis, the two numbers agree:

$$\frac{1}{(1+0.0358064)^{1/365}} = 0.99990362. \tag{6.5}$$

This small calculation shows that the discount factor is not obtained from the labelled maturity alone. It is obtained from the actual maturity date, the day-count basis, and the compounding convention.

The same calculation can be inverted to infer the year fraction that is implicitly used by a yield–discount-factor pair. Under annual compounding, the implied maturity is

$$T = \frac{\log(1/P(0,T))}{\log(1+y(T))}, \tag{6.6}$$

and under continuous compounding it is

$$T = -\frac{\log(P(0,T))}{y(T)}. \tag{6.7}$$

In this convention check, the implied maturities recovered the actual short-end calendar structure rather than the rounded tenor labels: the same snapshot implied one day for ON, two days for TN, nine days for 1W, 34 days for 1M, 94 days for 3M, 369 days for 1Y, and 1462 days for 4Y. These implied maturities document the role of actual payment dates and market calendars in the curve reconstruction. If the labelled tenors were used mechanically as exact year fractions, the resulting discount factors, forward rates, and later derivative target prices would already be inconsistent before the G2++ parameters are touched.

6.3 Constructing the initial term structure

In an idealised model, zero-coupon bond prices are available for every maturity. In actual market data, this is not the case. One observes deposits, swaps, overnight indexed swaps, government bond yields, or published curve parameters, depending on the data source and the application. The term structure used for calibration must therefore be constructed.

There are two broad approaches. The first approach is to use a published term structure. Examples include central-bank or supervisory curves, such as Bundesbank or EIOPA term structures. The advantage is transparency and reproducibility: the curve is defined by an external institution, and its construction is documented. The disadvantage is that the published curve may not match the exact collateralization, tenor, or instrument conventions of the derivative data used for calibration.

The second approach is to reconstruct the curve from market instruments. This is closer to derivative-pricing practice. In such a procedure, liquid money-market and swap instruments are used to bootstrap discount factors or zero rates. The advantage is consistency with the market quotes used on the valuation date. The disadvantage is that the procedure requires more convention choices and more data cleaning. It also introduces a stronger dependence on the quality and liquidity of the selected instruments.

For the present calibration framework, the chosen curve approach must be adequate for the later application. A curve used only to price short-dated derivatives can be judged differently from a curve that serves as the basis for a forty-year pension-product simulation. In the latter case, long-end behaviour, extrapolation, and stability over time are part of the scientific object.

In insurance applications, this long-end question has a specific regulatory counterpart. Solvency-II-style valuation requires a risk-free term structure also beyond maturities for which deep, liquid, and transparent market information is available. The Smith–Wilson framework and the Ultimate Forward Rate methodology are examples of how supervisory practice turns extrapolation into a controlled modelling convention [86, 42, 43, 41]. For long-horizon calibration, this regulatory perspective implies a stricter methodological requirement: the analysis must document where market information ends, where extrapolation begins, and how this choice affects subsequent simulation and validation.

6.3.1 Nelson-Siegel-Svensson method

A common parametric representation of yield curves is the Nelson-Siegel-Svensson family. The Nelson-Siegel model was introduced in [79], and the Svensson extension adds a second curvature component; see [87]. The model is widely used because it is flexible enough to represent the typical level, slope, and curvature patterns of yield curves while remaining low-dimensional. Practical calibration aspects are discussed, for example, in [51], and the Bundesbank documentation and daily time-series catalogue provide institutional references for its use in German term-structure estimation [32, 33].

In the Svensson specification, the continuously compounded spot rate for maturity T is written as

$$i(0,T) = \beta_0 + \beta_1 \frac{1-e^{-T/\tau_1}}{T/\tau_1} + \beta_2 \left(\frac{1-e^{-T/\tau_1}}{T/\tau_1} - e^{-T/\tau_1} \right) \tag{6.8}$$

$$+ \beta_3 \left(\frac{1-e^{-T/\tau_2}}{T/\tau_2} - e^{-T/\tau_2} \right). \tag{6.9}$$

The corresponding instantaneous forward rate is obtained from

$$f(0,T) = i(0,T) + T\frac{\partial i(0,T)}{\partial T},$$

and therefore

$$f(0,T) = \beta_0 + \beta_1 e^{-T/\tau_1} + \beta_2 \frac{T}{\tau_1} e^{-T/\tau_1} + \beta_3 \frac{T}{\tau_2} e^{-T/\tau_2}. \tag{6.10}$$

The coefficients provide a compact description of the fitted curve shape on a fixed date. The parameter β_0 determines the long-run level of the spot and instantaneous forward curves. The parameter β_1 affects the short end relative to the long end and therefore the slope. The parameters β_2 and β_3 generate curvature effects, while τ_1 and τ_2 locate these effects along the maturity axis. This curve-shape reading concerns the fitted curve itself rather than separately observable economic state variables. In daily time series, changes in the NSS coordinates may reflect yield-curve movements, coordinate effects of the nonlinear parametrisation, and interactions among the basis functions.

The Nelson-Siegel-Svensson specification also imposes a smoothing and extrapolation structure. This structure can suppress local market information, and the long end may be driven by the parametric form once observed maturities become sparse. In the present calibration framework, the NSS curve is therefore interpreted as a curve-generating convention whose suitability depends on the intended application.

Daily NSS Parameter Coordinates. This perspective also applies to the daily Bundesbank NSS parameters. The Bundesbank NSS series is a transparent institutional reference, but strong

movements in individual NSS coordinates, especially in β_1, β_2, β_3, τ_1, or τ_2, are not automatically evidence of equally strong movements of the fitted curve. They may instead reflect coordinate instability of the NSS parameterisation. The calibration problem is nonlinear and may contain shallow local optima or flat valleys; in addition, the loading functions can become strongly co-dependent or nearly collinear, so materially different parameter vectors may generate very similar fitted curves [51, 5, 74]. Coordinate instability alone does not imply that the implied curve is unstable.

For the derivative calibration below, the relevant objects live in maturity space: discount factors, zero rates, forward rates, ATM strikes, and monetary option price targets. NSS coordinate instability becomes a direct calibration-input issue only if it translates into unstable or convention-inconsistent curve-implied quantities. Thus the daily Bundesbank NSS series is used here as a transparent curve-generating benchmark. The individual coordinates are not treated as directly interpretable economic state variables.

6.3.2 Interpolation and extrapolation methods

Even when a parametric curve is not used, interpolation is unavoidable. Market instruments are available at discrete maturities, but the model requires discount factors for all reset, payment, expiry, and maturity dates appearing in the calibration instruments. Caps require discount factors at caplet reset and payment dates. Swaptions require discount factors for all fixed-leg cash-flow dates of the underlying swap. If these dates do not coincide exactly with quoted market maturities, the missing values must be interpolated.

Interpolation can be performed on zero rates, discount factors, logarithmic discount factors, or forward rates. These choices are not equivalent. Linear interpolation of zero rates produces a different forward curve from linear interpolation of logarithmic discount factors. Smooth interpolation of forward rates can make the deterministic shift function smoother, but it may also create spurious distortions if the observed data are sparse. A calibration report must therefore record the interpolated quantity and the interpolation method. This is the practical point made in the curve-construction literature: interpolation choices affect the curve's economic and hedging properties; graphical smoothness alone is not a sufficient criterion [55].

Extrapolation requires separate attention. It determines the curve outside the liquid market segment. For short-horizon pricing this may be of secondary importance. For actuarial applications, especially PIA- or openIRM-related long-horizon simulations, the long end can directly affect the distribution of bond returns and the plausibility of simulated yield-curve shapes. Supervisory frameworks such as EIOPA's risk-free interest-rate term structures explicitly address long-end construction and ultimate forward rate methodology [86, 44, 45, 42, 43, 41]. In the present chapter, these frameworks serve as methodological references for documenting long-end choices. A long-horizon calibration must state where market information ends, how extrapolation is imposed, and how this choice enters subsequent simulation and validation.

6.3.3 Numerical reconstruction of the curve behind cap quotes

The available cap data contained market prices, ATM strikes, log-normal volatilities, and shifted log-normal volatilities. The cap names also identified the relevant tenor structure. The short cap instruments, such as EU1Y3LATM, EU18M3LATM, and EU2Y3LATM, refer to 3-month EURIBOR. The remaining cap instruments in the studied portfolio, from 3Y to 30Y, use a 6-month tenor. Smoothness, external documentation, tenor consistency, and quote reconstruction have to be assessed jointly. In the empirical check below, the decisive criterion is whether a curve, together with the relevant tenor convention, reproduces the quoted ATM strikes and prices.

Under the maintained single-curve ATM convention, the cap strike over the schedule $\mathcal{T} = \{t_m, \ldots, t_n\}$ is reconstructed as the forward swap rate, with $\delta_j = t_j - t_{j-1}$,

$$K = S_{m,n}(0) = \frac{P(0, t_m) - P(0, t_n)}{\sum_{j=m+1}^{n} \delta_j P(0, t_j)}. \tag{6.11}$$

In the empirical part of the chapter, this identity serves as a diagnostic check for the initial curve. The quoted cap strikes are compared with strikes reconstructed from candidate curves and tenor schedules. A candidate curve that reproduces the quoted strikes is internally close to the convention under which the market data were generated. A systematic reconstruction error indicates a convention mismatch that can later appear as a pricing residual or as a parameter effect.

The reconstruction compared published NSS curves, OIS 3M and OIS 6M discount curves, and a EUR standard 6M zero curve. The OIS and EUR curves were available at discrete maturities. For long cap schedules this creates missing cash-flow dates, for instance at 25.5Y or 26.5Y in a 30Y cap with 6M tenor. Two completion methods were compared. The first used the directly quoted discount factors wherever available and filled only missing maturities by an NSS fit. The second replaced the full curve by the fitted NSS curve. The direct-discount completion gave materially smaller deviations. Figure 6.1 illustrates the reason: the fitted Svensson curve is useful as a smooth continuation device, while observed discount factors remain the preferred input at quoted maturities.

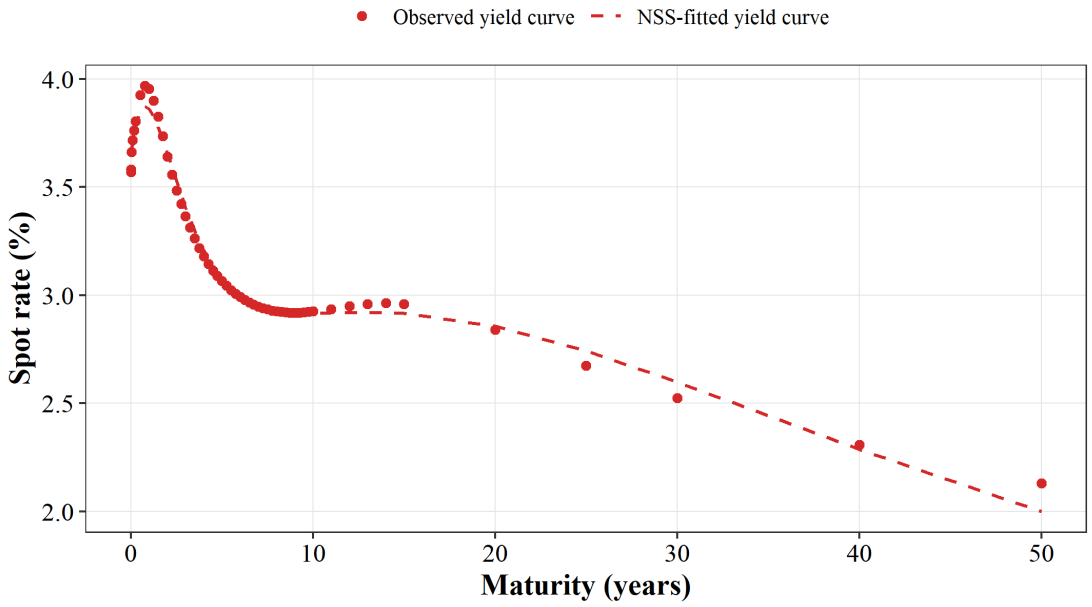


Figure 6.1: Observed and fitted Svensson yield curve for an OIS 3M curve snapshot. The figure illustrates the empirical distinction between using observed curve points directly and using a parametric curve as a smoothing or completion device.

This point becomes clearer when the distribution of strike-reconstruction deviations is examined across the full complete-case window from 1 September 2016 to 2 June 2025. Here, complete-case means that a date contains the cap quotes, tenor schedules, shifted-volatility information, and curve inputs required for the corresponding reconstruction. Figure 6.2 makes the role of curve choice visible over this long complete-case sample. The Bundesbank and ECB NSS curves generate materially larger deviations across almost the whole cap spectrum, the OIS 3M curve is the clear match for the 3M-tenor caps, and the direct EUR Standard 6M and OIS 6M curves form the two clear best-performing conventions for the 6M-tenor caps.

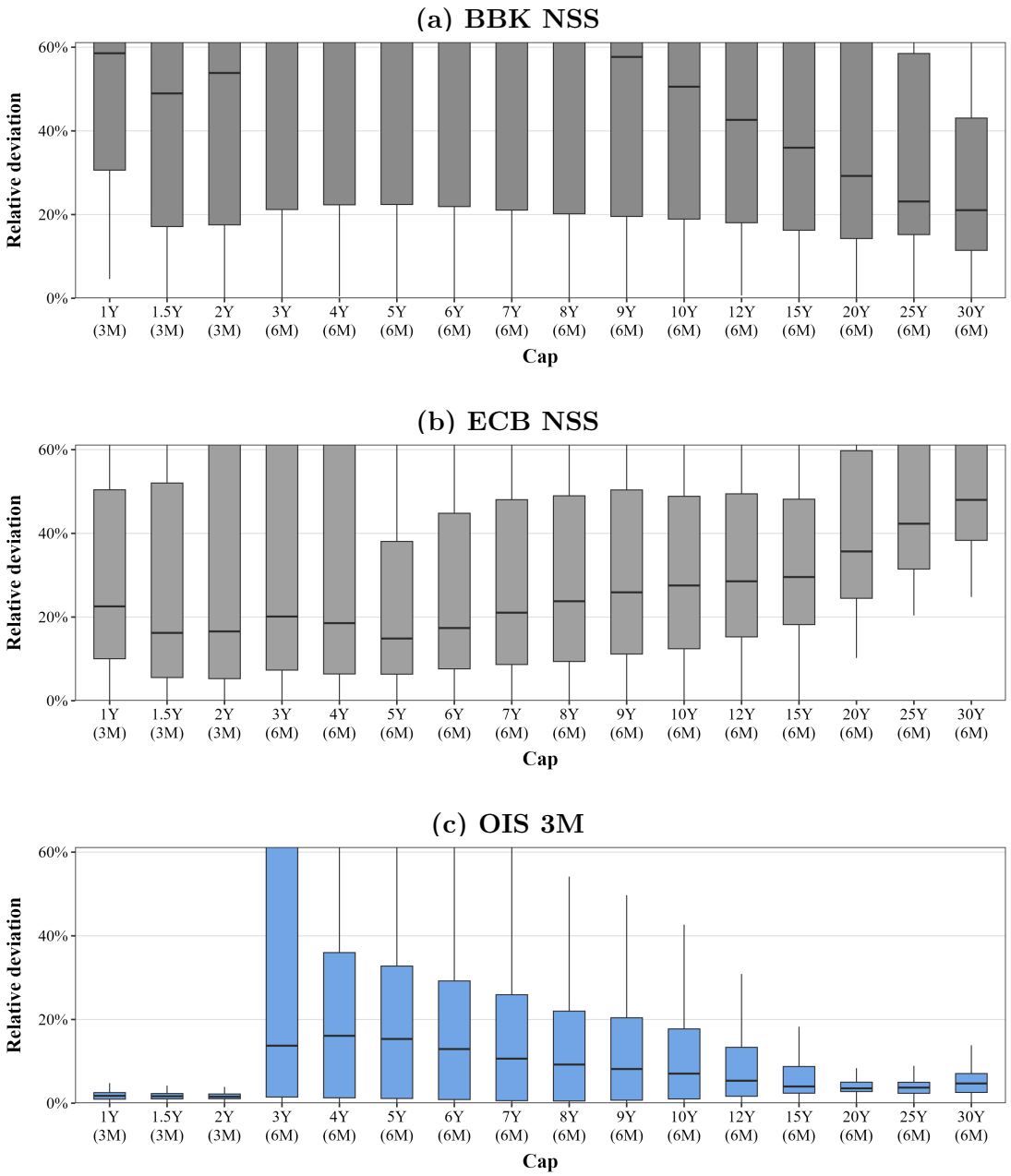


Figure 6.2: Boxplots of absolute relative ATM-strike reconstruction deviations under five curve choices on the complete-case sample from 1 September 2016 to 2 June 2025. The labels on the x-axis report cap maturity and underlying tenor; for example, 3Y (6M) denotes a three-year cap with a 6-month tenor. Published NSS curves such as BBK and ECB are visibly less consistent with the cap quotation convention, OIS 3M is the appropriate reconstruction curve for the 3M-tenor caps, and the direct EUR Standard 6M and OIS 6M curves form the two best-performing conventions for the 6M-tenor caps.

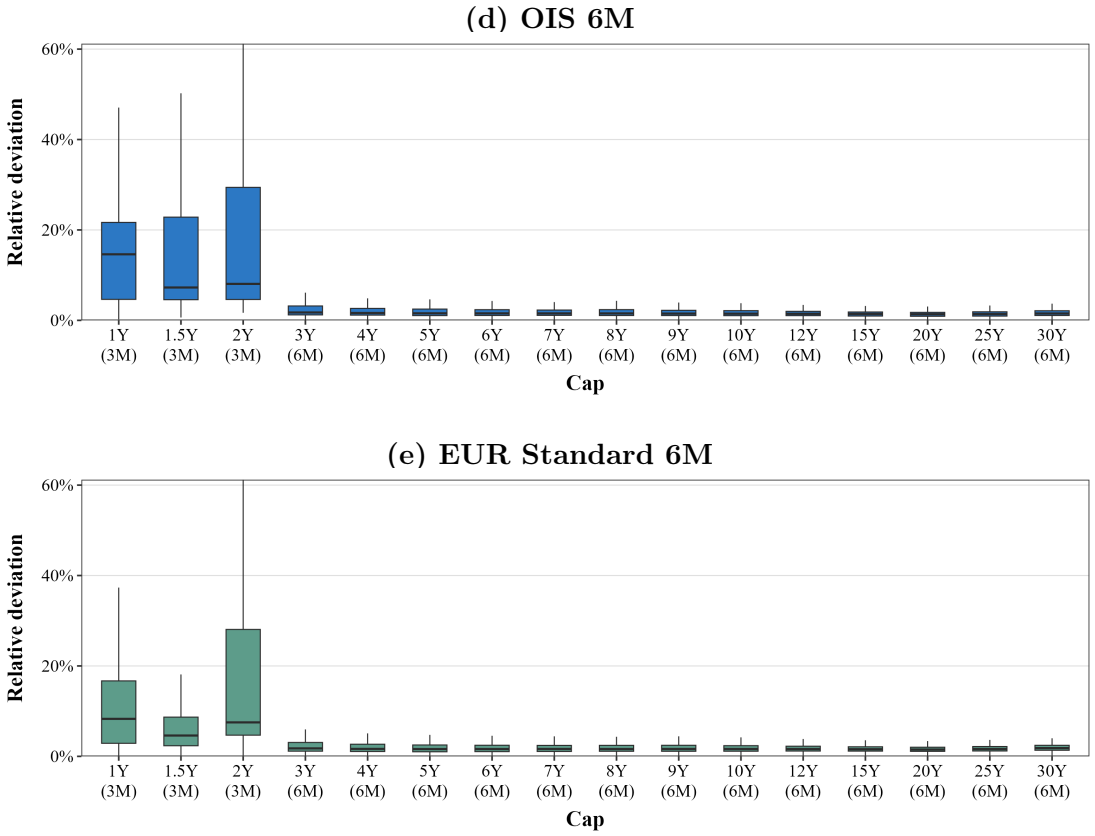


Figure 6.2: Continued.

The numerical evidence is summarised in Table 6.1. The deviations are absolute relative deviations, defined by

$$\left| \frac{\text{market value} - \text{reconstructed value}}{\text{market value}} \right|. \tag{6.12}$$

The summary table uses the same synchronized market-data basis as the swaption quote-to-price reproduction window, namely the complete-case sample from 1 September 2016 to 2 June 2025. Figure 6.2 and Table 6.1 therefore refer to the same long empirical window. The ECB NSS curve remains visible in the figure as an additional published-curve comparison, while the table restricts attention to the conventions for which strike and shifted-price reconstruction can be reported jointly.

The table has three main implications. First, the tenor information in the instrument name is empirically active over the full synchronized sample. The 3M-tenor caps are best reconstructed with the OIS 3M curve. The 6M-tenor caps are best reconstructed by the direct 6M market-curve family. Within that 6M family, OIS 6M is marginally tighter in strike reconstruction, and EUR Standard 6M is marginally tighter in the reported median price deviation. Second, published NSS curves such as the Bundesbank curve remain valuable as transparent actuarial references, but they reconstruct the cap quotation convention much less accurately. On the long window, the Bundesbank strike deviations are an order of magnitude larger than those of the direct OIS or EUR market curves. Third, full NSS replacement can distort monetary price targets despite its mathematical smoothness. The large relative price deviations in the off-tenor 3M rows and especially in the BBK and full-NSS rows are best read as a warning about price-space diagnostics with small denominators. They make the same practical point more sharply: quote reconstruction must be controlled before residuals are attributed to the stochastic interest-rate model. The later calibration evidence therefore treats the OIS-based

Instrument group	Curve convention	Median strike deviation	90% strike deviation	Median price deviation
3M-tenor caps	OIS 3M direct discounts	1.64%	3.15%	5.39%
3M-tenor caps	OIS 3M NSS replacement	3.37%	11.3%	11.4%
3M-tenor caps	OIS 6M direct discounts	9.23%	36.6%	31.4%
3M-tenor caps	BBK NSS published reference	53.8%	199.2%	90.8%
6M-tenor caps	EUR 6M direct discounts	1.64%	5.07%	1.15%
6M-tenor caps	OIS 6M direct discounts	1.50%	5.10%	1.31%
6M-tenor caps	OIS 6M NSS replacement	3.19%	12.9%	2.34%
6M-tenor caps	BBK NSS published reference	51.0%	378.8%	33.9%

Table 6.1: Reconstruction quality for ATM cap strikes and shifted-Black cap prices under selected curve conventions on the synchronized long window from 1 September 2016 to 2 June 2025. The table reports absolute relative deviations. Direct discount-factor use means that observed discount factors are used where available and the fitted curve is used only to complete missing cash-flow dates.

curve as the main market-consistent front end. Published-curve variants serve as alternative actuarial or robustness conventions.

6.3.4 From volatility quotes to monetary prices

The second numerical layer concerns the conversion from implied volatility to price. Market data for caps and swaptions are commonly communicated in volatility units, but the G2++ calibration compares prices. The Black–76 formula is therefore a translator between the market quote and the model objective [13]. For later reference it is convenient to write the Black–76 building block with a sign parameter χ, where $\chi = 1$ gives the call-type payoff used for caplets:

$$B_{\chi}^{\mathsf{Black}}(F, K, \nu) = \chi \left[F\Phi\left(\chi d_{+}\right) - K\Phi\left(\chi d_{-}\right) \right], \qquad d_{\pm} = \frac{\log(F/K)}{\nu} \pm \frac{\nu}{2}. \tag{6.13}$$

In a caplet application, F is the forward rate $L(0; t_{i-1}, t_i)$, $\nu = \sigma\sqrt{t_{i-1}}$, and the cap price is obtained by summing the discounted caplet prices

$$\mathrm{Cap}^{\mathsf{Mkt}}(0) = N \sum_{i=m+1}^{n} P(0, t_i)\delta_i B_1^{\mathsf{Black}}\left(L(0; t_{i-1}, t_i), K, \sigma_{m,n}\sqrt{t_{i-1}}\right). \tag{6.14}$$

For swaptions, the corresponding Black–76 translation uses the forward swap rate $S_{m,n}(0)$ and the swap annuity $A_{m,n}(0)$ induced by the same curve:

$$\mathrm{Swpt}^{\mathsf{Mkt}}(0) = NA_{m,n}(0)B_{\chi}^{\mathsf{Black}}\left(S_{m,n}(0), K, \sigma_{m,n}^{\mathsf{swpt}}\sqrt{T_m}\right), \tag{6.15}$$

where $\chi = 1$ corresponds to a payer swaption and $\chi = -1$ to a receiver swaption. These formulas are deliberately written in terms of the same discount factors, forwards, and annuities that appeared in the curve reconstruction step. A volatility quote cannot be turned into a unique price without these curve inputs.

Negative-rate periods require one further convention. The unshifted Black–76 formula treats the forward rate entering the option formula as a positive lognormal quantity and contains the term $\log(F/K)$. It is therefore not suitable as a quote-to-price translator when forwards or strikes are non-positive. The market-data work therefore uses shifted log-normal volatilities, with a shift s applied both to the strike and to the forward rate:

$$\mathrm{Cap}^{\mathsf{Mkt,shift}}(0) = N \sum_{i=m+1}^{n} P(0, t_i)\delta_i B_1^{\mathsf{Black}}\left(L(0; t_{i-1}, t_i) + s, K + s, \widetilde{\sigma}_{m,n}\sqrt{ti-1}\right). \tag{6.16}$$

In the available cap data, the shifted-volatility quote is the relevant continuous bridge through the negative-rate years; the working convention in this chapter is a 3% shift. As an additional single-year consistency check, included to isolate the quote-to-price layer from the long-window strike reconstruction, the 2023 shifted-volatility comparison confirms that the curve convention remains visible even after moving from prices back to shifted volatilities. Using the OIS 6M curve gives a median shifted-volatility deviation of 0.91%, compared with 4.83% under the OIS 3M curve over the same check. This comparison concerns the market-data front end and the internal consistency of the quote-to-price conversion. The stochastic G2++ parameters can be assessed only after this conversion has been fixed.

A worked single-date cap example

To illustrate the quote-to-price chain on one concrete observation, consider the 5Y cap with 6M tenor on 1 June 2021. This date lies inside the same synchronized 2016–2025 market-data window that is later used for the swaption calibration. The stored cap panel reports an ATM strike of −0.23%, a shifted volatility of 14.46%, and a market price of 119 basis points.

Using the direct OIS 6M curve, the quantities entering the forward-swap-rate formula are

$$P(0, 0.5) = 1.0026, \qquad P(0, 5) = 1.0133,$$
$$A = \sum_{j=2}^{10} 0.5 P(0, t_j) = 4.5500,$$

so that the internally consistent ATM strike reconstruction becomes

$$K = S_{0.5,5}(0) = \frac{P(0, 0.5) - P(0, 5)}{A} = -0.2345\%.$$

This is essentially identical to the stored market strike. Feeding this strike, the observed shifted volatility, the 3% shift, and the same OIS 6M discount curve into (6.16) gives a reconstructed cap price of 117.7 basis points, that is, a deviation of about 1.1% from the stored market price. The direct EUR 6M curve gives 118.5 basis points on the same observation, confirming the near-tie of the two direct 6M market curves already visible in Table 6.1. By contrast, the published Bundesbank NSS curve implies a strike of −0.5682% and a cap price of only 101.9 basis points, a deviation of about 14.4%. This example shows that the monetary calibration target is already affected by the front-end curve convention before the stochastic model is calibrated.

Swaption repricing on the synchronized window

The same quote-to-price issue appears on the swaption side. The swaption market-data reproduction covers 1,968 complete-case dates from 1 September 2016 to 2 June 2025 and 196 swaptions per date, hence 385,728 swaption observations in total. Using the observed Black–76 swaption volatilities together with the curve-implied forward swap rates and annuities, the direct EUR 6M curve gives a median relative price deviation of 1.03%, and the direct OIS 6M curve gives 1.12%. The corresponding median strike deviations are 0.91% and 1.08%. The qualitative message is the same as for caps: the direct 6M market-curve family remains close to the quotation convention, and the residual differences between the two direct 6M curves are economically small.

The cap and swaption checks differ in their numerical implementation. For caps, the negative-rate period requires an additional shifted-volatility consistency check. For swaptions, the quote-side reproduction is already implicit in the observed Black–76 volatility surface. The reconstruction check is therefore reported directly through the monetary price translation and the associated ATM-strike consistency. This Chapter-6 reproduction window is broader than the later

numerical-Jacobian diagnostics panel, which additionally requires aligned calibrated parameter vectors, datewise valid swaption subsets, finite-difference evaluations, and numerical-governance checks.

The empirical front end is fixed by a small set of curve and derivative-market inputs. This matters for the transition to Chapter 7: the later risk-neutral calibration takes this synchronized input layer as fixed and does not reopen the front-end convention.

More generally, the market data used later for caps and swaptions are convention-dependent inputs. The empirical task is to reconstruct economically meaningful target prices from these quotes before calibrating the model to the resulting price targets. Curve choice, interpolation, ATM interpretation, day-count and accrual conventions, volatility scaling, vendor premium units, and the treatment of missing observations are therefore part of the empirical model-building problem itself.

In the logic established in Chapter 3, Black-type formulas serve as benchmark quote translators. The present chapter fixes that front end. The diagnostics in Chapter 7 can then be read as statements about the chosen pricing map and its calibrated parameters.

6.4 On the choice of the initial term structure

The choice of the initial term structure is evaluated by its role in the full calibration chain. A curve is application-adequate if it gives a coherent starting point for market-consistent pricing under $\mathbb{Q}$, for the transition to $\mathbb{P}$, and for the validation of simulated scenarios. The numerical reconstruction above provides the empirical basis for this choice. The following subsections summarise the modelling implications of the main curve families.

6.4.1 Published term structures

Published term structures are attractive because they provide a clear external reference. Their construction is usually documented, and the data can often be reproduced. Bundesbank and EIOPA data, for example, provide structured curve information that can be used to reconstruct zero rates, discount factors, and forward rates over time. This transparency is valuable in actuarial applications, where calibration choices have to be documented and explained.

For EIOPA risk-free-rate curves, the external reference has an additional regulatory meaning. EIOPA publishes the relevant risk-free interest-rate term structures on a monthly basis for the calculation of technical provisions under Solvency II; the accompanying technical documentation and UFR methodology make the construction auditable and application-specific [42, 43, 41]. This gives the curve a role that differs from desk-level derivative pricing. It can serve as a transparent actuarial reference, even when the derivative quotation convention points to a different market curve.

The limitation is convention consistency. A published curve may not match the collateralisation, tenor, and instrument conventions embedded in the derivative quotes used for calibration. If swaption volatilities are quoted relative to a market swap curve and the G2++ calibration uses a different published curve for discounting and forwarding, the calibration target becomes a hybrid object. Such a choice can still be defensible in an actuarial setting, especially when transparency and long-term stability are central. It has to be stated explicitly, because part of the residual pattern may then come from curve inconsistency instead of the stochastic interest-rate model.

6.4.2 Swap-derived and OIS-derived term structures

Swap-derived and OIS-derived curves are closer to the usual derivative-pricing setting. In a collateralised market, OIS discounting is the standard reference for many derivatives, and tenor-specific forwarding curves may be needed for floating-rate projections. Modern interest-rate modelling therefore distinguishes discounting and forwarding curves; see, for example, the broader modelling discussions in [4, 12, 59].

Using such curves can improve consistency between the calibration instruments and the monetary target prices. The numerical reconstruction in this chapter supports this point: the direct 6M market-curve family is closest to the 6M cap and swaption quotation convention, with OIS 6M and EUR Standard 6M giving very similar reconstruction quality. For the short 3M-tenor caps, the OIS 3M curve is the relevant convention. These findings make the curve choice an empirical front-end decision, not a purely cosmetic smoothing choice.

The same market proximity increases the number of conventions that must be controlled. The curve construction depends on selected instruments, calendars, day-count conventions, interpolation rules, and market-data cleaning. On a single valuation date these details may appear minor. Over a long empirical period such as 2016–2025, changes in available instruments, liquidity, or quote conventions can introduce artificial jumps or regime-dependent effects. The market-data line is therefore part of the empirical contribution of the dissertation: the chosen curve construction has to remain stable, interpretable, and application-adequate over the period used for calibration diagnostics and validation.

6.4.3 Single-curve simplification and its limits

The broader market benchmark is multi-curve: discounting and forwarding are, in general, distinct objects [59]. The present calibration framework uses a more restrictive single-curve representation built around an OIS-based risk-free reference curve. This is an application-driven simplification. The actuarial target picture is long-horizon simulation, scenario generation, and validation for products and reference portfolios whose behaviour depends mainly on the overall level and shape of the risk-free term structure.

This simplification has concrete limitations. A single-curve setup suppresses basis spreads as separate state variables. It cannot reproduce all cross-curve effects that may matter in collateralised derivative pricing. Part of the residual structure of a cap or swaption calibration may therefore reflect the simplified curve convention used to reconstruct the target prices. The calibration exercise remains useful, but the interpretation of residuals changes.

The same limitation affects parameter interpretation. If a single-curve short-rate model is calibrated to market prices that embed multi-curve basis effects, the stochastic G2++ parameters may absorb part of the omitted basis-spread structure. This can distort volatility and correlation estimates relative to a full multi-curve valuation model. The chosen design accepts this limitation for tractability, ALM-style consistency, and comparability across the calibration chain. A full multi-curve G2++ calibration is a natural extension outside the empirical design used here.

The scientific justification of the chosen simplification rests on internal consistency. The same curve convention is carried through discount-factor construction, forward-rate and swap-rate reconstruction, ATM-strike determination, and the translation of quoted implied volatilities into monetary target prices. Under this fixed convention, the diagnostics in Chapter 7 can be interpreted as statements about the chosen pricing map, its calibrated parameters, and its local stability.

This also clarifies the assignment of roles between chapters. Chapter 3 explains the multi-curve framework as the natural market benchmark. The present chapter fixes the single-curve OIS-based empirical front end. Chapter 7 then takes this convention as given and studies the consequences for fit, identifiability, leverage, influence, and stability. In this sense, the single-curve choice is the first controlled modelling restriction in the calibration chain.

Operationally, the later cap and swaption calibrations use the direct OIS-based curve family as the main market-consistent baseline, with the OIS 6M convention serving as the main reference for the 6M-tenor portfolio. The direct EUR 6M curve is treated as a near-equivalent 6M robustness convention. The published BBK and ECB NSS families remain explicit comparison cases.

6.5 Connection to calibration instruments

The initial curve enters the calibration of risk-neutral parameters in two ways. First, it determines the discount factors used in the pricing formulas. Second, it determines the forward rates and swap rates used to define at-the-money strikes and to translate market volatility quotes into prices.

For a caplet over the period $[t_{i-1}, t_i]$, with $\delta_i = t_i - t_{i-1}$, the relevant simple forward rate in a single-curve setting is

$$L(0; t_{i-1}, t_i) = \frac{1}{\delta_i}\left(\frac{P(0, t_{i-1})}{P(0, t_i)} - 1\right). \tag{6.17}$$

For a swaption, the at-the-money strike is the forward swap rate, which is also a function of zero-coupon bond prices. Thus, even before the G2++ parameters are optimised, the curve determines part of the calibration target.

This dependency explains why calibration diagnostics must later be interpreted with the initial curve in mind. A large residual for a particular maturity may indicate that the stochastic model cannot reproduce the volatility structure at that maturity. But it may also indicate a mismatch between the curve construction, the quote conversion, and the derivative instrument used as a target. A robust calibration analysis must therefore document the curve and the instruments together.

6.6 Checklist for an application-adequate initial curve

The following checklist summarises the practical role of the initial curve in the maintained calibration chain. It is placed before the risk-neutral calibration chapter because these questions must be answered before calibration residuals are interpreted.

1. **Source:** Which market or institutional data source provides the curve information?
2. **Market object:** Is the curve represented as discount factors, zero rates, forward rates, or parametric curve parameters?
3. **Conventions:** Which compounding, day-count, calendar, and tenor conventions are used?
4. **Interpolation:** Which quantity is interpolated between observed maturities?
5. **Extrapolation:** How is the long end determined, and where does liquid market information end?

6. **Instrument consistency:** Are the discount and forward curves consistent with the cap or swaption quotes used for calibration?

7. **Application adequacy:** Is the curve suitable for the long-horizon actuarial use case rather than only for short-horizon pricing?

8. **Empirical stability:** Does the curve construction remain stable over the 2016–2025 sample, including the structural break after 2022?

Once these questions are settled, the risk-neutral calibration under $\mathbb{Q}$ can be read as a statement about the stochastic parameters of the G2++ model. Discounting, forwarding, ATM-strike reconstruction, and quote-to-price translation are then part of the fixed empirical front end, not additional uncontrolled sources of variation. Chapter 7 starts from this point and studies how the G2++ parameters behave under this fixed market-data convention.

7 Calibration Step II: The Risk-Neutral Parameters

The preceding chapters fixed the theoretical short-rate model and the initial term structure convention. This chapter determines the risk-neutral G2++ parameters from liquid market instruments. It translates cap and swaption quotes into model-price targets, formulates the constrained calibration problem, and then studies whether the resulting parameter vector is locally stable, identifiable, and suitable as an input for the later real-world calibration and actuarial application.

The market instruments and Black-type quotation conventions were introduced in Chapter 3, while Chapter 6 fixed the initial-curve convention and the quote-to-price translation used for the empirical sample. Using that benchmark-quote interpretation, the present chapter calibrates G2++ to reconstructed monetary prices; Black volatilities enter only through the earlier translation layer, not as structural parameters of the short-rate model.

Having established the pricing formulas, the calibration problem focuses on parameter selection and diagnostic analysis. This includes how the constrained calibration problem should be posed, how the RMSRE criterion behaves as a weighted least-squares problem, which maturities anchor the fitted parameter vector, and whether the resulting local geometry is stable enough for the later transition to the calibration under $\mathbb{P}$ and the actuarial application.

7.1 Calibrating the risk-neutral parameters

After the initial market forward-rate curve $f^M(0,t)$ has been fixed, the stochastic risk-neutral parameters can be calibrated. This step uses the market price vector constructed above and searches for the G2++ parameter vector that reproduces it as closely as possible under the chosen loss function.

7.1.1 The optimisation problem

Using market data, namely prices of a set of at-the-money interest rate derivatives, denoted by IRD_i^M, the risk-neutral parameters $\Pi = \{a_x, a_y, \sigma_x, \sigma_y, \rho_{xy}\}$ can be calibrated by fitting the model prices to the market prices. The choice of the objective function is part of the modelling design. The option-calibration literature shows that loss functions materially affect valuation and parameter interpretation, so the objective is not a neutral numerical detail [21]. To avoid a bias towards high-priced instruments with long maturities, which would dominate a simple sum of squared errors, a relative error metric is chosen. The Root Mean Squared Relative Error (RMSRE) is used here. The optimisation is constrained to a financially admissible parameter space: mean-reversion speeds and volatilities are strictly positive, and the correlation is bounded within $[-1, 1]$.

Hence, the calibration becomes a constrained minimisation problem, which can be solved by a variety of numerical optimisation algorithms. These can be classified into:

- **Gradient-free methods**, which do not require derivative information. Examples include stochastic methods like Adaptive Simulated Annealing (ASA) or local, model-based approaches like BOBYQA or Nelder-Mead.
- **Gradient-based methods**, which use derivative information to converge more efficiently for smooth objective functions, like the Levenberg-Marquardt algorithm.

For a set of ν different instruments, the optimisation problem is of the following form:

$$\begin{aligned} \arg\min_{\Pi=\{a_x,a_y,\sigma_x,\sigma_y,\rho_{xy}\}} & \sqrt{\frac{1}{\nu}\sum_{i=1}^{\nu}\left(\frac{IRD_i^M - IRD_i(\Omega_i,\Pi)}{IRD_i^M}\right)^2} \\ \text{s.t.} \quad a_x &> 0, \\ a_y &> 0, \\ \sigma_x &> 0, \\ \sigma_y &> 0, \\ -1 \leq \rho_{xy} &\leq 1. \end{aligned}$$

Here, IRD_i^M represents the observed market price of the i-th instrument. The model price $IRD_i(\Omega_i, \Pi)$ is a function of the risk-neutral parameters Π and a set of instrument-specific characteristics Ω_i. This set contains all necessary information to price the instrument, such as its maturity, strike rate K_i, notional N_i, and tenor structure $\mathcal{T}$. This framework is applied below to the calibration of European ATM caps and ATM swaptions.

The optimisation problem is only the starting point. The following diagnostic analysis treats risk-neutral calibration as a quality problem after fit. With curve construction, quote conversion, discounting convention, calibration instruments, and objective function fixed, it studies the local geometry induced by these choices through the Jacobian of the G2++ pricing map. This is where fit, leverage, influence, effective degrees of freedom, boundary activity, and parameter uncertainty become scientifically relevant.

Parts of the diagnostic framework and of the cap-based empirical analysis used in the next section build on joint work with Peter Ruckdeschel on advanced calibration diagnostics and are also available in the joint preprint [76]. Here this material is reorganised, expanded, and embedded into the broader calibration-and-validation chain instead of being presented as a stand-alone paper.

7.2 Diagnostic Tools for Advanced Calibration Analysis

The following diagnostics assess the fitted parameter vector beyond the residual norm.

They serve three tasks. First, they translate the RMSRE calibration criterion into a fixed-design weighted nonlinear least-squares problem. This translation matters because the loss function itself affects the object being calibrated, a point emphasised in the option-calibration literature [21]. Second, they turn the local Jacobian geometry of the G2++ pricing map into interpretable objects: leverage, effective degrees of freedom, influence, conditioning, and local covariance approximations. In this reading the RMSRE weights define the local price space in which distances, projections, and uncertainty are measured. Third, they study these objects

on a unified Euro ATM cap sample from 2016 to 2025. The historical cap panel provides the empirical base, while the Hessian and boundary-leverage arguments add the additional technical material needed for the monograph.

This section takes the earlier market-data choices as fixed and studies the local calibration map induced by these choices. The answer matters because the calibrated risk-neutral history later enters the calibration under $\mathbb{P}$, actuarial scenario generation, and ex-post validation.

At this point the technical material connects back to the practical application chapters. In a PIA- or PRIIPs-related application setting, and likewise in an openIRM-style application chain, the calibration is used by application-level projections, product-level assessments, validation reports, and governance decisions. A diagnostic warning therefore has a practical meaning: it marks a day or a market segment where the calibrated $\mathbb{Q}$ input requires additional scrutiny before being used as an ordinary, well-identified observation. This practical reading remains in view throughout the section.

The following sections proceed in the same order in which the diagnostic objects are later used. First, they explain why the calibration can be read as a nonlinear regression problem, why asymptotic local geometry remains useful even in this fixed-design financial setting, and why constrained-parameter uncertainty requires transformation-based reasoning. Only then do we turn to the empirical diagnostics themselves.

These sections use statistical terminology only to describe the local calibration problem. The calibrated parameter vector minimises the weighted least-squares objective and can therefore be read as an M-estimator. At an interior regular solution, the first-order condition gives the corresponding Z-estimator view. The term score refers here to this local gradient condition, not to a likelihood score. Influence and covariance are used as first-order sensitivity diagnostics for the realised weighted design. Appendix E gives the supporting background.

7.2.1 Key Concepts: Nonlinear Regression

Let

$$\Pi = (a_x, a_y, \sigma_x, \sigma_y, \rho_{xy})^\top$$

denote the vector of risk-neutral G2++ parameters and let

$$g_k(\Pi) = IRD_k(\Omega_k, \Pi), \qquad k = 1, \ldots, m,$$

be the model price of the k-th calibration instrument. The observed market price is denoted by y_k^{Mkt}. The calibration estimator is then a nonlinear regression estimator with fixed design,

$$\hat{\Pi} = \arg\min_{\Pi \in \mathcal{D}} \sum_{k=1}^{m} \ell_k\big(y_k^{\mathsf{Mkt}}, g_k(\Pi)\big),$$

where $\mathcal{D}$ is the constrained parameter domain and ℓ_k is the contribution of the k-th instrument to the loss function. The maturities, strikes, tenors, notionals, curve convention, and quote conversion are treated as the realised empirical design. The statistical question is conditional on this design: given these instruments and these market prices, how stable is the fitted parameter vector?

This is the standard viewpoint of nonlinear least-squares regression; see, for example, [85]. The G2++ pricing function is not linear in Π. Hence the global model surface in the observation space $\mathbb{R}^m$ is curved. Nevertheless, a local linear approximation around the calibrated solution is

available through the Jacobian

$$J(\hat{\Pi}) = \left(\frac{\partial g_k(\hat{\Pi})}{\partial \Pi_j} \right)_{k=1,\dots,m;\ j=1,\dots,5}.$$

This Jacobian plays the role that the design matrix plays in ordinary linear regression. Its column space describes the local tangent space of the model price surface. All local diagnostics below are therefore diagnostics of this tangent space, not of the full nonlinear model globally.

The diagnostics below use this local geometric view. Let

$$\mathcal{M} = \{g(\Pi) : \Pi \in \mathcal{D}\} \subset \mathbb{R}^m$$

be the model price surface in price space. Calibration does not search for a parameter in isolation; it searches for the point on $\mathcal{M}$ that lies closest to the realised market-price vector under the metric induced by the WLS weights. Around $\hat{\Pi}$, this surface is approximated by its tangent plane,

$$g(\Pi) \approx g(\hat{\Pi}) + J(\hat{\Pi})(\Pi - \hat{\Pi}).$$

This first-order expansion is where regression geometry enters the calibration problem. The leverage analysis below makes this concrete. A maturity has high leverage when its weighted Jacobian row contributes a direction that the remaining rows cannot easily reconstruct. A parameter direction is weakly identified when the corresponding tangent direction is nearly redundant. These quantities are different interpretations of the same local linearisation; compare [85, 53].

This is the local geometric view used in this chapter. The RMSRE weights define the scale on which price perturbations are compared. The Jacobian maps parameter perturbations into this weighted price space. Consequently, the matrix $J^\top W J$ describes how strongly small parameter changes move model prices under the chosen calibration weights. It contains local curvature, identification, and uncertainty information at the same time. If two columns of the weighted Jacobian point almost in the same direction, the corresponding parameter directions are hard to separate. If one instrument row is unusually aligned with the fitted tangent space, that instrument can have high leverage. The diagnostics below are different ways of reading this same local geometry.

This local nature is a strength and a limitation. It is a strength because it reveals precisely which directions in price space the model can reproduce by small parameter changes around $\hat{\Pi}$. It is a limitation because it does not prove that the global objective is convex or that a different local optimum cannot exist. The diagnostic reading must therefore always be combined with optimisation checks, alternative starting values, and later ex-post validation.

Other diagnostic routes are possible. Bootstrap resampling, leave-one-out refits, robust refits, out-of-sample pricing checks, and the later ex-post validation layer are all useful complements. They answer different questions, however. Refits test empirical stability under changes in the data, out-of-sample checks test transfer, and ex-post validation tests simulated target objects. The Jacobian-based analysis used here is the direct local reading of the calibration map.

7.2.2 Key Concepts: Asymptotic Statistics

The second conceptual layer is asymptotic statistics. In a regular nonlinear regression problem, the estimator behaves locally like a linear estimator after suitable normalisation. Formally, under standard regularity conditions, an estimator defined by a smooth criterion or by smooth estimating equations admits an asymptotic linear representation and converges to a normal limit;

see [61, 57, 81, 89]. Here, this asymptotic language is used conditionally on the observed daily instrument set. The cap maturities are fixed by the calibration design and are not interpreted as randomly sampled observations.

A more precise statistical reading is useful because it clarifies why the same matrix $J^\top W J$ reappears in leverage, covariance, and influence diagnostics. Conditional on the realised instrument design, define the daily empirical criterion

$$Q_m(\Pi) = \frac{1}{m}\sum_{k=1}^{m} \rho_k(\Pi), \qquad \rho_k(\Pi) = w_k\big(y_k^{\mathsf{Mkt}} - g_k(\Pi)\big)^2.$$

The calibration estimator is therefore an M-estimator: it minimises an empirical objective. At the same time, differentiating Q_m yields the estimating equation

$$S_m(\Pi) = \frac{1}{m}\sum_{k=1}^{m} \psi_k(\Pi) = 0,$$

with

$$\psi_k(\Pi) = -2w_k J_k(\Pi)^\top \big(y_k^{\mathsf{Mkt}} - g_k(\Pi)\big),$$

where $J_k(\Pi)$ denotes the k-th Jacobian row. In this second reading, the same calibration is a Z-estimator because it is characterized as the root of a system of estimating equations. The objective function and the estimating equation therefore give two equivalent local descriptions of the WLS calibration problem.

The underlying mechanism is the same as in the local theory of smooth M- and Z-estimators. Write

$$\varepsilon(\Pi_0) = y^{\mathsf{Mkt}} - g(\Pi_0)$$

for the vector of pricing errors at the reference point. Then

$$S_m(\Pi_0) = -\frac{2}{m} J(\Pi_0)^\top W \varepsilon(\Pi_0).$$

A first-order Taylor expansion of S_m around Π_0 gives, up to the common factor $2/m$,

$$0 \approx -J(\Pi_0)^\top W \varepsilon(\Pi_0) + J(\Pi_0)^\top W J(\Pi_0)(\hat{\Pi} - \Pi_0),$$

and therefore motivates the approximation

$$\hat{\Pi} - \Pi_0 \approx \big(J^\top W J\big)^+ J^\top W \varepsilon,$$

where $J = J(\Pi_0)$, $\varepsilon = \varepsilon(\Pi_0)$, W is the realised weight matrix, and $(\cdot)^+$ denotes the Moore–Penrose pseudo-inverse. Hence the same matrix $J^\top W J$ appears simultaneously as the linearised curvature of the objective, as the local sensitivity matrix of the estimating equations, and as the central matrix in the covariance approximation. This common matrix explains why later subsections move between leverage, influence, Fisher-information-type arguments, and local parameter uncertainty. These are different diagnostics of the same local score linearisation.

Robust statistics expresses the same idea in the language of statistical functionals; see [61, 57, 81]. One writes $\hat{\Pi}_m = T(G_m)$, where G_m is the empirical measure of the realised quote objects and T is the estimator functional. Since the daily calibration minimises the WLS criterion, this functional is of M-estimator type. Since the same solution satisfies $S_m(\hat{\Pi}_m) = 0$, it also has a Z-estimator reading. The influence function measures how the estimator $T(F)$ changes when F is perturbed slightly towards a quote object z:

$$\mathrm{IF}(z; T, F) := \lim_{\varepsilon \downarrow 0} \frac{T\big((1-\varepsilon)F + \varepsilon\delta_z\big) - T(F)}{\varepsilon}.$$

This is the directional derivative of T along the perturbation path $(1-\varepsilon)F + \varepsilon\delta_z$, equivalently along the signed direction $\delta_z - F$. If T is Hadamard differentiable at F in the relevant directions and the usual regularity conditions for the corresponding Z-estimator hold, one obtains the asymptotic linear representation

$$\sqrt{m}(\hat{\Pi}_m - \Pi) = \frac{1}{\sqrt{m}} \sum_{k=1}^{m} \mathrm{IF}(z_k; T, F) + o_p(1),$$

see [81, 89]. This is the formal background of the finite-sample influence vectors used later in the chapter. Appendix E records the corresponding fixed-design estimator status, including the Gâteaux- and Hadamard-differentiability terminology, for the realised daily WLS problem.

A likelihood interpretation requires additional assumptions. For a genuine maximum-likelihood estimator, the estimating equation is the likelihood score and the information matrix is the Fisher information matrix. For general M- and Z-estimators, the covariance theory is usually expressed through the sensitivity and variability of the estimating equations, leading to sandwich-type covariance formulas; see [61, 81, 89]. In the present local WLS setting, conditional on the realised weights and under the homoscedastic relative-error approximation, the sensitivity part reduces to the Jacobian curvature matrix $J^\top W J$. Since the calibration objective does not specify a full joint distribution of quote errors, bid–ask effects, data revisions, or dependence across maturities, the analysis refers to $J^\top W J$ as a Fisher-information-type matrix rather than as a literal Fisher information matrix.

The diagnostic interpretation is subject to two qualifications. First, the calibration sample on a single day is not a large independent sample in the classical sense. It is a portfolio of market quotes across maturities. Hence the asymptotic language is used as a local approximation and as a diagnostic geometry, not as an unconditional statement about repeated independent sampling. Second, the parameters are constrained. Volatilities and mean-reversion speeds are positive, and the correlation is bounded by $[-1, 1]$. Near active constraints, the usual unconstrained normal approximation has to be interpreted with care.

After these qualifications, the relevant local object is the curvature matrix

$$J(\hat{\Pi})^\top W J(\hat{\Pi}).$$

It measures how strongly the observed calibration instruments identify infinitesimal parameter directions. If the matrix is well conditioned and full rank, small changes in market prices translate into relatively stable changes in parameters. If it is nearly singular or loses rank, some parameter directions are weakly identified or not locally identified at all.

This subsection explains why a low price error can coexist with a diagnostically weak calibration. Beyond the distance between the fitted price vector and the market price vector, the local analysis studies the curvature available around the fitted point and its distribution across maturities and parameter directions.

7.2.3 Key Concepts: Functional Delta Method

The third conceptual layer concerns parameter constraints. The local covariance approximation first describes uncertainty on the original parameter scale. For a scalar coordinate θ of the parameter vector, let $\widehat{\Sigma}_{\theta\theta}$ denote the corresponding diagonal entry of the local covariance approximation. The direct symmetric interval generated by the local normal approximation is

$$I_\theta^{\mathrm{direct}} = \left[\hat{\theta} - z_{1-\alpha/2}\sqrt{\widehat{\Sigma}_{\theta\theta}},\ \hat{\theta} + z_{1-\alpha/2}\sqrt{\widehat{\Sigma}_{\theta\theta}}\right],$$

where $z_{1-\alpha/2}$ is the standard-normal quantile. This interval is constructed directly on the original parameter scale. For the G2++ parameters, this scale is constrained: mean-reversion speeds and volatilities must be positive, and the correlation must lie in $[-1, 1]$. A direct interval for ρ_{xy} may therefore extend below -1 or above 1, and a direct interval for a volatility parameter may include negative values. Such intervals are incompatible with the model domain; they can become misleading in later simulation.

The functional Delta Method provides a standard construction for boundary-respecting intervals. If a scalar estimator satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, V(\theta)),$$

and if η is differentiable at θ, then

$$\sqrt{n}\big(\eta(\hat{\theta}_n) - \eta(\theta)\big) \xrightarrow{d} \mathcal{N}\big(0, [\eta'(\theta)]^2 V(\theta)\big),$$

see [89, Ch. 3]. The practical idea is to construct intervals in a transformed space where the parameter is unconstrained and the local variance is better behaved, and then map the interval back to the original parameter scale.

The conceptual point is broader than interval construction. Near a boundary, the original parameter scale is often the least informative scale for local uncertainty. The same absolute change in ρ_{xy} has a very different meaning when the correlation is near 0 than when it is near -1. The same issue appears for positive parameters such as mean-reversion speeds or volatilities. Additive intervals on the original scale can include inadmissible negative values. Constructing intervals on a transformed scale avoids impossible endpoints and gives uncertainty bands that respect the constrained parameter domain.

For the G2++ calibration we use the following componentwise transformations:

$$\eta(a_x) = \log(a_x), \quad \eta(a_y) = \log(a_y), \quad \eta(\sigma_x) = \log(\sigma_x), \quad \eta(\sigma_y) = \log(\sigma_y),$$

and

$$\eta(\rho_{xy}) = \operatorname{arctanh}(\rho_{xy}).$$

The logarithm maps positive parameters to the real line. The Fisher z-transformation maps the correlation interval $(-1, 1)$ to the real line. In the multivariate case, the covariance matrix in transformed coordinates is approximated by

$$\operatorname{Cov}\big(\eta(\hat{\Pi})\big) \approx D\eta(\hat{\Pi})\, \widehat{\operatorname{Cov}}(\hat{\Pi})\, D\eta(\hat{\Pi})^{\top},$$

where $D\eta(\hat{\Pi})$ is the diagonal Jacobian of the transformation. Intervals are constructed in the transformed coordinates and then mapped back by η^{-1}. The resulting intervals respect the natural parameter boundaries by construction.

This is a post-calibration diagnostic. It does not change the point estimate $\hat{\Pi}$ and it does not solve an optimisation problem. Its role is to make local parameter uncertainty compatible with the constrained geometry of the model.

The chosen transformation is therefore part of the uncertainty display, not a second calibration step. For positive parameters, alternatives such as square-root transformations would be possible; for the correlation coordinate, a logit or probit transform of $(\rho_{xy}+1)/2$ would play a similar boundary-respecting role. The logarithm and the Fisher z-transformation are used because they are standard, simple, and easy to interpret. No invariance claim is made with respect to all possible transformations.

7.2.4 RMSRE as a Weighted Least-Squares Problem

The optimisation criterion in the risk-neutral calibration is the root mean squared relative error. In the notation of this section,

$$\mathsf{RMSRE}(\Pi) = \sqrt{\frac{1}{m}\sum_{k=1}^{m}\left(\frac{y_k^{\mathsf{Mkt}} - g_k(\Pi)}{y_k^{\mathsf{Mkt}}}\right)^2}.$$

Because the square root is strictly increasing and the factor $1/m$ is independent of Π, minimising the RMSRE is equivalent to minimising

$$\begin{aligned}\hat{\Pi} &= \arg\min_{\Pi}\sum_{k=1}^{m}\left(\frac{y_k^{\mathsf{Mkt}} - g_k(\Pi)}{y_k^{\mathsf{Mkt}}}\right)^2 \\ &= \arg\min_{\Pi}\sum_{k=1}^{m}\frac{1}{(y_k^{\mathsf{Mkt}})^2}\big(y_k^{\mathsf{Mkt}} - g_k(\Pi)\big)^2.\end{aligned} \tag{7.1}$$

Thus the RMSRE calibration is a weighted least-squares problem with weights

$$w_k = \frac{1}{(y_k^{\mathsf{Mkt}})^2}. \tag{7.2}$$

The realised weights are treated as fixed in the diagnostic analysis. In the implementation of the cap study, the inverse-price weights are formed with a numerical guard

$$y_k^{\mathsf{safe}} = \max(y_k^{\mathsf{Mkt}}, 10^{-5}).$$

This floor is inactive in the cap portfolio: after vendor-to-model scaling, the smallest selected cap price is $2.6 \cdot 10^{-3}$. It is therefore a numerical safeguard against near-zero prices and does not change the weighting scheme in the analysed cap panel.

The relative-error objective has a statistical interpretation. Locally, it can be read as a heteroskedastic error model in price units. If absolute quote errors tend to scale with the level of the instrument price, then a model of the form

$$y_k^{\mathsf{Mkt}} = g_k(\Pi)(1 + \varepsilon_k),$$

or, equivalently, $\mathsf{Var}(u_k) \approx \sigma^2 (y_k^{\mathsf{Mkt}})^2$ for the absolute residual $u_k = y_k^{\mathsf{Mkt}} - g_k(\Pi)$, motivates a relative-error weighting. The WLS formulation then treats the relative residuals as the objects that should be comparable across maturities. This is the local error picture behind RMSRE: the errors are heteroskedastic in original price units and approximately homoskedastic in relative price units.

This interpretation is conditional on the chosen weights and observed market prices. The observed market prices are used to set the weights, and no full probability law for the quote vector is specified. The RMSRE criterion is therefore not a complete likelihood and the resulting estimator is not automatically an MLE. It is a conditional WLS criterion with a clear economic motivation: maturities with very different price levels are compared in relative price units. It also changes the functional being targeted by the loss function. As discussed by [52], losses based on relative errors elicit different distributional functionals from ordinary squared-error losses.

A different weighting rule would define a different calibration target. Weights based on traded volume, liquidity, or repeated quote information would be meaningful if that information is observed and intended to enter the calibration objective. They would not be neutral replacements for RMSRE. The RMSE-versus-RMSRE sensitivity check in Section 7.3.2 keeps this point explicit for the OIS 6M cap panel.

Proposition 7.1 (RMSRE as a weighted least-squares local design)**.** *For fixed positive market prices and fixed weights* $w_k = (y_k^{\mathsf{Mkt}})^{-2}$*, minimising RMSRE is equivalent, up to the constant factor* $1/m$ *and the monotone square-root transformation, to a weighted nonlinear least-squares problem. At a locally regular solution for which* $W^{1/2}J(\hat{\Pi})$ *has full column rank, set*

$$X = W^{1/2}J(\hat{\Pi}).$$

The matrix

$$H_w = X(X^\top X)^{-1}X^\top = W^{1/2}J(\hat{\Pi})\big(J(\hat{\Pi})^\top W J(\hat{\Pi})\big)^{-1}J(\hat{\Pi})^\top W^{1/2}$$

is the orthogonal projector onto the column space of the weighted Jacobian. It describes which first-order weighted price movements can be generated by small parameter changes. Consequently, leverage, EDoF, influence, and local covariance or Fisher-information-type quantities are functions of the same local weighted design.

7.2.5 Tangent-Space Geometry and the Weighted Hat Matrix

Let

$$y = (y_1^{\mathsf{Mkt}}, \dots, y_m^{\mathsf{Mkt}})^\top, \qquad g(\Pi) = (g_1(\Pi), \dots, g_m(\Pi))^\top,$$

and let $W = \mathsf{diag}(w_1, \dots, w_m)$. Around the calibrated parameter vector $\hat{\Pi}$, the pricing map is approximated by

$$g(\Pi) \approx g(\hat{\Pi}) + J(\hat{\Pi})(\Pi - \hat{\Pi}). \tag{7.3}$$

The columns of $J(\hat{\Pi})$ span the tangent space of the model price surface at $g(\hat{\Pi})$. In weighted coordinates, the relevant design matrix is

$$X = W^{1/2}J(\hat{\Pi}).$$

This is the Jacobian written in the coordinates in which the RMSRE objective measures price errors. The orthogonal projector onto the column space of X is

$$H_w = W^{1/2}J(\hat{\Pi})\big(J(\hat{\Pi})^\top W J(\hat{\Pi})\big)^{+}J(\hat{\Pi})^\top W^{1/2}. \tag{7.4}$$

This is the symmetric weighted hat matrix. Its diagonal entries

$$h_{kk} = (H_w)_{kk}$$

are the leverage scores. These diagonal values are algebraically identical to the diagonal values of the corresponding oblique projector in the original price coordinates, but the symmetric formulation makes clear that the projection is orthogonal after weighting.

In the same notation, $J^\top W J = X^\top X$ is the pullback of the weighted price metric to parameter space. It measures the lengths and angles of model-implied price movements generated by small parameter changes. Small eigenvalues correspond to parameter directions that barely move prices in the weighted calibration space. Large diagonal entries of H_w correspond to observations that are strongly aligned with that local tangent space. This explains why the same matrix family appears in leverage, effective degrees of freedom, conditioning, influence, and covariance diagnostics.

The trace of H_w is the rank of the weighted Jacobian,

$$\mathsf{tr}(H_w) = \mathsf{rank}(W^{1/2}J).$$

If all five G2++ parameter directions are locally identified, the trace equals five. If a direction is lost because the Jacobian is singular or because an active constraint removes a local degree

of freedom, the trace falls to a lower integer value under the Moore–Penrose construction. This gives the effective degrees of freedom diagnostic used below.

The hat-matrix construction is the algebraic bridge to the classical regression-diagnostics literature; see [60, 8, 25]. For the present calibration problem, very high individual leverage means that one market quote is no longer small relative to the total local information represented by $W^{1/2}J$. The practical interpretation is taken up again in the diagnostic-tool summary below.

7.2.6 Analytical Derivatives for Caps

The diagnostic framework requires a reliable Jacobian. For this reason, the empirical diagnostic study is built on caps instead of swaptions. In the G2++ model, cap prices admit an analytical caplet decomposition, while swaption prices require a semi-analytical formula with numerical integration. Differentiating the cap formula analytically avoids finite-difference noise in $J^\top W J$, which is precisely the matrix that later has to be inverted or pseudo-inverted.

For a caplet with rate-setting date t_{i-1}, payment date t_i, accrual period δ_i, strike K, and integrated G2++ variance Σ_i^2, write

$$L_i = \log\left(\frac{P(0,t_{i-1})}{(1+K\delta_i)P(0,t_i)}\right),$$

and

$$h_{+,i} = \frac{L_i}{\Sigma_i} + \frac{\Sigma_i}{2}, \qquad h_{-,i} = \frac{L_i}{\Sigma_i} - \frac{\Sigma_i}{2}.$$

Let $p \in \Pi = (a_x, a_y, \sigma_x, \sigma_y, \rho_{xy})$ denote one risk-neutral G2++ parameter. Under the convention that the initial curve and the strike are fixed inputs, the G2++ parameter dependence of the caplet price enters through Σ_i^2. The derivative is organized as follows:

Chain Rule Along a Path.

$$\mathsf{Caplet}_i \;\to\; \Phi \;\to\; h_{\pm,i} \;\to\; \Sigma_i \;\to\; \Sigma_i^2 \;\to\; p,$$

or, in compressed chain-rule notation,

$$\frac{\partial\,\mathsf{Caplet}_i}{\partial p} = \frac{\partial\,\mathsf{Caplet}_i}{\partial \Phi}\frac{\partial \Phi}{\partial h}\frac{\partial h}{\partial \Sigma_i}\frac{\partial \Sigma_i}{\partial \Sigma_i^2}\frac{\partial \Sigma_i^2}{\partial p}. \tag{7.5}$$

The last factor is the only parameter-specific part of the chain.

Step 1: The Price-to-Variance Sensitivity. Write

$$A_i = P(0,t_{i-1}), \qquad B_i = (1+K\delta_i)P(0,t_i), \qquad L_i = \log\left(\frac{A_i}{B_i}\right).$$

The curve and strike inputs are fixed in this derivative, so A_i, B_i, and L_i do not depend on the G2++ parameters. Set $x_i = \Sigma_i^2$. Then

$$h_{+,i} = L_i x_i^{-1/2} + \frac{1}{2}x_i^{1/2}, \qquad h_{-,i} = L_i x_i^{-1/2} - \frac{1}{2}x_i^{1/2}.$$

Hence

$$\frac{\partial h_{+,i}}{\partial x_i} = \frac{1}{2\Sigma_i}\left(-\frac{L_i}{\Sigma_i^2} + \frac{1}{2}\right), \qquad \frac{\partial h_{-,i}}{\partial x_i} = \frac{1}{2\Sigma_i}\left(-\frac{L_i}{\Sigma_i^2} - \frac{1}{2}\right).$$

Differentiating the caplet price with respect to $x_i = \Sigma_i^2$ gives

$$\begin{aligned}\frac{\partial \,\mathsf{Caplet}_i}{\partial \Sigma_i^2} &= N\left[A_i\varphi(h_{+,i})\frac{\partial h_{+,i}}{\partial \Sigma_i^2} - B_i\varphi(h_{-,i})\frac{\partial h_{-,i}}{\partial \Sigma_i^2}\right] \\ &= \frac{N}{2\Sigma_i}\left[A_i\varphi(h_{+,i})\left(-\frac{L_i}{\Sigma_i^2}+\frac{1}{2}\right) - B_i\varphi(h_{-,i})\left(-\frac{L_i}{\Sigma_i^2}-\frac{1}{2}\right)\right].\end{aligned}$$

The apparent dependence on L_i cancels. Indeed,

$$h_{+,i} - h_{-,i} = \Sigma_i, \qquad h_{+,i} + h_{-,i} = \frac{2L_i}{\Sigma_i},$$

so that

$$h_{+,i}^2 - h_{-,i}^2 = (h_{+,i} - h_{-,i})(h_{+,i} + h_{-,i}) = 2L_i.$$

Since $\varphi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, this implies

$$\frac{\varphi(h_{+,i})}{\varphi(h_{-,i})} = \exp\left(-\frac{1}{2}(h_{+,i}^2 - h_{-,i}^2)\right) = \exp(-L_i) = \frac{B_i}{A_i},$$

and therefore

$$A_i\varphi(h_{+,i}) = B_i\varphi(h_{-,i}).$$

Using this identity in the derivative expression above yields

$$\frac{\partial \,\mathsf{Caplet}_i}{\partial \Sigma_i^2} = \frac{NA_i\varphi(h_{+,i})}{2\Sigma_i} = \frac{NB_i\varphi(h_{-,i})}{2\Sigma_i}.$$

Substituting back $A_i = P(0,t_{i-1})$ and $B_i = (1+K\delta_i)P(0,t_i)$ gives

$$\mathcal{M}_i(\Pi) := \frac{\partial \,\mathsf{Caplet}_i}{\partial \Sigma_i^2} = \frac{NP(0,t_{i-1})\varphi(h_{+,i})}{2\Sigma_i} = \frac{N(1+K\delta_i)P(0,t_i)\varphi(h_{-,i})}{2\Sigma_i}. \tag{7.6}$$

Step 2: The Variance-to-Parameter Sensitivity. The previous step gives the derivative of the caplet price with respect to the integrated variance Σ_i^2. The dependence on a G2++ parameter $p \in \Pi$ is then obtained by the chain rule:

$$\frac{\partial \,\mathsf{Caplet}_i}{\partial p} = \frac{\partial \,\mathsf{Caplet}_i}{\partial \Sigma_i^2}\frac{\partial \Sigma_i^2}{\partial p} = \mathcal{M}_i(\Pi)\frac{\partial \Sigma_i^2}{\partial p}. \tag{7.7}$$

The factor $\mathcal{M}_i(\Pi)$ contains the market-data-specific price-to-variance sensitivity, while $\partial\Sigma_i^2/\partial p$ carries the parameter-specific variance sensitivity. The explicit formulas for $\partial\Sigma_i^2/\partial p$, $p \in \Pi$, are collected in Appendix A.

Step 3: The Factorized Jacobian. For a cap k consisting of caplets $i \in \mathcal{I}_k$, the Jacobian row is therefore

$$\frac{\partial \,\mathsf{Cap}_k}{\partial p} = \sum_{i\in\mathcal{I}_k}\mathcal{M}_i(\Pi)\frac{\partial \Sigma_i^2}{\partial p}. \tag{7.8}$$

This factorization is more than an implementation shortcut. It separates three pieces of information that have different meanings: the caplet-composition structure of a cap, the strike- and curve-dependent price-to-variance weights, and the parameter-specific variance sensitivities. If $\mathbf{c}_k$ denotes the caplet-composition vector of cap k, if

$$M_{\mathsf{diag}}(K_k;\Pi) = \mathsf{diag}\big(\mathcal{M}_i(K_k;\Pi)\big)_i$$

collects the caplet price-to-variance sensitivities for the strike K_k, and if

$$S_{\mathsf{deriv}}(\Pi) = \left(\frac{\partial \Sigma_i^2}{\partial p_j}\right)_{i,j}$$

collects the variance sensitivities for $p_j \in \Pi$, then the full row of the cap Jacobian can be assembled as

$$J_{k,\cdot} = \mathbf{c}_k^\top M_{\mathsf{diag}}(K_k; \Pi) S_{\mathsf{deriv}}(\Pi). \tag{7.9}$$

Equation (7.9) has a diagnostic and computational role. The matrix S_{deriv} describes how the G2++ variance surface reacts to parameter changes; M_{diag} translates those variance movements into price movements under the current curve and strike; and $\mathbf{c}_k$ records which caplets enter the quoted cap. Hence row differences in the Jacobian can be traced back to economically distinct sources instead of an undifferentiated numerical derivative.

Cap–floor Jacobian identity. The same G2++ geometry applies to floors with the same curve, strike convention, and payment schedule. At the caplet level, no-arbitrage gives

$$\mathsf{Caplet}_i(0) - \mathsf{Floorlet}_i(0) = N\left[P(0,t_{i-1}) - (1+K\delta_i)P(0,t_i)\right]. \tag{7.10}$$

Summing over the payment dates yields the cap–floor parity relation

$$\begin{aligned}\mathsf{Cap}(0,\mathcal{T},N,K,\Pi) - \mathsf{Floor}(0,\mathcal{T},N,K,\Pi) &= \mathsf{PayerSwap}(0,\mathcal{T},N,K)\\ &= N\big[P(0,t_m) - P(0,t_n)\big] - NK\sum_{i=m+1}^{n}\delta_i P(0,t_i).\end{aligned} \tag{7.11}$$

The right-hand side depends on the initial curve and the fixed strike and payment schedule, but not on Π. Differentiating (7.11) with respect to any $p \in \Pi$ therefore gives

$$\frac{\partial\,\mathsf{Cap}}{\partial p} = \frac{\partial\,\mathsf{Floor}}{\partial p}. \tag{7.12}$$

The same identity also follows directly from the analytical floorlet derivative. The G2++ floorlet price is

$$\mathsf{Floorlet}_i = N\left[(1+K\delta_i)P(0,t_i)\Phi(-h_{-,i}) - P(0,t_{i-1})\Phi(-h_{+,i})\right].$$

Using $\varphi(-x) = \varphi(x)$ and differentiating only the normal-distribution terms gives

$$\begin{aligned}\frac{\partial\,\mathsf{Floorlet}_i}{\partial p} &= N\left[-(1+K\delta_i)P(0,t_i)\varphi(h_{-,i})\frac{\partial h_{-,i}}{\partial p} + P(0,t_{i-1})\varphi(h_{+,i})\frac{\partial h_{+,i}}{\partial p}\right]\\ &= \frac{\partial\,\mathsf{Caplet}_i}{\partial p}.\end{aligned} \tag{7.13}$$

Thus $\mathcal{M}_i^{\mathsf{Floor}}(\Pi) = \mathcal{M}_i^{\mathsf{Cap}}(\Pi)$, and summing over matching payment schedules gives $J_{\mathsf{Floor}} = J_{\mathsf{Cap}}$.

For an ATM instrument, the strike is the par swap rate for the same payment schedule,

$$K_{\mathsf{ATM}} = \frac{P(0,t_m) - P(0,t_n)}{\sum_{i=m+1}^{n}\delta_i P(0,t_i)}.$$

Substituting this strike into the right-hand side of (7.11) gives

$$\mathsf{PayerSwap}(0,\mathcal{T},N,K_{\mathsf{ATM}}) = 0.$$

Hence, under the same curve and schedule conventions,

$$\mathsf{Cap}(0,\mathcal{T},N,K_{\mathsf{ATM}},\Pi) = \mathsf{Floor}(0,\mathcal{T},N,K_{\mathsf{ATM}},\Pi).$$

For ATM instruments with identical observation and maturity structure, the parity-implied prices therefore coincide as well. Since the WLS weights are functions of these prices, the weights coincide, and the weighted hat matrix and the diagnostics derived from it are unchanged.

7.2.7 Exact Hessian and Gauss–Newton Adequacy

The first-order Jacobian is sufficient for the main diagnostics, but the exact Hessian clarifies why the Gauss–Newton curvature approximation is empirically reliable in the cap calibration. Starting from (7.7), the second derivative of a caplet price with respect to two parameter coordinates p_u and p_v is obtained by the product rule:

$$\frac{\partial^2 \mathsf{Caplet}_i}{\partial p_u \partial p_v} = \mathcal{M}_i'(\Pi)\frac{\partial \Sigma_i^2}{\partial p_u}\frac{\partial \Sigma_i^2}{\partial p_v} + \mathcal{M}_i(\Pi)\frac{\partial^2 \Sigma_i^2}{\partial p_u \partial p_v}, \tag{7.14}$$

where

$$\mathcal{M}_i'(\Pi) = \frac{NP(0,t_{i-1})\varphi(h_{+,i})}{4\Sigma_i^3}\big(h_{+,i}h_{-,i}-1\big). \tag{7.15}$$

The compact form (7.15) is used because it makes the derivative of $\mathcal{M}_i$ readable without changing the analytical object. The cap Hessian is then obtained by summing (7.14) over the caplets of the cap. The remaining second derivatives of Σ_i^2 are listed in Appendix A.

The validity of Jacobian-based diagnostics depends on the accuracy of the local linearisation in the G2++ cap calibration problem. Cap prices are nonlinear functions of the model parameters: they involve exponential Ornstein–Uhlenbeck kernels, nonlinear variance aggregations, logarithmic terms, and cumulative normal distribution functions. Therefore, it cannot be taken for granted that a first-order approximation around the calibrated parameter captures the local geometry of the problem well enough for diagnostic purposes. For the diagnostics developed here, the relevant object is the local second-order geometry of the weighted least-squares objective at the calibrated point, rather than the global nonlinearity of the cap price function alone.

For the weighted least-squares objective

$$Q(\Pi) = \sum_{k=1}^{m} w_k\big(y_k^{\mathsf{Mkt}} - g_k(\Pi)\big)^2,$$

the exact objective Hessian at $\hat{\Pi}$ is

$$\nabla^2 Q(\hat{\Pi}) = 2J(\hat{\Pi})^\top W J(\hat{\Pi}) - 2\sum_{k=1}^{m} w_k r_k(\hat{\Pi})\nabla^2 g_k(\hat{\Pi}), \tag{7.16}$$

where

$$r_k(\hat{\Pi}) = y_k^{\mathsf{Mkt}} - g_k(\hat{\Pi}).$$

The first term is the Gauss–Newton curvature. The second term is the residual-weighted second-order correction. If the residuals are small at the calibrated point, this correction is expected to be small even though the individual price functions are nonlinear. The decomposition therefore separates two questions. The first question is whether the individual price functions $g_k(\Pi)$ are nonlinear. The second question is whether this nonlinearity materially changes the local curvature of the residual-based objective at the fitted point. The empirical answer is that the price function is nonlinear, but the objective Hessian is nevertheless almost indistinguishable from its Gauss–Newton part.

The cap price map is nonlinear

The exact cap price Hessians

$$\nabla^2 g_k(\hat{\Pi}_t)$$

are computed for every cap maturity k and every calibration day t in the sample. They are evaluated at the daily calibrated parameter vector $\hat{\Pi}_t$, using the corresponding day-specific curve

construction, ATM strike, and price target convention from the OIS 6M ATM-cap calibration. Figures 7.1 and 7.2 are therefore empirical aggregation plots based on realised daily calibration inputs. No separate illustrative parameter vector is introduced.

The first empirical finding is that the price map is nonlinear in a non-trivial way. This is visible in the maturity-wise distribution of the cap-price Hessian norms. The Frobenius norms of $\nabla^2 g_k(\hat{\Pi}_t)$ are clearly nonzero for all maturities and tend to increase towards the long end. On the median level, the cap-price Hessian norm rises from roughly 307 at 3Y to roughly 920 at 30Y. At the 90% quantile, the increase is even more pronounced, from approximately 1,064 at 3Y to approximately 45,672 at 30Y. Hence, the price function itself is not close to globally linear. The adequacy of the Gauss–Newton approximation is therefore treated as an empirical diagnostic below. The availability of analytical cap derivatives does not by itself imply that the second-order terms are negligible.

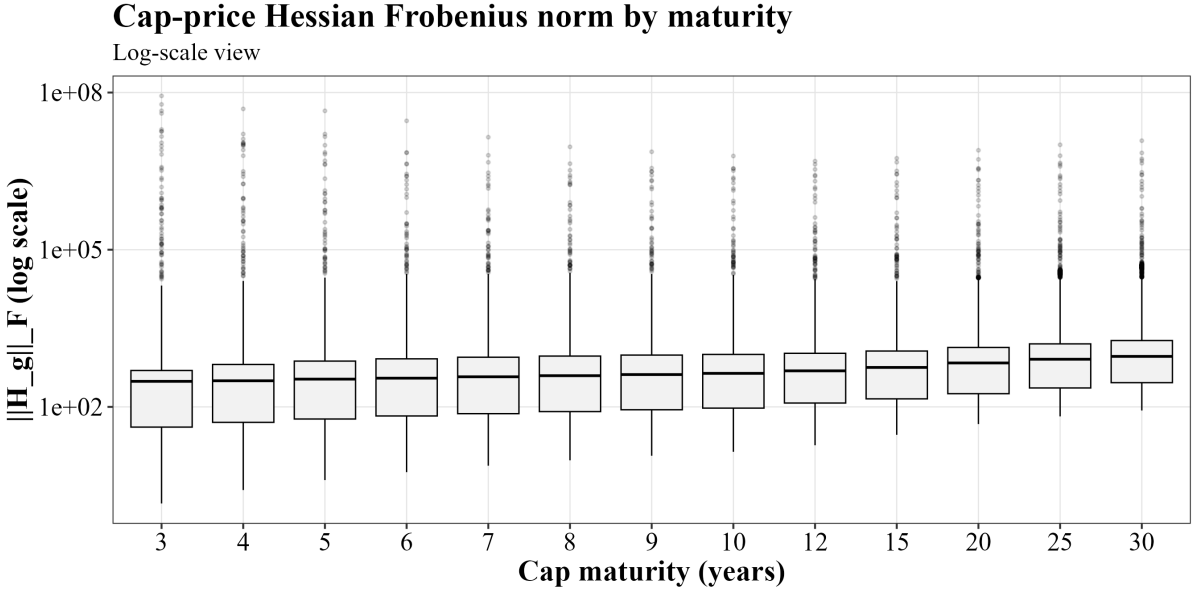


Figure 7.1: **Cap-price Hessian Frobenius norm by maturity (log-scale view).** The figure shows the empirical distribution of $\|\nabla^2 g_k(\hat{\Pi}_t)\|_F$ across all OIS 6M ATM-cap calibration days in the unified 2016–2025 sample for each cap maturity. The norms are clearly nonzero and tend to increase towards longer maturities, confirming that the cap price map is nonlinear.

Because the log-scale view is dominated by the upper tail, a zoomed linear-scale version is useful to visualise the typical range of curvature levels. This confirms the same message on the scale of the bulk of the data: the price Hessians are economically and numerically non-negligible, and the nonlinearity is systematically stronger at longer maturities.

The exact objective Hessian is almost Gauss–Newton

The central finding is that the strong nonlinearity of the individual price map does not propagate into an important deviation between the exact objective Hessian and its Gauss–Newton part at the calibrated point. This is most clearly seen in the relative norm comparison of the residual-weighted correction term to the Gauss–Newton term. Over the full sample, the Frobenius-norm ratio

$$\frac{\left\| -2 \sum_k w_k r_k(\hat{\Pi}) \nabla^2 g_k(\hat{\Pi}) \right\|_F}{\left\| 2 J(\hat{\Pi})^\top W J(\hat{\Pi}) \right\|_F}$$

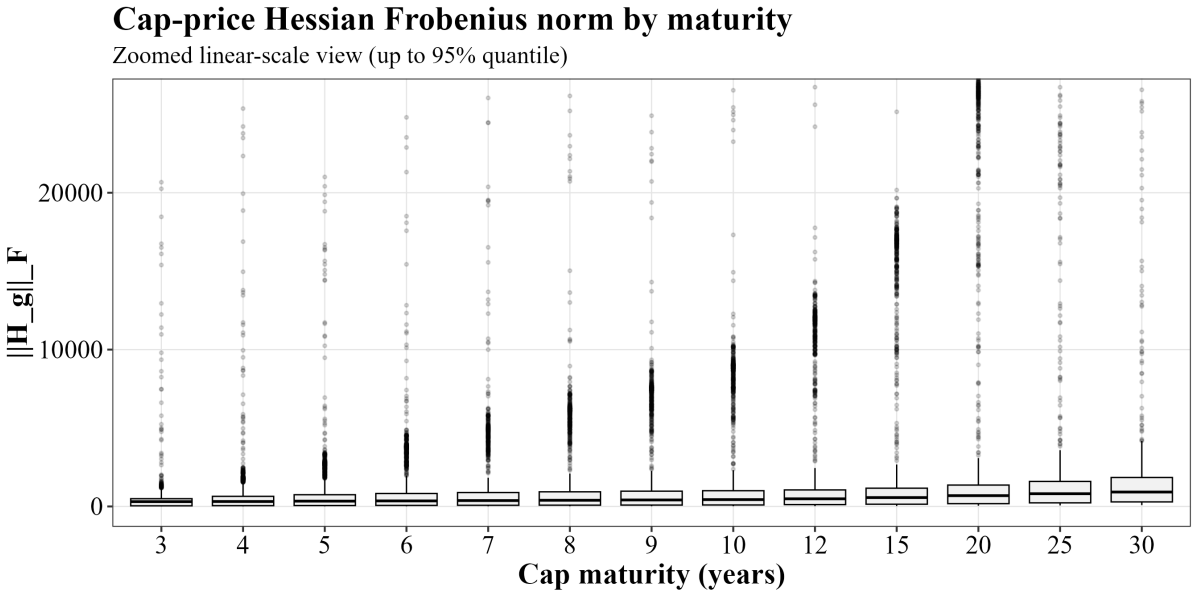


Figure 7.2: **Cap-price Hessian Frobenius norm by maturity (zoomed linear-scale view).** The figure shows the same empirical OIS 6M ATM-cap Hessian quantity as Figure 7.1, but restricted to the central range of the distribution. This makes the typical curvature levels visible without being dominated by rare extreme outliers.

has mean approximately 0.00190, median approximately 0.000123, and 90% quantile approximately 0.00627. Thus, on average, the second-order correction contributes only about 0.19% of the Gauss–Newton term in Frobenius norm; on the median day the contribution is only about 0.0123%; and even at the 90% quantile it is still only about 0.627%. The local objective geometry is therefore dominated by the Gauss–Newton term.

The same conclusion is obtained from the absolute levels. The median Frobenius norm of the Gauss–Newton term is approximately 237,364, whereas the median norm of the correction term is only approximately 29.6. The median norm of the exact Hessian is approximately 237,372, i.e. essentially identical to the Gauss–Newton level. The corresponding share statistics lead to the same conclusion: the Gauss–Newton term accounts for almost all of the exact Hessian norm, while the residual-weighted correction is norm-wise negligible.

Figure 7.3 gives a compact visual summary of the result. If the Gauss–Newton approximation were poor, the cloud would deviate visibly from the 45-degree line. Instead, the points lie almost exactly on that line across the entire sample and across several orders of magnitude. Thus, at the calibrated solutions, the exact Hessian and the Gauss–Newton term are practically the same object on the norm scale relevant for local diagnostics.

A complementary time-series view leads to the same interpretation. Across the entire sample, the exact objective Hessian and the Gauss–Newton term move almost on top of one another, whereas the correction term remains several orders of magnitude smaller for most days.

The direct relative time-series diagnostic confirms that the result is not an artefact of averaging. The correction-to-Gauss–Newton ratio is small in the aggregate and day by day. Although occasional spikes occur, they remain small on the scale of the main term. The result is also stable across subsamples: even in the years with the comparatively largest deviations, the 90% quantile of the ratio remains around one percent.

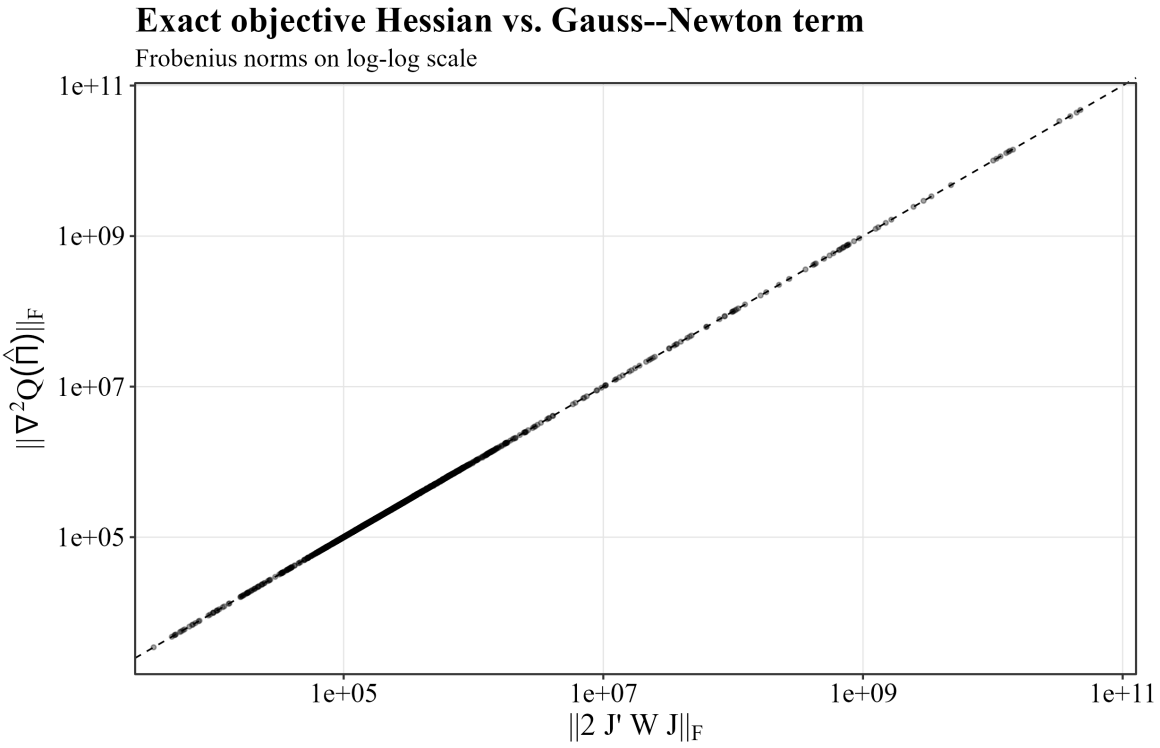


Figure 7.3: **Exact objective Hessian versus Gauss–Newton term.** The figure plots the Frobenius norm of the exact objective Hessian against the Frobenius norm of the Gauss–Newton term on log–log scale, together with the 45-degree line. The cloud lies almost exactly on the diagonal, showing that the two quantities are nearly indistinguishable on the scale relevant for local calibration geometry.

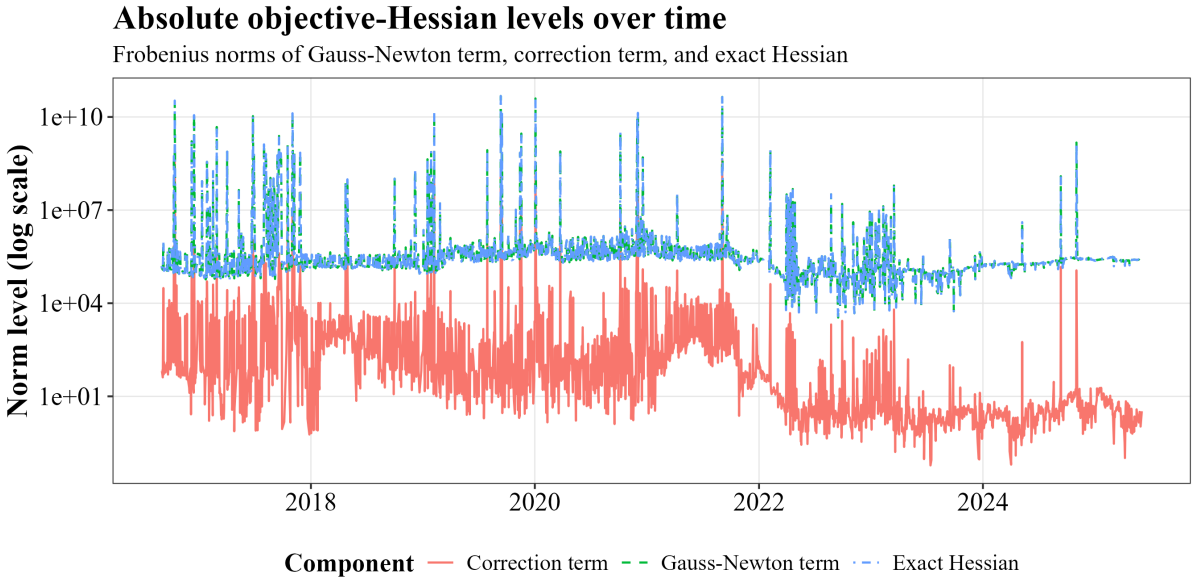


Figure 7.4: **Absolute objective-Hessian levels over time.** The figure shows the Frobenius norms of the Gauss–Newton term, the residual-weighted correction term, and the exact objective Hessian on log scale. The exact Hessian and the Gauss–Newton term are visually almost indistinguishable, while the correction term remains much smaller.

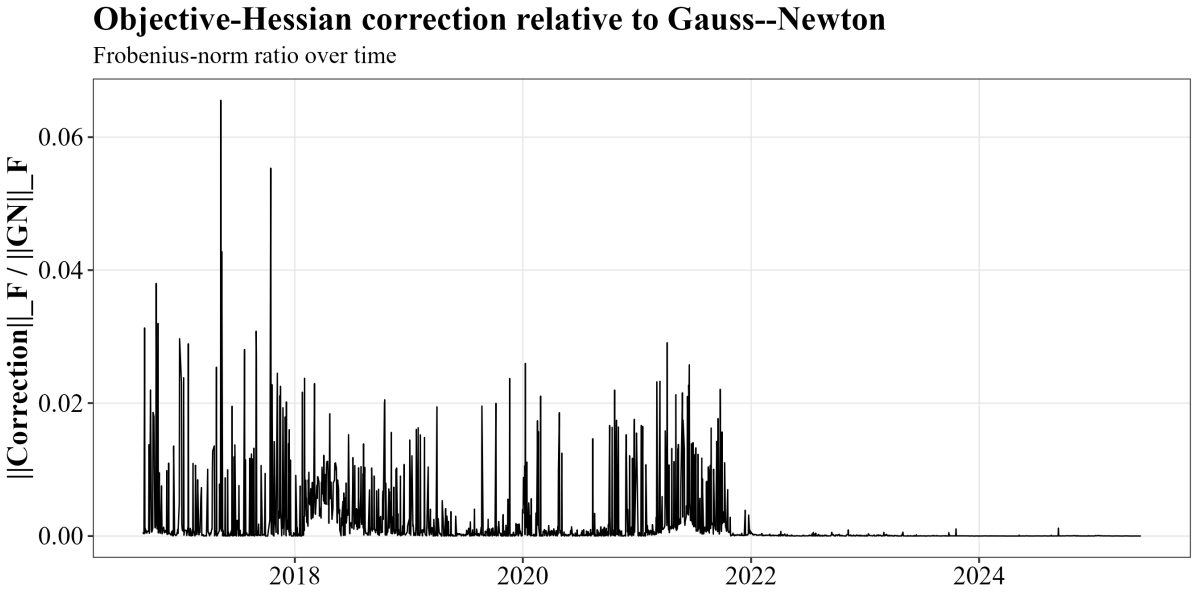


Figure 7.5: **Relative size of the residual-weighted correction term over time.** The figure shows the Frobenius-norm ratio of the exact second-order correction term to the Gauss–Newton term. The ratio stays very small throughout the sample, confirming that the Gauss–Newton geometry dominates the exact local curvature of the objective.

Structural interpretation and limitation

The empirical result has a simple structural explanation. As shown by the exact Hessian decomposition, the second-order correction to the Gauss–Newton term is multiplied by the fitted residuals. At a well-calibrated solution, these residuals are small. Consequently, even visible nonlinearity in individual cap prices can have only a minor effect on the local Hessian of the weighted least-squares objective.

This distinguishes two objects. The pricing map $g(\Pi)$ is nonlinear. The local objective geometry at the fitted parameter is governed by the Hessian of the residual-based WLS criterion. The empirical comparison above shows that, for the daily G2++ cap calibrations studied here, the residual-weighted nonlinear correction is very small. The Gauss–Newton term therefore provides an accurate description of the local objective curvature.

Taylor diagnostics along parameter shocks can still be useful, but ratios such as

$$\frac{|\text{quadratic term}|}{|\text{linear term}|}$$

are unstable when the linear term is close to zero. Large ratios may then reflect directional cancellation rather than a practically relevant failure of the first-order approximation. For this reason, the direct comparison between the exact objective Hessian and the Gauss–Newton term is the more relevant diagnostic in this setting.

The result is local. It does not imply global linearity, convexity, or the absence of second-order effects under arbitrary perturbations. It supports the use of $J(\hat{\Pi})^\top W J(\hat{\Pi})$ as the central local curvature object for the regression diagnostics developed below: Fisher-information-type geometry, leverage and hat-matrix diagnostics, local influence approximations, and covariance approximation.

There is also a related asymptotic interpretation. One-step correction results show that, in

smooth finite-dimensional estimation problems, a single Newton- or scoring-type update from a sufficiently accurate preliminary estimator can already have the same first-order asymptotic distribution as the fully iterated estimator; see [89, Sec. 5.7, Thm. 5.45]. The theorem provides asymptotic background for the role of local derivative objects in smooth estimation problems. In the present empirical setting, the exact objective Hessian and the Gauss–Newton term are almost identical at the calibrated solutions. Thus the Gauss–Newton information matrix is also the relevant local object for the diagnostic geometry.

7.2.8 Main Diagnostic Tools

The diagnostic tools follow directly from the WLS geometry and the analytical Jacobian.

The empirical block below is organised around four questions. First, does the fit remain good over time, or do diagnostic problems simply mirror poor pricing performance? Second, which maturities anchor the local tangent-space geometry, and which ones generate realised parameter movements once residuals are taken into account? Third, when does the local rank collapse, especially near active constraints? Fourth, how do these geometric changes translate into local parameter uncertainty? The diagnostic definitions therefore start with leverage, then add influence as the residual-weighted realised effect, then introduce EDoF as the rank diagnostic, and finally collect the covariance and uncertainty layer.

Leverage

The leverage score of instrument k is

$$h_{kk} = (H_w)_{kk}.$$

It measures the potential geometric impact of the k-th quote. Since H_w maps weighted market-price perturbations to weighted fitted-price perturbations, the diagonal entry h_{kk} measures the local self-impact of this quote. A high leverage score means that the fitted model quote at this maturity is strongly determined by its own market quote. If h_{kk} is close to one, the local fit has little freedom to adjust this quote using information from the remaining maturities.

The same point can be expressed in row-reconstruction terms. If the k-th weighted Jacobian row is difficult to reproduce from the remaining rows, the local fit cannot easily absorb a change in this quote through the other maturities. Generalized inverses provide the algebraic language for this statement. In particular, the Moore–Penrose inverse selects least-squares and minimum-norm solutions in the corresponding projection geometry [9, Ch. 2, Sec. 7; Ch. 3, Secs. 1–2]. In this setting, this applies to the row system $X_{-k}^{\top} a = x_k$. The diagnostic question is whether the weighted Jacobian row associated with a maturity can be reconstructed from the remaining weighted Jacobian rows.

The leverage interpretation follows the same logic as classical regression diagnostics [60, 8, 25]. High leverage is not automatically a data problem. It says that the observation sits in a geometrically important position. Whether this becomes a practical concern depends on the residual and on the stability of the surrounding calibration design.

Proposition 7.2 (Leverage-reconstruction identity)**.** *Let $X = W^{1/2} J$ have full column rank, let $x_i^{\top}$ denote its i-th row, and let μ_i be the squared minimum-norm coefficient norm needed to reconstruct x_i from the remaining rows, with $\mu_i = +\infty$ if exact reconstruction fails. Then the leverage score satisfies*

$$h_i = \frac{\mu_i}{1 + \mu_i},$$

with the usual interpretation $h_i = 1$ when $\mu_i = +\infty$.

Appendix D gives the proof and states the clean full-rank result. Rank-deficient daily designs are treated diagnostically through the Moore–Penrose construction used in (7.4), where applicable.

Influence

Leverage measures potential impact based on geometry alone. Influence combines geometry with the realised residual. The empirical influence vector of the k-th quote is

$$\mathsf{IF}_k(\hat{\Pi}) = \big(y_k^{\mathsf{Mkt}} - g_k(\hat{\Pi})\big)\big(J^\top W J\big)^+ J_k^\top w_k, \tag{7.17}$$

where J_k is the k-th row of the Jacobian. The scalar influence score used in the empirical plots is

$$\| \mathsf{IF}_k(\hat{\Pi})\|_2.$$

This score ranks instruments by their realised local impact on the parameter vector. It is especially important because a high-leverage quote with a negligible residual need not move the parameters much, while a moderate-leverage quote with a large residual can become influential.

This interpretation is related to classical influence diagnostics in regression, where leverage and residual size jointly determine the effect of an observation on the fitted model; see [23, 24, 8].

A limited conceptual parallel to model-interpretability work in machine learning is that both make a complex input–output relation locally readable; compare [78]. Here, the diagnostics are derived analytically from the WLS geometry of the G2++ pricing map. Their role is to identify which market quotes and parameter directions carry the local calibration geometry. This differs from model-agnostic black-box explanations and from SHAP-type additive decompositions.

Effective Degrees of Freedom

The effective degrees of freedom are

$$\mathsf{EDoF} = \mathsf{tr}(H_w) = \mathsf{rank}(W^{1/2} J).$$

Under the singular-value-decomposition-based (SVD-based) Moore–Penrose construction, the trace equals the numerical rank of the weighted Jacobian and therefore falls in integer steps when local parameter directions are lost. A Tikhonov-regularised smoother would yield non-integer traces, but then the diagnostic would mix genuine rank loss with the chosen regularisation level; see [9, 53] for the relevant matrix background. Here the integer EDoF are meant to make rank loss visible.

Fisher-Information-Type Matrix, Covariance, and Robust Scale

Conditional on the realised calibration design and weights, the Fisher-information-type matrix is approximated by

$$\mathcal{I}(\hat{\Pi}) \approx \frac{1}{\sigma^2} J(\hat{\Pi})^\top W J(\hat{\Pi}).$$

The corresponding covariance approximation is

$$\widehat{\mathsf{Cov}}(\hat{\Pi}) \approx \sigma^2 \big(J(\hat{\Pi})^\top W J(\hat{\Pi})\big)^+. \tag{7.18}$$

The residual scale σ can be estimated classically by an MSE-type estimator, but this is sensitive to outlying residuals. In the cap study, short-end residuals can occasionally be much larger than the typical residual level. Squaring these residuals can inflate all confidence intervals, a phenomenon closely related to masking in regression diagnostics; see [6].

For this reason, the main uncertainty analysis uses a MAD-based scale estimate. With weighted residuals

$$e_k = \sqrt{w_k}\big(y_k^{\mathrm{Mkt}} - g_k(\hat{\Pi})\big),$$

the robust scale is

$$\hat{\sigma}_{\mathrm{MAD}} = 1.4826 \cdot \mathrm{median}_k \left| e_k - \mathrm{median}_j(e_j) \right|.$$

The factor 1.4826 is the normal-consistency constant $1/\Phi^{-1}(0.75)$. Indeed, if $Z \sim \mathcal{N}(0,1)$, then by symmetry $\mathrm{median}(|Z|) = \Phi^{-1}(0.75) \approx 0.67449$. For $X \sim \mathcal{N}(0,\sigma^2)$, the raw median absolute deviation therefore satisfies $\mathrm{median}(|X - \mathrm{median}(X)|) = \sigma\,\Phi^{-1}(0.75)$. Multiplying by $1/\Phi^{-1}(0.75) \approx 1.4826$ makes the MAD consistent for the Gaussian standard deviation. The Gaussian reference model enters only through this scaling constant. The use of the MAD is motivated by robustness: it limits the effect of a small number of extreme residuals on the global uncertainty estimate; compare [26].

Boundary-Respecting Confidence Intervals

The covariance matrix (7.18) is first propagated to the transformed parameter space described in Section 7.2.3. Confidence intervals are then constructed in the transformed space and mapped back. The interpretation remains local: the intervals are conditional on the realised calibration design, the local WLS approximation, the chosen residual scale, and the pseudo-inverse treatment of rank deficiency. Near active constraints, very narrow or collapsed intervals do not necessarily indicate high statistical precision. They may also reflect a direction that has been removed or truncated by the pseudo-inverse construction and should therefore be interpreted jointly with the EDoF diagnostic.

7.2.9 Empirical Diagnostic Study Design

The empirical diagnostic results are based on Euro ATM caps with six-month tenor. Throughout this empirical diagnostic study, all investigations are reported on one unified sample from 1 September 2016 to 2 June 2025, containing 2157 trading days. This common window is used deliberately so that fit measures, leverage, influence, regime sensitivity, and the later calibration-and-validation discussion all refer to the same realised market history. It contains the negative-rate environment, the COVID stress episode, the structural break after 2022, the hiking cycle, and the first phase of monetary easing, making the sample diagnostically demanding.

The post-2022 regime shift serves as a stress period for the calibration methodology. It shows how good price fit can coexist with weaker local identification and more fragile parameter interpretation.

The diagnostics developed above have two readings. Pointwise diagnostics assess the local geometry of one calibrated daily solution. Historical contextual diagnostics compare the diagnostic values of a given calibration date with the empirical distribution of the same diagnostics over the calibration history. They show whether a date is ordinary or unusual relative to previous calibrations of the same model on the same instrument panel. This adds a validation view beyond aggregate fit.

The break after 2022 remains analytically central, but it is studied within the unified 2016–2025 sample. This matters for interpretation: local instability after 2022 belongs to the main empirical evidence within the unified sample.

The empirical comparison is organised around three temporal landmarks: 18 March 2020 for the COVID shock, 21 July 2022 for the first ECB rate hike, and 6 June 2024 for the first ECB rate cut. The three dates serve as common temporal landmarks across the diagnostic plots. They

are carried through the RMSRE, leverage, influence, EDoF, PCA, and uncertainty analyses so that the diagnostic layers remain directly comparable.

The cap maturity grid is

$$T \in \{3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30\} \text{ years.}$$

The term structure is represented in the single-curve OIS/ESTR/EONIA setup fixed in Chapters 3 and 6, with Svensson interpolation between observed OIS maturities. The diagnostic question is therefore how the local G2++ calibration geometry behaves under that documented convention.

For each trading day, the G2++ parameters are calibrated by minimising the RMSRE. The baseline optimisation uses BOBYQA with bounds

$$\sigma_x, \sigma_y \in [10^{-5}, 1], \qquad a_x, a_y \in [10^{-5}, 10], \qquad \rho_{xy} \in [-1, 1],$$

relative tolerance 10^{-8}, and at most 20000 function evaluations. Robustness checks include alternative starting values, a comparison with Nelder–Mead as alternative optimisation algorithm, and reduced portfolios in which the 3Y cap or the 25Y/30Y caps are removed. The reduced portfolios are used as diagnostic what-if experiments. They show how the diagnostics change when the short end is temporarily removed or when the long end is de-emphasised. The resulting diagnostics are qualitatively stable across these variants: the boundary concentration of leverage and the main EDoF patterns remain visible. The baseline specification reported below is therefore interpreted as a representative implementation of the calibration design, supported by stability checks across nearby portfolio choices.

7.2.10 Goodness of Fit over Time

The first empirical diagnostic is the price fit. Figure 7.6 reports the daily RMSRE. The three marked temporal landmarks are March 2020 for the COVID shock, July 2022 for the first ECB rate hike, and June 2024 for the first ECB rate cut.

The figure makes the central distinction visible. The earlier part of the sample contains extended periods with higher RMSRE and several extreme outliers. From 2022 onward, the RMSRE is low on most days, often below roughly 1% to 1.5%. The figure therefore shows what a fit-based diagnostic can and cannot detect. It shows the improvement in price fit after the 2022 market transition, but it does not reveal the parameter-space instability analysed in the following sections. Thus the post-2022 issue is not simply a fit issue. It is a local-identification issue: prices can be fitted well while the parameter vector becomes more weakly or more unevenly determined.

All three marked dates lie in periods with relatively good aggregate fit. A fit-based assessment alone would therefore classify them as adequate calibration dates. Their role in the subsequent analysis is to provide fixed reference points across the leverage, influence, EDoF, and uncertainty diagnostics, where differences in the local calibration geometry become visible.

This distinction matters directly for actuarial use. A fit-only gate would accept many low-RMSRE daily calibrations without noticing that they belong to different local states. For scenario generation under $\mathbb{P}$, and later for PIA- or openIRM-related use cases, these states are not equivalent. The later diagnostics therefore assess low-RMSRE days through their leverage, influence, EDoF, and uncertainty profiles.

7.2.11 Leverage and Influence

The leverage distribution in Figure 7.7 shows a clear boundary pattern. The highest leverage values occur at the maturity edges, especially at the 3Y and 30Y caps. In the 3Y cap, leverage

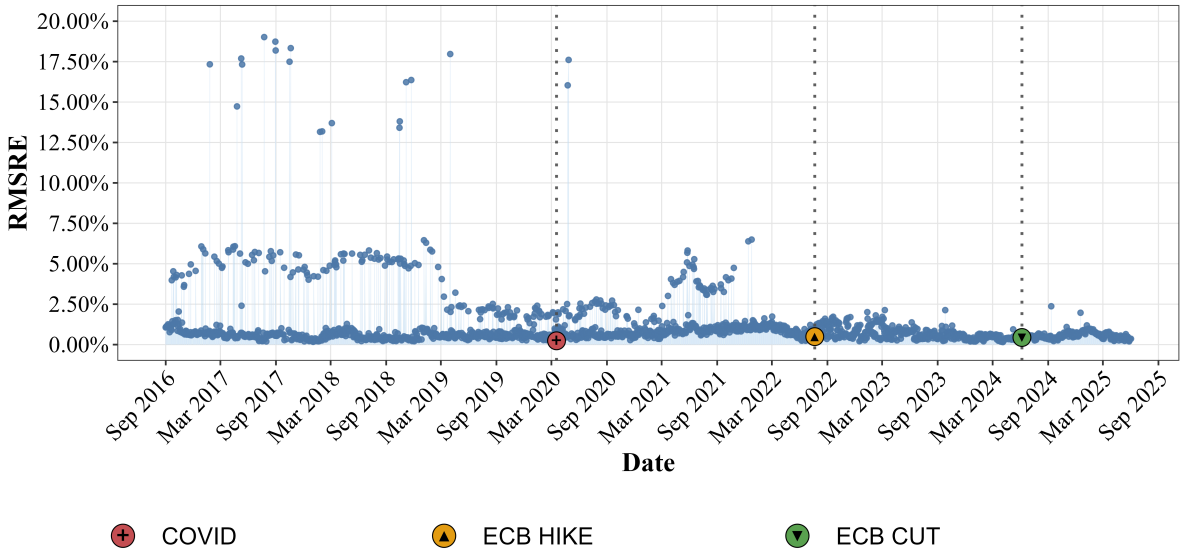


Figure 7.6: Daily goodness of fit of the G2++ cap calibration in the unified 2016–2025 empirical window. The RMSRE is clearly time-varying: the early sample contains extended periods of elevated fit error and several extreme outliers, whereas from 2022 onward the RMSRE is low on most days. The three event badges mark the reference dates used throughout the diagnostic comparison. From the RMSRE perspective alone, these dates do not appear as exceptional calibration failures.

exceeds 0.95 on 73.1% of all days and exceeds 0.99 on 32.7% of all days in the sample. This means that the short end is often nearly indispensable in the local tangent-space reconstruction.

The influence diagnostic captures a different aspect of the local calibration geometry. Figure 7.8 combines the leverage geometry with realised residuals. The influence scores are more diffuse than the leverage scores. No single long-end maturity acts as the unique realised driver across the full sample. The short end, however, exhibits the most extreme tail events. This is exactly what the formula (7.17) predicts: influence is large when high leverage is combined with a sizable residual.

The practical interpretation separates geometric support from realised parameter movement. Boundary maturities provide the main geometric support of the fit. Realised parameter movement arises from the interaction of this support with maturity-specific residuals. The short end is often the main stressor because it combines very high leverage with occasional large residuals.

Scope of the boundary-dominance explanation. The boundary-dominance result is interpreted for the ATM cap design studied here. It follows from the leverage-reconstruction identity, the one-sided position of boundary maturities on the cap grid, the cumulative G2++ cap variance-sensitivity structure, and the empirical verification over the calibration history. Different weights, strike conventions, or maturity grids may lead to different leverage profiles.

The event markers show that low-RMSRE dates can still have different local diagnostic profiles, especially in leverage and influence. The detailed numerical comparison is reported in Table 7.1.

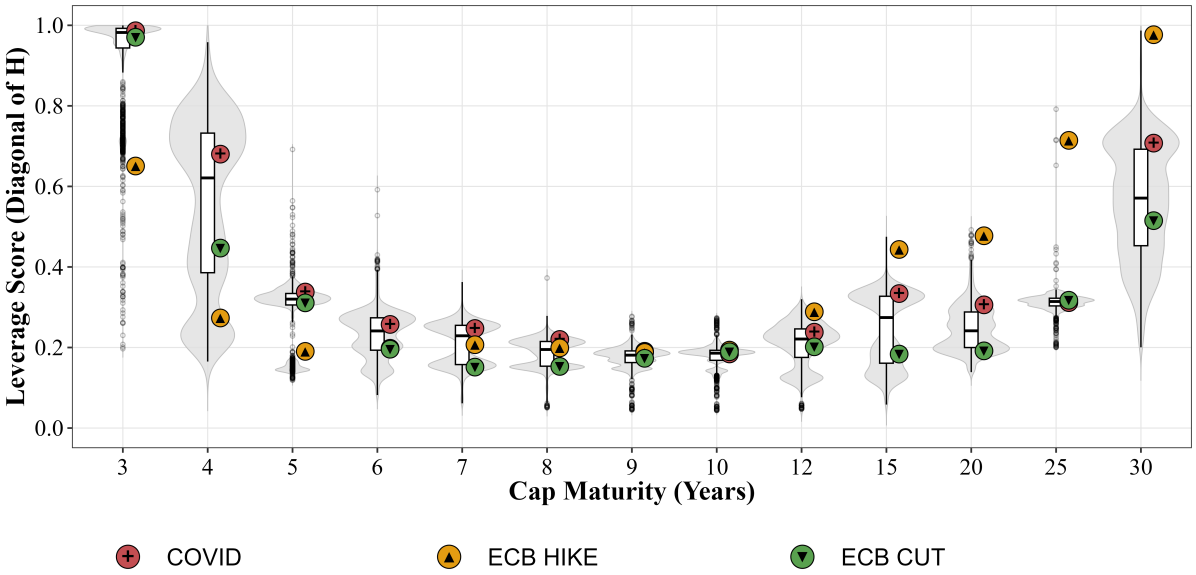


Figure 7.7: Distribution of leverage scores across maturities. The violin layer shows the cross-day distribution, while the boxplot summarises the centre and spread. The profile is boundary dominated rather than perfectly monotone.

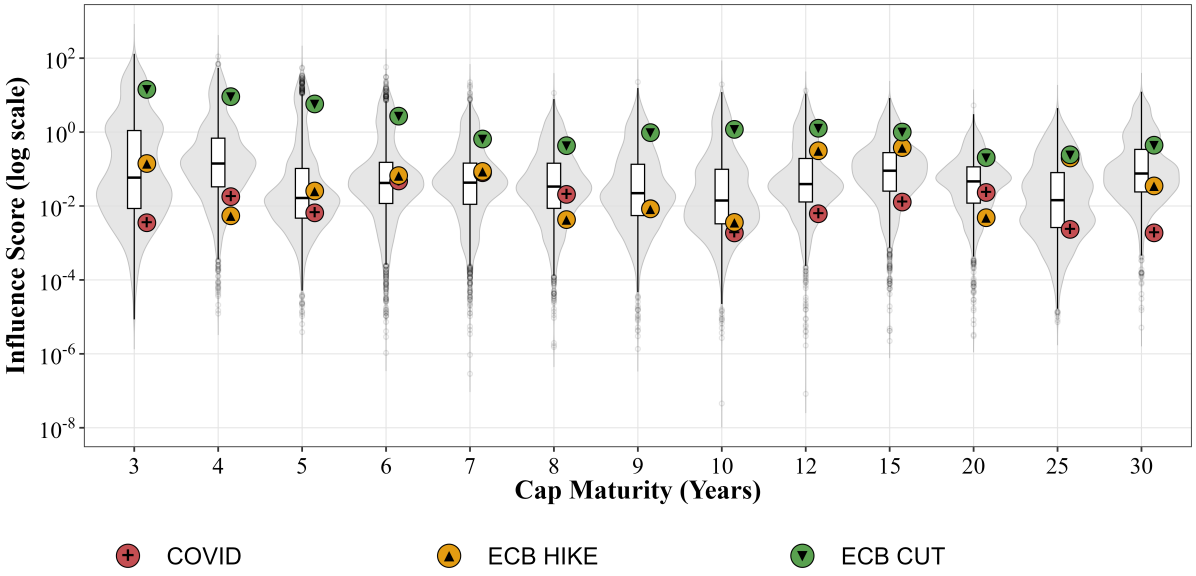


Figure 7.8: Distribution of influence scores across maturities on a logarithmic scale. Influence is less concentrated than leverage because it combines the local geometry with the realised residuals.

7.2.12 Effective Degrees of Freedom and Active Constraints

Figure 7.9 reports the trace of the weighted hat matrix over time. With five G2++ parameters, the full local dimension is five. The distinct lower integer bands indicate rank loss in the Moore–Penrose tangent-space diagnostic.

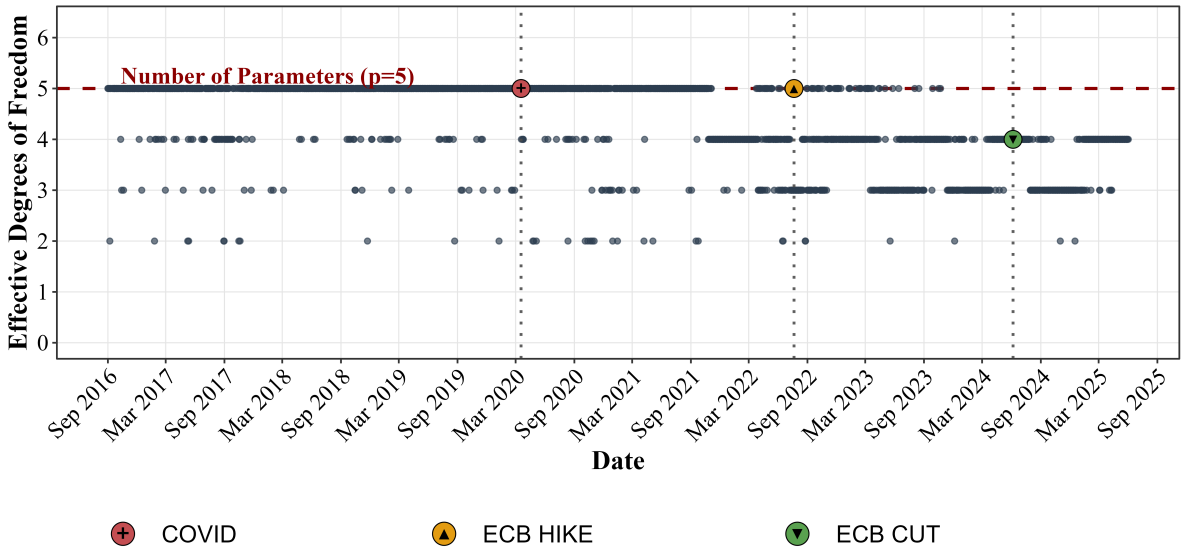


Figure 7.9: Effective degrees of freedom, measured as the trace of the weighted hat matrix. Most daily calibrations lie exactly on the theoretical maximum of 5, indicating full local rank of the Jacobian. The distinct lower integer bands reveal instances of rank loss, not a smooth regularisation effect. Switching between integer levels becomes more frequent after the 2022 market transition.

The Tikhonov-regularised EDoF plot provides a complementary view of the same local geometry. The Moore–Penrose EDoF are integer-valued because, under the SVD-based pseudo-inverse, the trace reflects the numerical rank of the weighted Jacobian. Under Tikhonov regularisation, the trace no longer coincides with this numerical rank. It becomes a regularisation-dependent effective dimension and can therefore vary continuously between integer values. This gives a smoother stability view than Figure 7.9, while its numerical level also depends on the chosen regularisation parameter. For this reason, the Moore–Penrose EDoF remain the primary rank-loss diagnostic in this dissertation. The Tikhonov version is used as an additional comparison.

In the Moore–Penrose diagnostic, most days remain at the full value of five. However, the periods with lower integer values are diagnostically important. They indicate that the calibration uses fewer than five locally independent directions, even when the price fit is good.

This often coincides with active parameter constraints, especially the lower correlation bound $\rho_{xy} = -1$. In the baseline calibration, ρ_{xy} lies at that lower bound to within 10^{-8} on 66.4% of trading days and to within 10^{-6} on 69.6% of trading days. The main EDoF and leverage plots use the ordinary full-Jacobian Moore–Penrose construction: all five Jacobian columns are retained, while singular directions are handled by the pseudo-inverse. Under this convention, the maturity-boundary concentration of leverage persists on lower-correlation-bound days. As an additional active-set check, the correlation coordinate can be removed from the local set of free parameter directions when ρ_{xy} is at its lower bound. With this convention, the maturity-boundary concentration remains visible on most of these days, although exceptions occur. The result supports the observed leverage pattern and shows that its precise form depends on the

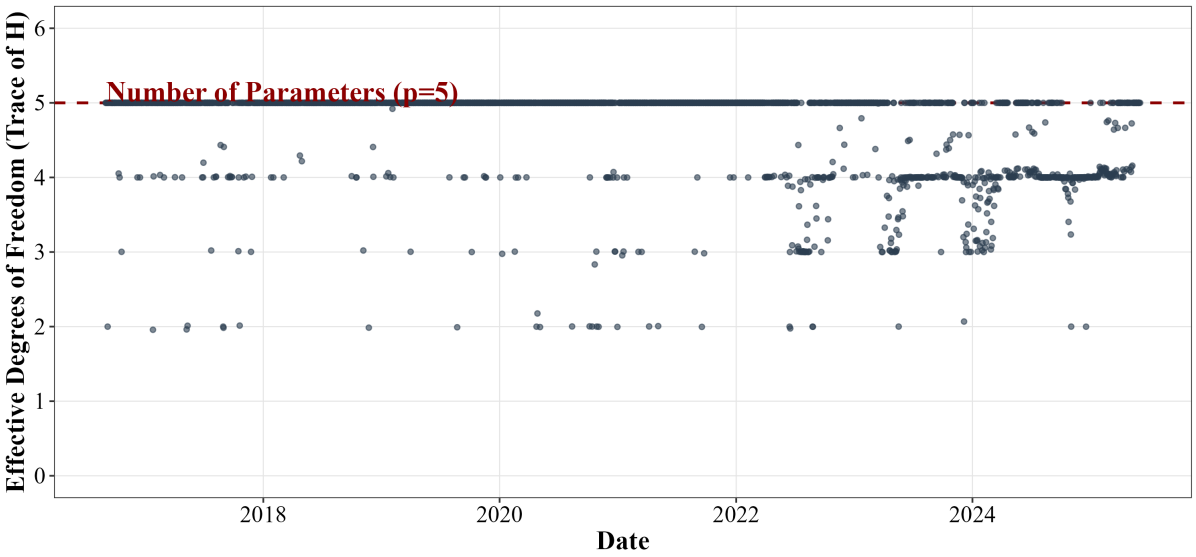


Figure 7.10: Tikhonov-regularised effective degrees of freedom for the same cap calibration design. The regularised trace varies continuously between integer values because the rank transition is smoothed by the regularisation term.

realised calibration design and active-set convention.

The event dates make this distinction concrete. Table 7.1 reports three low-RMSRE days that appear similar from an aggregate fit perspective, while their leverage, influence, EDoF, and uncertainty diagnostics differ.

Table 7.1: Three calibration reference cases. The three dates are marked consistently in the time-series and violin-boxplots. Although all three have low aggregate RMSRE, the local diagnostics lead to different interpretations.

Date	Event	RMSRE	EDoF	Max lev.	Max infl.	Diagnostic reading
2020-03-18	COVID	0.24%	5.00	0.99 (3Y)	0.08 (7Y)	Full EDoF; 3Y leverage appears structural.
2022-07-21	ECB hike	0.49%	5.00	0.98 (30Y)	0.38 (15Y)	Good fit with long-end dominance.
2024-06-06	ECB cut	0.44%	4.00	0.97 (3Y)	14.27 (3Y)	Reduced EDoF with large 3Y influence.

The ECB-hike date illustrates the historical contextual reading. Its RMSRE and EDoF are unremarkable on their own. The historical leverage comparison, however, shows a long-end-dominated local geometry: on this date the 30Y cap carries the maximum leverage, whereas the median historical pattern is more concentrated at the short end. Several maturity-specific leverage values lie close to the outer empirical boxplot range. The date is therefore mainly distinctive in its leverage geometry; its aggregate fit remains ordinary.

The ECB-cut date shows the strongest separation between aggregate fit and local diagnostics. The aggregate RMSRE remains small, while the largest single relative residual reaches 47.9% at the 3Y cap. Similar aggregate fit quality can therefore correspond to materially different local diagnostic states.

Taken together, the three dates provide fixed reference points for three diagnostic situations: a

COVID shock date with comparatively mild local diagnostics, an ECB-hike date with long-end leverage concentration, and an ECB-cut date with low RMSRE and clear local instability. This is why the same dates are carried through the empirical diagnostic layers.

7.2.13 Regime Sensitivity and Parameter Uncertainty

The PCA plots in Figures 7.11 and 7.12 visualize the movement of the calibrated parameter vectors. The first principal component mainly captures volatility-scale movements through σ_x and σ_y. The second component is driven by mean reversion and correlation. The 2016–2021 part of the sample is concentrated in a relatively stable corridor, with most movement explained by volatility-scale variation. The post-2022 phase shows a clear vertical expansion in the PCA space, which aligns with the increased switching in the EDoF plot. Within this later phase, the 2024–2025 observations show much stronger movement in mean-reversion directions, in particular through a_x and a_y.

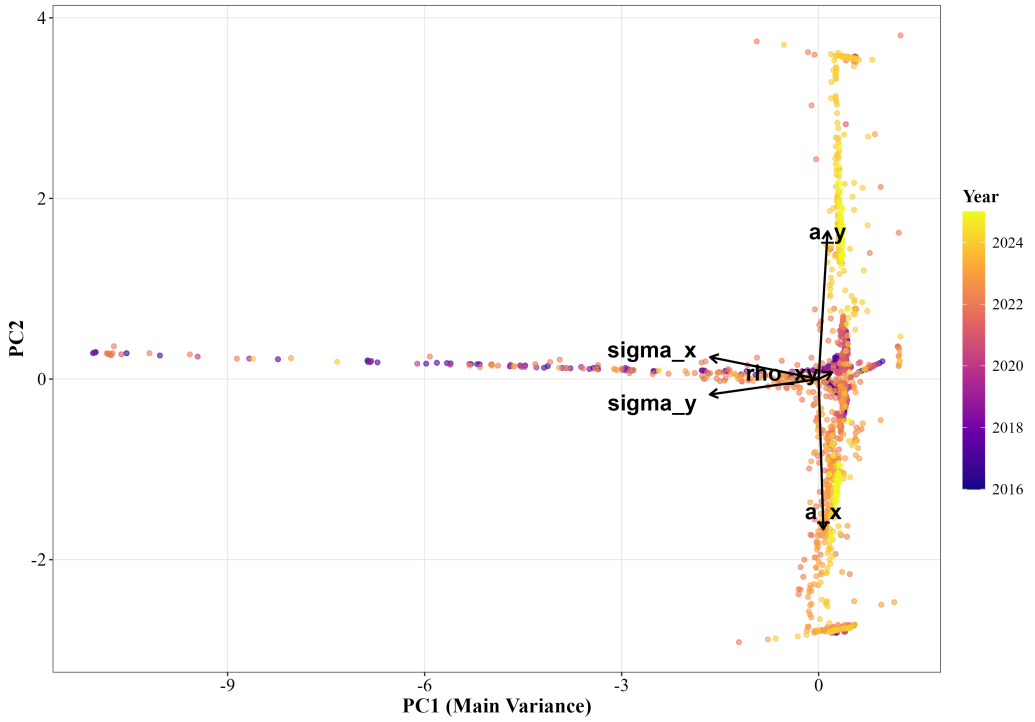


Figure 7.11: PCA biplot of calibrated G2++ parameters. The horizontal direction mainly captures volatility scale, while the vertical direction reflects mean-reversion and correlation movements.

Parameter uncertainty is then quantified through the Fisher-information-type matrix, the corresponding local covariance approximation, robust MAD scaling, and the transformations from Section 7.2.3. Figure 7.13 reports the robust fan chart for the mean-reversion speed a_y.

The break in $\hat{a}_y$ visible in Figure 7.13 is the coordinate-level expression of the regime movement shown in Figures 7.11 and 7.12. Before the post-2022 regime, the main parameter movement is largely captured by the volatility-scale coordinates. From 2024 onward, the mean-reversion coordinates become more variable, and $\hat{a}_y$ moves repeatedly into values clearly above one. Since a_y is a mean-reversion speed, these high values correspond to a short y-factor time scale $1/a_y$. In this calibration, the movement in a_y should therefore be read jointly with a_x, σ_y, ρ_{xy}, and the local rank diagnostics. Large values can reflect compensation between mean reversion, volatility scale, and correlation in a weakly separated parameter direction.

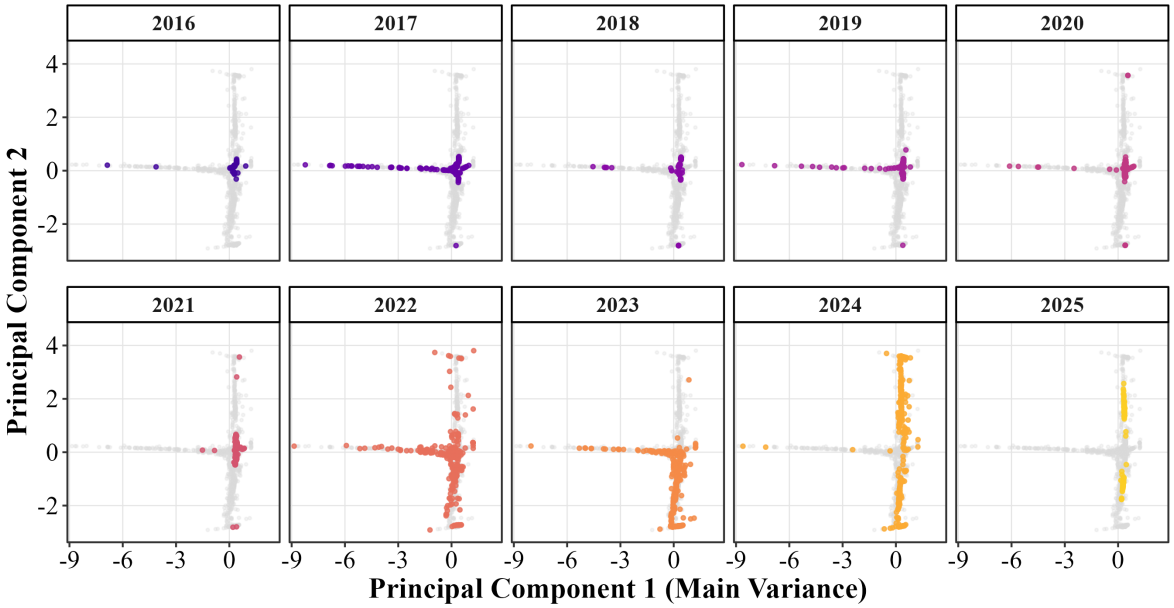


Figure 7.12: Temporal evolution of calibration regimes. The faceted view shows the transition from the relatively stable 2016–2021 corridor to the post-2022 regime with stronger movement in structural parameter directions.

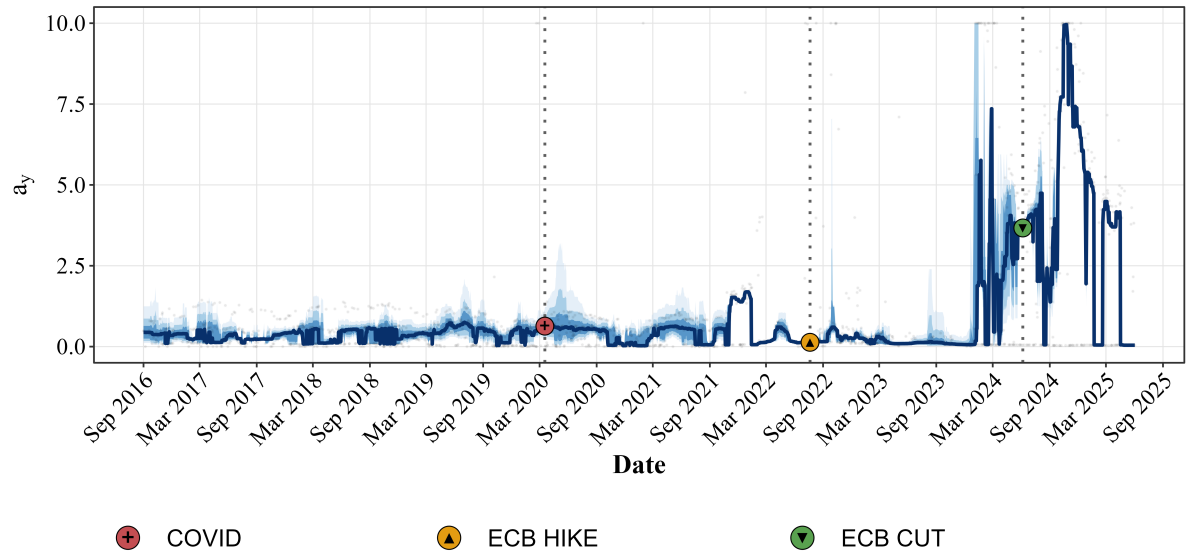


Figure 7.13: Time-varying confidence intervals for a_y. The solid line shows the daily calibrated value $\hat{a}_{y,t}$, while the shaded areas represent the 50%, 75%, and 95% confidence intervals obtained via the Variance Stabilizing Transformations, the Functional Delta Method, and robust MAD-based covariance scaling.

The robust intervals are asymmetric and remain in the admissible parameter domain. They also avoid a persistent full-range uncertainty band that arises under classical MSE-based scaling. Figure 7.14 shows the classical scaling for comparison. The classical MSE scale is much more sensitive to extreme residuals and therefore inflates the interval bands over wider periods.

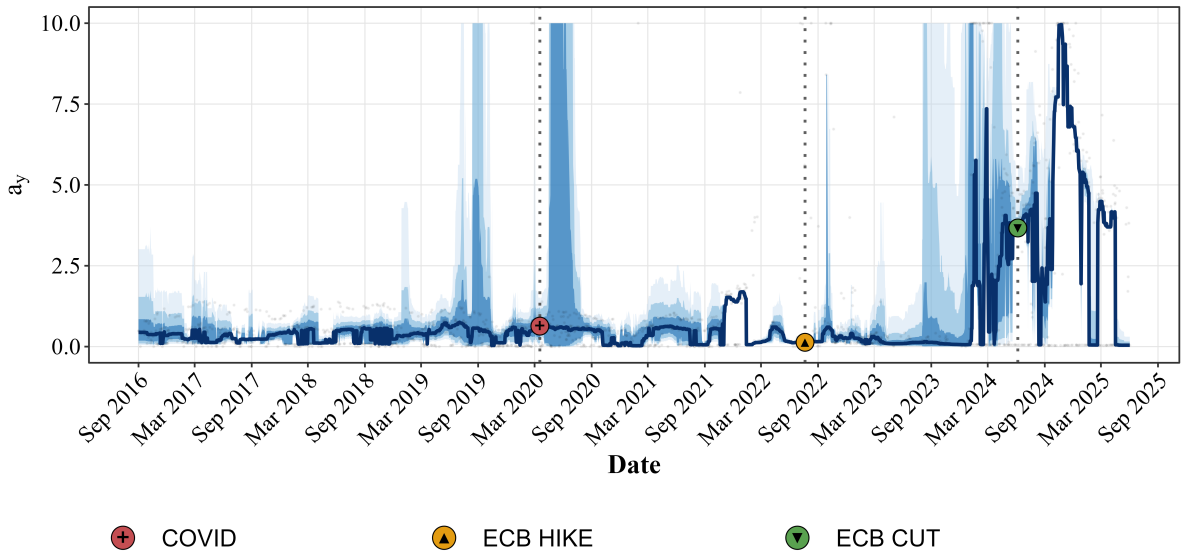


Figure 7.14: Classical MSE-based confidence intervals for a_y. Compared with the robust MAD-based version in Figure 7.13, the classical bands react much more strongly to a small number of extreme residuals.

Collapsed confidence bands require particular care. Under the SVD-based pseudo-inverse, directions that are not locally identified are truncated instead of inflated. A collapsed coordinate-wise interval can therefore mean that the local covariance approximation no longer treats the coordinate as a free direction. It should never be read as high precision without checking the EDoF and active-constraint diagnostics. For the confidence intervals, the event dates mainly serve as common reference points, not as the dates with the largest uncertainty bands.

7.2.14 Boundary-Dominated Leverage

The leverage profile in Figure 7.7 often appears U-shaped. The claim made here is narrower: in the realised Euro ATM cap calibration, the leverage geometry is boundary-dominated, and this boundary-dominance can be explained by the cumulative structure of the G2++ cap Jacobian.

The claim is conditional on the realised calibration design. Leverage depends on the weighted Jacobian rows and is therefore affected by weights, strikes, market prices, and the maturity grid. The boundary-dominance interpretation is therefore built from the analytical caplet factorisation, the cumulative cap-row structure, the reconstruction identity, and empirical verification on the realised design.

First, the caplet sensitivity factorisation (7.7) shows that each caplet contributes a positive price-to-variance factor times a variance-sensitivity vector. Second, a cap Jacobian row is the sum of all caplet contributions belonging to that cap. For the six-month cap structure, these rows are cumulative maturity-ordered sums of G2++ variance-sensitivity vectors. The shortest-maturity cap has no shorter neighbouring cap from which its Jacobian row can be interpolated. The longest-maturity cap has no longer neighbouring cap. Interior caps, by contrast, overlap strongly with neighbouring caps on both sides.

This intuition is related to the special role of boundary and support points in optimal-design and Chebyshev-system arguments [70, 69, 28]. These references provide structural motivation for why endpoint rows can become difficult to replace. The G2++ cap-specific claim is then verified directly for the calibration design studied here.

Third, ordinary least-squares projection geometry gives an exact identity linking leverage to row reconstruction. Let $X = W^{1/2}J$ be a full-column-rank weighted design, and write $x_k^\top$ for its k-th row. The reconstruction equation asks whether x_k can be written as a linear combination of the remaining rows:

$$X_{-k}^\top a = x_k, \qquad a \in \mathbb{R}^{m-1}.$$

Define the minimum-norm reconstruction cost

$$\mu_k = \inf\left\{\|a\|_2^2 : a \in \mathbb{R}^{m-1}, \quad X_{-k}^\top a = x_k\right\},$$

with $\mu_k = +\infty$ if this reconstruction equation has no solution. This convention corresponds to the rank-drop case: after deleting row k, the remaining rows no longer span the deleted row direction. In that case no finite coefficient vector a can reconstruct x_k from the remaining rows. Then

$$h_k = \frac{\mu_k}{1+\mu_k}, \tag{7.19}$$

where the right-hand side is interpreted through the continuous limit $\mu_k \to +\infty$, which gives $h_k = 1$. Thus high leverage means high reconstruction difficulty. The boundary interpretation is therefore supported by reconstruction geometry: boundary rows have high leverage when they are difficult to reconstruct from the remaining maturity rows.

The corresponding realised reconstruction costs are shown in Figure 7.15. The scale is different from the leverage plots because the relation $h_k = \mu_k/(1+\mu_k)$ compresses large reconstruction costs into leverage values close to one. The economic reading is nevertheless the same: the short boundary maturity is the hardest row to reconstruct, and the long boundary maturity is again harder to reconstruct than the central cap maturities.

Appendix D contains the algebraic reconstruction details and the exact benchmark-row definitions. Table 7.2 summarises the numerical evidence used for the maturity-boundary-dominance interpretation. It separates two levels of evidence: a row-normalised benchmark that removes scale effects, and realised daily diagnostics from the full calibration history.

The first two rows of Table 7.2 concern the row-normalised cumulative G2++ sensitivity-vector benchmark. In this benchmark, row-scale effects from WLS weights and strike-dependent price-to-variance sensitivities are removed. What remains is the directional cumulative variance-sensitivity structure of the cap Jacobian. The benchmark verification is carried out on the empirical maturity grid

$$(3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30)$$

with six-month tenor. The positive maturity-boundary gap throughout the valid grid, together with the local-search check around the worst grid points, supports the structural benchmark explanation.

The last two rows return to the realised daily calibration history. They compare the full-Jacobian Moore–Penrose diagnostic used in the main leverage analysis with an active-set diagnostic in which the binding correlation direction is removed on days with $\hat{\rho}_{xy} = -1$. The full-Jacobian diagnostic remains maturity-boundary dominated on the lower-correlation-bound subset. The active-set diagnostic gives a stricter constrained tangent-space reading and preserves the maturity-boundary maximum on 96.1% of these days.

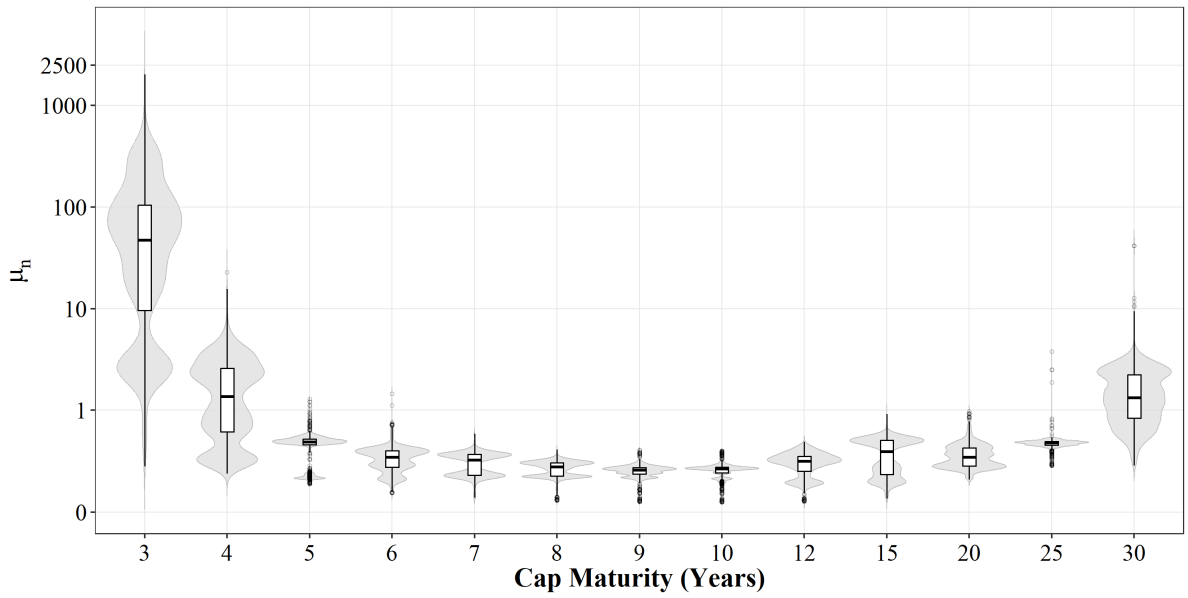


Figure 7.15: Minimum-norm reconstruction difficulty by cap maturity on the OIS 6M ATM-cap calibration panel. The violin-boxplot shows the Moore–Penrose reconstruction proxy on a pseudo-logarithmic scale, because the short-boundary costs are much larger than the central-maturity costs. The figure gives the reconstruction-side reading of the leverage pattern: boundary maturities are harder to recover from the remaining maturity rows.

Table 7.2: Evidence for maturity-boundary-dominated leverage in the G2++ cap design. The table separates the row-normalised sensitivity-vector benchmark from the realised daily full-Jacobian and active-set diagnostics.

Diagnostic check	Sample / grid	Result	Interpretation
Row-normalised cumulative G2++ benchmark grid	11,200 candidate points; 9,800 valid points on the 13-maturity six-month cap grid	All valid points have positive maturity-boundary gap; $\min_{\text{grid}} \bar{\eta} = 0.2611$	Maturity-boundary dominance remains after removing row-scale, weight, and strike-specific price-to-variance effects.
Local search around the benchmark minimum	Local searches started from the worst finite-grid points	$\min_{\text{local search}} \bar{\eta} = 0.1941$	The finite-grid minimum is stable in the tested neighbourhoods.
Full-Jacobian diagnostic on lower-correlation-bound days	1,432 days with $\hat{\rho}_{xy} = -1$	Ordinary maturity-boundary gap remains positive; minimum 0.0559	The diagnostic used in the main leverage panel remains maturity-boundary dominated on lower-correlation-bound days.
Active-set diagnostic on lower-correlation-bound days	Same 1,432 days, with the active correlation direction removed	Maturity-boundary maximum present on 96.1% of days	The stricter constrained tangent-space reading supports the pattern on most lower-correlation-bound days, with exceptions.

These checks support the term maturity-boundary-dominated leverage for the chosen calibration design. The statement is conditional on the realised WLS design and on the benchmark restrictions stated in Appendix D. Different weights, maturity grids, strike conventions, or parameter locations may lead to different leverage profiles. The algebraic reconstruction identity explains why endpoint rows can become hard to replace, and Table 7.2 reports the corresponding numerical evidence for the cap design studied here.

7.2.15 From Analytical Cap Diagnostics to Numerical Swaption Diagnostics

After the cap diagnostics have established the analytical benchmark, the same local WLS geometry can be transferred to swaptions through a numerically computed Jacobian. Caps remain the main empirical reference of this chapter because their analytical Jacobian gives a clean diagnostic geometry. For G2++ swaptions, prices are available in semi-analytical form: the price is evaluated by a one-dimensional integral, while the exercise boundary is determined implicitly. The diagnostic issue is therefore numerical stability of the local Jacobian.

The transfer is local. If the swaption pricing map is differentiable in a neighbourhood of the calibrated parameter vector $\hat{\Pi}$, and if the numerical Jacobian passes step-size, tolerance, and boundary checks, the weighted projector, the Fisher-information-type matrix, the influence approximation, and the covariance propagation are constructed exactly as in the cap case. The implementation difference is that the swaption Jacobian is obtained by finite differences. Appendix C records the corresponding local differentiability argument and the numerical-governance checks.

The swaption diagnostic layer is therefore a numerical extension of the cap-based framework. The numerical-Jacobian construction uses the OIS 6M swaption panel over the 2016–2025 record, central finite differences, and boundary-aware clipping. After aligning swaption quotes, OIS curves, calibrated parameter vectors, valid instruments, finite-difference evaluations, and governance checks, the panel contains 1,334 observation dates from 5 September 2016 to 2 June 2025. The selected universe contains 89 possible swaptions with option maturity plus swap tenor at most 20 years; the median number of valid instruments is 62 per date.

Table 7.3 summarises the resulting full-window diagnostics. The median EDoF is 4.0, and about 31.7% of dates reach full local rank five. The daily maximum leverage has median 0.718 and a 95% quantile close to one, so individual swaptions can become nearly irreplaceable in the local projection. Influence is usually small, but some dates show large realised parameter impact. Step-size checks are mostly stable on representative dates, although one stressed date records a relative Frobenius difference of 0.711. This identifies a limitation of the swaption diagnostic layer: its conclusions remain conditional on the finite-difference and quadrature setup.

Figures 7.16–7.19 show the corresponding distribution and time-series evidence. The leverage violin-boxplot separates option maturity by swap tenor and shows that the local projection is unevenly distributed across the swaption surface. The EDoF point plot shows that rank loss becomes more frequent in the later market regime. The influence plots add the realised residual layer: high leverage is a geometric property, whereas influence becomes large only when high-leverage directions interact with residuals and conditioning.

The full-window OIS swaption panel supports a limited empirical conclusion: leverage, EDoF, and influence diagnostics can be transferred to swaptions under explicit numerical governance. The evidence is based on a full-window numerical panel for the OIS swaption setup. It does not turn swaptions into a closed-form analytical benchmark. Caps remain the clean diagnostic reference; swaptions provide a numerical extension and an additional stress test of the same governance logic.

Table 7.3: Full-window numerical-Jacobian diagnostics for the OIS 6M swaption panel with option maturity plus swap tenor at most 20 years. The table reports daily WLS diagnostics computed from the finite-difference swaption Jacobian.

Metric	**Value**
Observation dates	1334
Date window	2016-09-05 to 2025-06-02
Median valid swaptions	62
Median RMSRE	3.41%
Mean RMSRE	4.26%
Median EDoF	4.0
Share with full rank	31.71%
Median daily max. leverage	0.718
95% quantile daily max. leverage	1.00
Median daily max. influence norm	0.0101
95% quantile daily max. influence norm	0.109

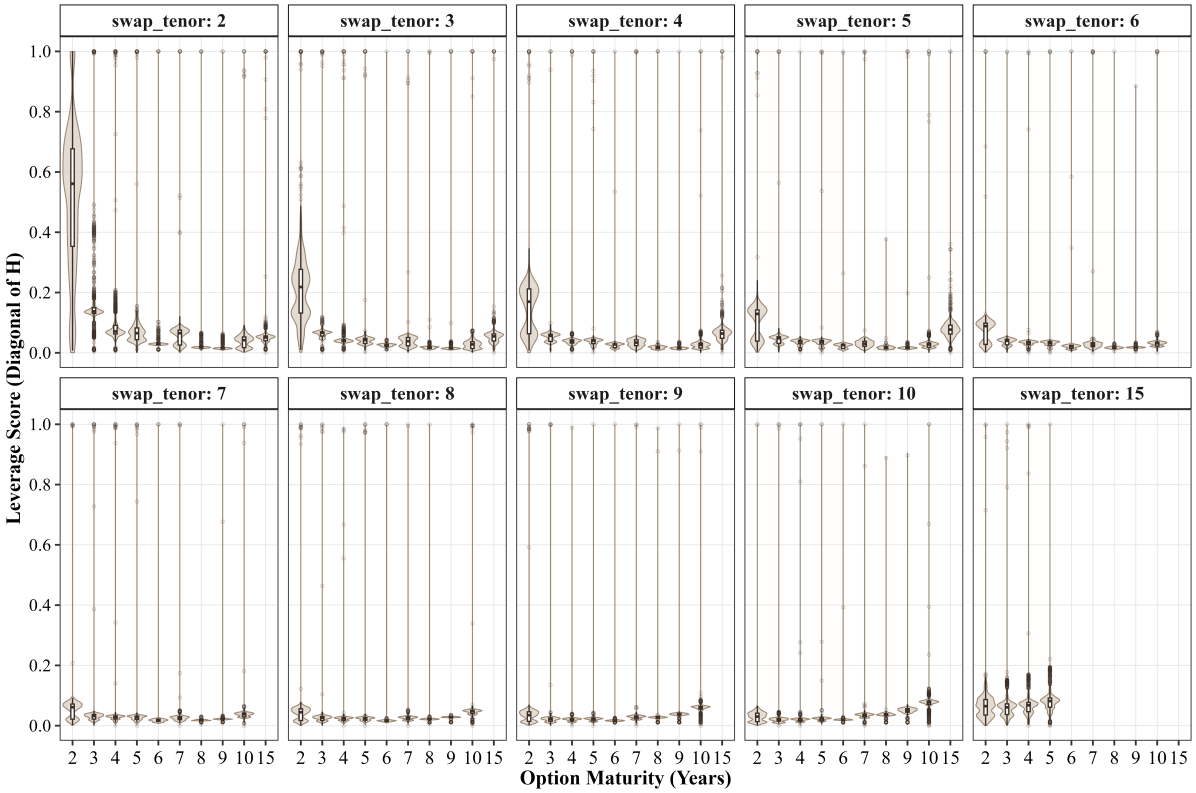


Figure 7.16: Instrument-level leverage distributions for the full-window OIS 6M swaption numerical-Jacobian panel. Violin-boxplots are grouped by option maturity and faceted by swap tenor.

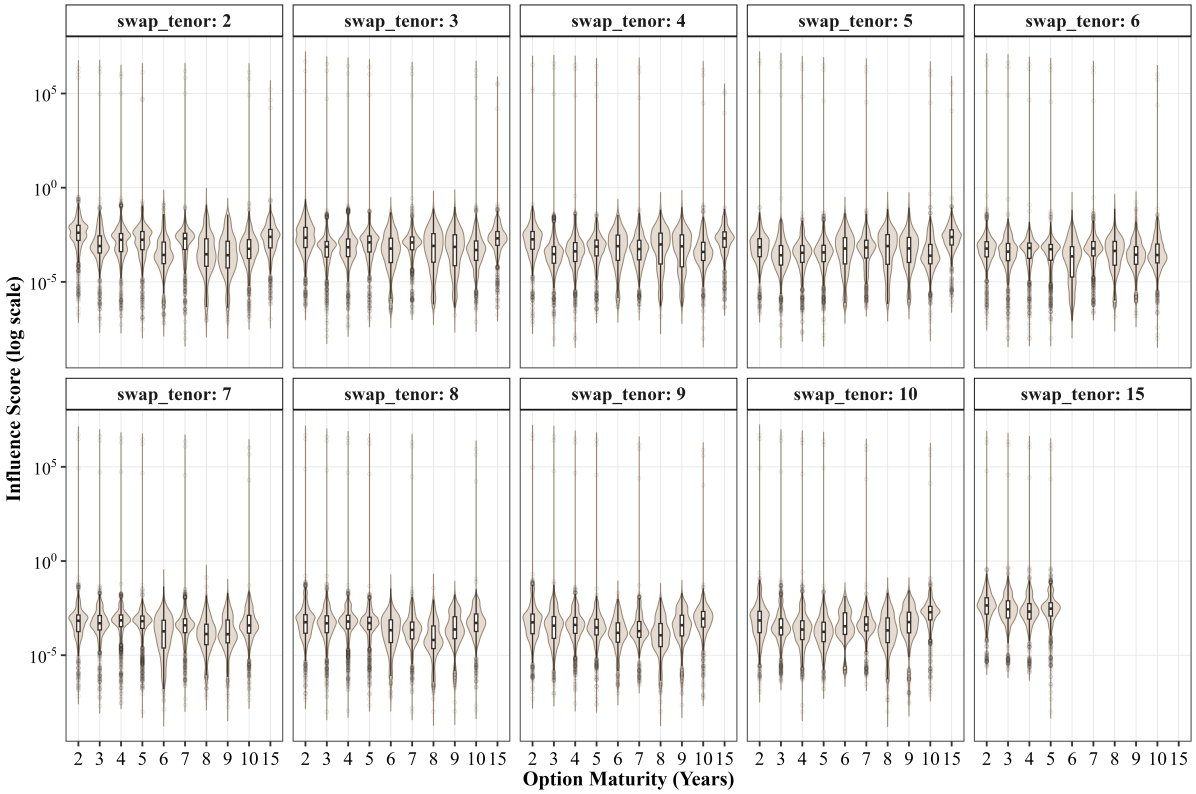


Figure 7.17: Instrument-level influence-norm distributions for the full-window OIS 6M swaption numerical-Jacobian panel. Violin-boxplots are shown on a logarithmic scale to separate ordinary observations from a small number of high-influence stress dates.

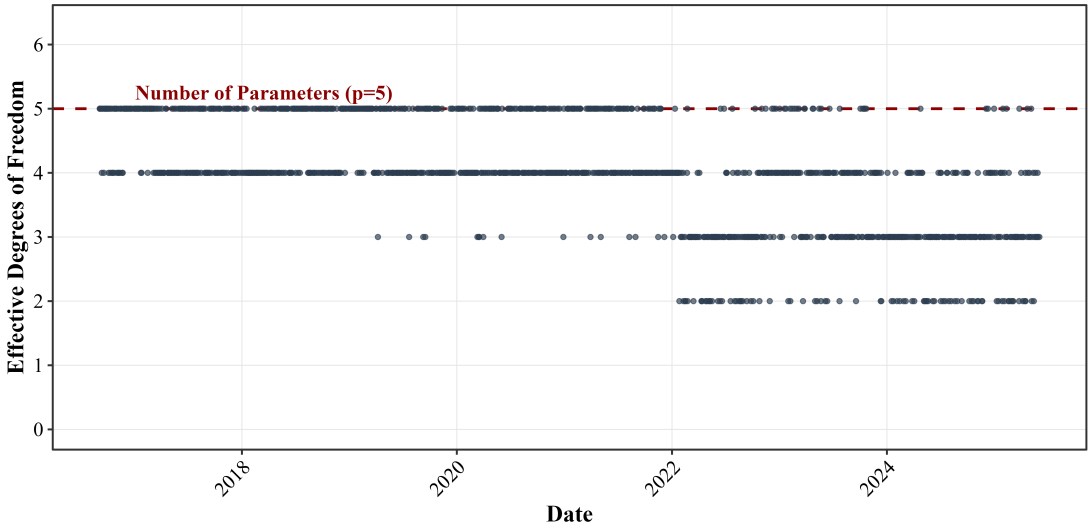


Figure 7.18: Effective degrees of freedom for the full-window OIS 6M swaption numerical-Jacobian panel. The point cloud shows the daily trace of the weighted hat matrix against the five-parameter reference line.

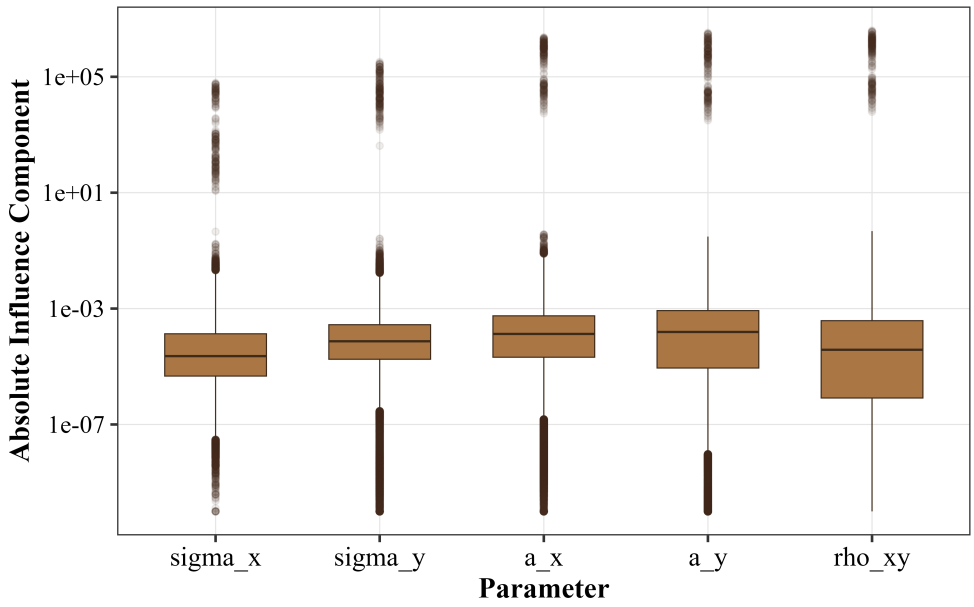


Figure 7.19: Absolute influence-function components by parameter for the full-window OIS 6M swaption numerical-Jacobian panel. The plot reports the realised WLS influence decomposition, not a global parameter-uncertainty statement.

7.2.16 Why the Daily Calibration Is Kept Stand-Alone

The local diagnostics above also motivate the empirical design of the study. In some practical G2++ applications, daily calibration is stabilised by adding a sequential penalty to the current-day objective. Schematically, the optimisation target becomes

$$Q_t^{\text{seq}}(\Pi_t) = Q_t^{\text{fit}}(\Pi_t) + c_t\, P(\Pi_t, \Pi_{t-1}),$$

where $P(\Pi_t, \Pi_{t-1})$ penalises day-to-day parameter movement and c_t controls the strength of the smoothing. Such a device can reduce visible parameter jumps and produce smoother reporting series. For the empirical diagnostics in this chapter, however, the primary object is the stand-alone daily calibration.

There are three reasons for this choice. First, a penalty based on absolute parameter differences can distort the relative importance of the G2++ coordinates. Mean-reversion speeds, volatilities, and correlations have different scales and different effects on option prices. A smoothing term in unstandardised parameter space may therefore penalise numerically large coordinates more strongly than economically relevant ones. If sequential stabilisation is used in practice, it should be formulated in relative or otherwise standardised coordinates.

Second, a sequential penalty changes the information content of a calibration date. A stand-alone risk-neutral calibration describes the market prices observed on that date. With a non-negligible penalty to Π_{t-1}, the estimate becomes a compromise between current market prices and historical parameter continuity. Around shocks or regime changes, this can create an information lag: the parameter vector reacts slowly because continuity has been assigned explicit weight in the objective.

Third, smoothing can mask the identification issues studied in this chapter. Low RMSRE can coexist with weak local identification, leverage concentration, active constraints, and large parameter uncertainty. A sequential penalty does not remove these features from the calibration geometry. It can keep the optimiser in a preferred region of parameter space and thereby create

a smoother parameter path even when the current market prices identify the parameters only weakly.

For this reason, the main empirical analysis is based on daily stand-alone calibrations. Diagnostically exceptional days are identified through local geometry, influence, effective rank, and uncertainty, rather than smoothed away before these diagnostics are applied. Sequential stabilisation may still be useful as a governance or reporting overlay. In that case, its scaling, weight choice, and effect on market fit should be documented explicitly. The scientific calibration object of this chapter remains the unsmoothed daily calibration.

7.2.17 From Local Diagnostics to Validation Governance

The diagnostics in this section are local diagnostics of the calibration under $\mathbb{Q}$. They describe the geometry of the fitted pricing map: which quotes span the tangent space, where influence concentrates, whether rank is lost, and how large local parameter uncertainty is under the chosen WLS approximation. Chapter 10 adds a different layer by testing the simulated outputs produced after the transition to $\mathbb{P}$.

The practical consequence is a clear ordering of the calibration workflow. A daily calibration is first screened by fit, local geometry, and realised sensitivity. The accepted $\mathbb{Q}$ parameter history is then carried forward to the $\mathbb{P}$ calibration and to actuarial validation.

The diagnostic layer also defines a diagnostic follow-up procedure. If a day is flagged by low EDoF, extreme leverage, large realised influence, or unstable local covariance, it is recorded as diagnostically exceptional. The next step is to perturb the dominant quotes or their weights and rerun the local diagnostics. This separates high leverage with accurate pricing from high leverage combined with large residuals. The final step is to compare application-level quantities, including the fitted $\mathbb{Q}$ parameter path, the subsequent $\mathbb{P}$ risk-premium calibration, and selected actuarial scenario outputs. The calibration is then accepted as ordinary input, accepted with qualification, recalibrated under a modified design, or quarantined from the historical training set.

This defines the transition criterion used later in the calibration chain. The same principle applies to the swaption and one-factor extensions: they are interpreted diagnostically first, while their simulation consequences are assessed in Chapter 10.

7.3 Empirical Comparison of Calibration Designs

The local diagnostic section above studies a fixed calibration design after the optimiser has returned a parameter vector. This section varies selected design choices and compares their empirical consequences. The comparison keeps the calibration target explicit: the risk-neutral calibration under $\mathbb{Q}$ provides the market-consistent input for the later transition to $\mathbb{P}$ and for actuarial validation. The relevant question is therefore which findings are stable across reasonable calibration designs and which depend on the curve convention, instrument universe, optimiser, parameter bounds, objective function, or model class.

7.3.1 A Common Comparison Protocol

All comparisons in this part of the chapter use the same basic calibration structure. Market quotes are translated into price targets under the relevant market convention, and the G2++ parameter vector is estimated by minimising a documented loss function. Unless stated otherwise, the baseline objective is RMSRE. It compares pricing errors in relative price units and thereby avoids a mechanical dominance of the most expensive instruments.

The main comparisons use the common empirical window from 1 September 2016 to 2 June 2025. This window covers the negative-rate period, the COVID episode, the post-2022 structural break, the hiking cycle, and the first phase of easing. Daily calibration results are therefore read as time series of fitted parameters, residuals, convergence states, boundary activity, diagnostics, and computation times.

The comparison is based on stand-alone daily calibrations. No sequential penalty is used in the main empirical design. This keeps the cross-sectional information content of each trading day separate from any later smoothing or reporting overlay.

7.3.2 Sensitivity to the Absolute-Error Objective

The baseline objective remains RMSRE because cap prices vary strongly by maturity. To assess the effect of this modelling choice, the OIS 6M ATM-cap panel is recalibrated with RMSE instead of RMSRE, while the instrument set, curve family, parameter bounds, and local solver policy are kept fixed. The comparison is a controlled objective-function sensitivity check for this panel.

Table 7.4: Full-window sensitivity of the OIS 6M ATM-cap calibration to replacing RMSRE by RMSE.

Metric	RMSRE solution	RMSE solution
Median RMSRE	0.66%	1.57%
Median RMSE (price bp)	8.207	5.191
Share ρ_{xy} near -1	69.59%	62.08%
Share $\rho_{xy} < -0.99$	86.46%	80.02%

Note: The comparison uses 2157 common dates from 1 September 2016 to 2 June 2025, with 13 instruments per date. Both columns use the same OIS 6M ATM-cap panel and the same local solver and bound policy.

Table 7.4 shows the expected scale trade-off. The RMSE solution improves the absolute-error criterion, while the RMSRE solution improves the relative-error criterion. The objective choice also changes the fitted parameter coordinates materially, especially in the volatility and mean-reversion coordinates. Near-boundary negative correlation remains visible under both objectives, although it is less frequent under RMSE. The objective function is therefore part of the calibration design. For the main diagnostic chain, RMSRE is chosen as the scale-normalised baseline

objective, and the RMSE comparison is used as an objective-function sensitivity check for the OIS 6M ATM-cap panel.

7.3.3 Curve, Subset, Solver, and Parameter-Box Axes

The empirical comparison varies only a small number of controlled axes. The first axis is the initial curve. The OIS 6M curve is the central market-consistent curve for the main comparison. The ECB AAA curve denotes the ECB Nelson–Siegel–Svensson curve for AAA-rated euro-area government bonds, while the BBK curve denotes the Bundesbank Nelson–Siegel–Svensson curve for German government bonds. These two curves therefore provide actuarially and institutionally relevant robustness axes.

The second axis is the instrument universe. For caps, the relevant comparison is between cap subsets that differ by tenor and maturity coverage. For swaptions, the comparison is between explicitly defined option-maturity and swap-tenor subsets. This distinction is necessary because caps are one-dimensional maturity instruments, whereas swaptions form a two-dimensional maturity-tenor surface. The instrument-set axis therefore tests whether a conclusion is a property of the G2++ model under the chosen market environment or mainly a property of the particular calibration portfolio.

The third axis is the optimiser. BOBYQA is used as the baseline solver for the final comparison because it is derivative-free, locally model-based, and well suited to bound-constrained smooth objectives when reliable analytical derivatives are not available for every pricing map. SBPLX is used as a targeted robustness solver, and Nelder–Mead remains a secondary optional check. Solver comparisons are interpreted cautiously. A different optimiser can expose flat valleys or local numerical irregularities, but it does not by itself prove that one economic parameter regime is more correct than another.

All solver-based evidence in this chapter is conditional on the daily calibration solution. The optimisers are local or derivative-free numerical procedures, and flat valleys or local minima are part of the calibration problem. The existing solver, objective, and subset comparisons are empirical safeguards against obvious numerical artefacts, but they are not exhaustive global-optimisation evidence.

The fourth axis is the parameter box. The empirical comparison uses the wide cap-consistent bounds

$$a_x, a_y \in [10^{-3}, 10], \qquad \sigma_x, \sigma_y \in [10^{-4}, 1], \qquad \rho_{xy} \in [-1, 1].$$

These bounds do not force economically convenient parameter values. They avoid narrow artificial upper bounds on the mean-reversion speeds and volatilities while preserving the mathematically necessary correlation interval. Numerical pricing routines may use internal safeguards near $\rho_{xy} = \pm 1$ when this is necessary for stable evaluation, but such safeguards do not change the reported calibrated value of ρ_{xy}.

The fifth axis is the model class. G2++ is the model carried forward into the calibration under $\mathbb{P}$. The one-factor Hull–White model enters only as a benchmark. It answers the narrower question of what is gained, and what diagnostic burden is added, when the calibration moves from a one-factor Gaussian short-rate model to a two-factor Gaussian short-rate model.

The diagnostic logic is not tied to G2++ in principle, but its implementation is model-specific. It requires a pricing map, a loss function, and a reliable local sensitivity matrix. G2++ is therefore the worked-through case in which these objects are available in enough detail to study calibration quality beyond fit.

7.3.4 Caps as the Analytical Diagnostic Benchmark

Caps are the analytical diagnostic benchmark of this chapter. In the G2++ model, cap prices admit a closed-form caplet decomposition in terms of integrated bond-option variances, and the cap-price Jacobian with respect to the five risk-neutral parameters can be derived analytically. The cap calibration therefore provides a clean weighted Jacobian for leverage, influence, EDoF, Fisher-information-type uncertainty, boundary-aware interpretation, and residual diagnostics.

This benchmark also fixes the reference point for the later instrument-class comparison. The swaption analysis asks whether the main qualitative cap findings, especially the near-boundary behaviour of ρ_{xy}, remain visible when the calibration instruments are replaced by a richer maturity-tenor surface. Identical parameter paths are not expected; the comparison concerns the persistence of the qualitative calibration regime.

Table 7.5 reports the cap configurations needed for the main reading. The OIS variants fit the ATM-cap panel closely across the common 2016–2025 window. The corresponding BBK variants fit less tightly, especially in the actuarial subset and all-caps specifications. This supports the use of OIS as the main market-consistent benchmark, while BBK is used as a historical-curve comparison.

The near-boundary correlation regime is visible across the reported cap designs. Across OIS and BBK, the median correlation remains close to -1, and a large share of dates lies below -0.99. The curve choice affects fit quality and the severity of boundary contact, but the broader boundary pattern remains present. This cap-side evidence is the reference level for the swaption comparison below.

Table 7.5: Selected cap calibration results on the common 2016–2025 comparison window. OIS provides the market-consistent benchmark, while BBK gives the corresponding historical-curve comparison.

Curve	Cap set	Role	Dates	Median RMSRE	Median ρ_{xy}	Share $\rho_{xy} < -0.99$
OIS 6M	6M caps	Main analytical benchmark	2074	0.63%	-1.0000	87.08%
OIS 6M	6M caps, $\leq$ 20Y	Maturity-restricted check	2100	0.42%	-1.0000	84.52%
OIS 6M	All caps, $\geq$ 2Y	Actuarial subset	1984	0.98%	-1.0000	86.69%
OIS 6M	All caps	Broad cap universe	1969	1.32%	-1.0000	83.60%
BBK	6M caps	Historical-curve comparison	2061	1.20%	-0.9999	80.16%
BBK	6M caps, $\leq$ 20Y	Maturity-restricted comparison	2087	0.78%	-0.9996	72.69%
BBK	All caps, $\geq$ 2Y	Actuarial historical-curve subset	1972	2.21%	-1.0000	92.09%
BBK	All caps	Broad historical-curve universe	1957	2.64%	-0.9997	67.40%

7.3.5 Swaptions as an Instrument-Class Extension

Swaptions enter as an instrument-class extension with explicit numerical controls. They are standard calibration instruments in practice and carry information that caps do not contain. A swaption combines an option maturity with the tenor of the underlying forward-starting swap. The market object is therefore a two-dimensional maturity-tenor surface rather than a one-dimensional cap maturity strip. This makes swaptions a useful comparison case for the cap-based G2++ parameter regime.

The need for numerical governance comes from the pricing map. The G2++ swaption formula is semi-analytical: it involves numerical integration and an implicit exercise-boundary calculation. The swaption Jacobian is therefore obtained by finite differences, whereas the cap Jacobian is available analytically. The resulting leverage, EDoF, influence, and related diagnostics are read as a numerical extension of the cap diagnostics, conditional on the quality of the finite-difference Jacobian. In the empirical study, this numerical layer is controlled through a fixed instrument

panel, common parameter bounds, explicit solver settings, finite-difference step choices, and step-stability checks.

The swaption study is kept to a small number of representative instrument universes, summarised in Table 7.6. The full swaption universe contains all 196 possible option-maturity and swap-tenor pairs in the vendor panel before datewise availability and pricing filters. It serves as the maximal information benchmark. The liquid total-maturity subset is the broad liquid subset with option maturity plus swap tenor at most 20 years. The central robust subset is the main swaption specification; it keeps the same total-maturity restriction but excludes one-year option maturities and one-year swap tenors. A representative grid with option maturities and swap tenors in $\{2, 5, 10, 15\}$ then checks how much of the surface information remains on a reduced but structured support.

Table 7.6: Swaption instrument universes used in the instrument-class comparison.

Instrument universe	Selection rule	Role in the comparison
Full swaption universe	All 196 possible option-maturity and swap-tenor pairs in the vendor panel, before datewise availability and pricing filters.	Maximal swaption-information benchmark; used as a computational and diagnostic stress test for the full quoted surface.
Liquid total-maturity subset	Option maturity plus swap tenor is at most 20 years.	Broad liquid swaption panel used to test whether the cap-side parameter regime survives a richer maturity-tenor surface.
Central robust subset	Same total-maturity restriction, excluding one-year option maturities and one-year swap tenors.	Main swaption specification; focuses on the liquid central region and reduces short-end sensitivity.
Representative grid	Option maturities and swap tenors restricted to $\{2, 5, 10, 15\}$.	Reduced structured support used to assess how much diagnostic information remains on a smaller surface grid.

7.3.6 The One-Factor Benchmark

The one-factor Hull–White comparison has a narrower role. It is not an attempt to reopen the model choice from first principles. It is a benchmark. It uses the same initial-curve logic, the same cap-price reconstruction logic, and the same RMSRE calibration criterion where possible, but restricts the stochastic structure to one Gaussian factor.

This benchmark is useful because it makes the benefit and cost of the second factor visible. If G2++ improves fit relative to Hull–White, the improvement has to be weighed against the additional local diagnostic burden: five parameters instead of two, a correlation parameter that can become boundary-active, and a richer but potentially more concentrated leverage structure. The detailed one-factor formulas and empirical comparison are therefore reserved for the separate Hull–White section below.

7.3.7 Scope of the Empirical Comparison

The empirical comparison is restricted to the curve choices, instrument subsets, solver variants, parameter boxes, and one-factor benchmark specified above. It tests whether the main cap-based conclusions are stable across those design axes and whether the swaption extension changes the severity of boundary activity in ρ_{xy}.

The results are comparative and sample-conditioned. The swaption side remains a numerical

instrument-class extension under explicit governance, while caps remain the analytical diagnostic benchmark. Any risk-neutral conclusion is then carried through the later calibration under $\mathbb{P}$ and the validation layer.

7.4 Caps versus Swaptions

The previous section fixed the empirical comparison protocol. This section turns to the instrument-class question itself. Caps and swaptions are both interest-rate derivatives, and both are used in market practice for risk-neutral calibration. The comparison determines what may be concluded once the cap-based diagnostic benchmark is confronted with swaption-based calibrations.

7.4.1 Why Compare Caps and Swaptions?

Caps and swaptions provide different views on the risk-neutral G2++ calibration. Caps are maturity instruments that decompose into caplets and mainly stress the model along the maturity direction of forward-rate periods. This makes them suitable for analytical diagnostics, but it also raises the question whether a cap-based parameter regime partly reflects the information structure of the cap portfolio.

Swaptions add a second instrument dimension. They are options on forward-starting swaps and combine option maturity with swap tenor, so their quotes form a maturity-tenor surface. They are therefore a natural instrument class for testing whether the cap-based correlation pattern survives a richer volatility surface.

The empirical question is whether the G2++ correlation parameter ρ_{xy} remains close to its lower boundary under swaption calibration. If this behaviour disappears, the cap result is mainly instrument-specific. If exact boundary hits become less frequent but ρ_{xy} still remains close to -1, the severity of boundary activity depends on the instrument set, while the near-boundary regime remains part of the empirical calibration evidence.

7.4.2 Common Sample and Calibration Protocol

The cap-versus-swaption comparison uses the same basic daily calibration protocol as the empirical comparison above. Each date is treated as a single cross-sectional calibration problem. Market quotes are converted into price targets, the G2++ parameter vector is estimated by RMSRE minimisation, and the result is interpreted through fit, convergence, parameter regimes, and boundary behaviour. The comparison does not add a sequential penalty term, because the aim is to compare instrument-class information instead of producing smoothed production parameter paths.

Within this protocol, the distinction between benchmark quote reconstruction and structural model pricing remains operative. Chapters 3 and 6 fix the quote conventions and the translation into monetary targets, while Chapter 4 provides the G2++ pricing map fitted here. This convention control is particularly important for swaptions because their market quotes describe a maturity-tenor surface, not a caplet strip.

The comparison therefore keeps the quote-to-price reconstruction instrument-specific and applies the same structural G2++ pricing map only after the corresponding monetary targets have been fixed.

The curve axes are OIS 6M, ECB AAA, and BBK. OIS 6M provides the main market-consistent benchmark. ECB AAA and BBK test whether the qualitative conclusion is sensitive to the

initial-curve convention. The baseline solver is BOBYQA under the wide cap-consistent parameter box, supplemented by one targeted Nelder–Mead check on the central OIS no-1Y subset. The comparison is intentionally limited in scope. BOBYQA is used as the baseline solver. The analysis first compares the instrument subsets, then reports the main curve sensitivity, and finally checks solver sensitivity for the subset in which BOBYQA gives the most interior correlation estimates.

7.4.3 Swaption Instrument Subsets

The swaption subsets are fixed before the results are interpreted. The full swaption universe contains all 196 possible option-maturity and swap-tenor pairs in the vendor panel before date-wise filters and is used as the maximal information benchmark. The liquid total-maturity subset restricts option maturity plus swap tenor to at most 20 years. The central robust specification keeps this restriction but excludes one-year option maturities and one-year swap tenors. This leaves a liquid medium-term subset and excludes the most price-sensitive short-end instruments. A reduced $2, 5, 10, 15$ grid is added as a structured support check. ECB AAA and BBK provide the curve comparisons, and the Nelder–Mead run on the central OIS subset gives the targeted solver check.

7.4.4 Fit and Convergence

Table 7.7 compares the selected swaption designs on the same relative-price scale. The central OIS 6M robust subset without one-year instruments gives the best BOBYQA fit, with median RMSRE 3.43%. The reduced $2, 5, 10, 15$ grid remains close at 4.32%, and the broader OIS liquid subset reaches 5.74%. The OIS full-universe run is less favourable, with median RMSRE 10.94%. The full universe therefore does not improve the median RMSRE relative to the selected liquid OIS subsets. It adds instruments from regions that are more exposed to sparse quotes, liquidity differences, and numerical pricing stress.

The full-universe run covers 2,244 attempted OIS observation dates from 1 September 2016 to 2 June 2025, with 2,219 successful calibrations. The daily panel has median 154 instruments and median 153 valid prices after the availability and pricing filters. The failures are mainly missing OIS discount-factor dates rather than a separate empirical regime.

The curve comparison is also clear. ECB AAA fits less tightly than the central OIS subset, and BBK is visibly weaker with median RMSRE 12.28%. This is the swaption analogue of the cap result: BBK is useful as a historical-curve contrast, but it is not the best-fitting calibration curve in the market sample. The Nelder–Mead check reaches almost the same fit as BOBYQA on the central OIS subset, so the fit ranking is not driven by one solver alone.

Table 7.7: Selected swaption calibration results. The reported rows cover the full OIS universe, the central OIS liquid specifications, a reduced grid, the main curve comparisons, and the targeted solver check.

Curve	Swaption set	Solver	Role	Median RMSRE	Median ρ_{xy}	Share $\rho_{xy} < -0.99$
OIS 6M	Full panel	BOBYQA	Full-universe benchmark	10.94%	-0.9788	44.89%
OIS 6M	≤ 20	BOBYQA	Liquid benchmark	5.74%	-0.9901	50.23%
OIS 6M	≤ 20, no 1Y	BOBYQA	Central robust specification	3.43%	-0.9528	33.00%
OIS 6M	Grid 2/5/10/15	BOBYQA	Reduced support grid	4.32%	-0.9875	46.60%
OIS 6M	≤ 20, no 1Y	Nelder–Mead	Solver check	3.42%	-0.9951	59.50%
ECB AAA	≤ 20, no 1Y	BOBYQA	Curve comparison	6.73%	-0.9911	50.36%
BBK	≤ 20, no 1Y	BOBYQA	Historical-curve comparison	12.28%	-0.8758	14.20%

7.4.5 Parameter Regimes

The next question concerns parameter regimes. The two instrument classes observe different parts of the interest-rate volatility structure. The parameter comparison therefore focuses on qualitative differences rather than pointwise agreement between cap-based and swaption-based parameters.

The five G2++ parameters have to be read jointly. Differences in a_x, a_y, σ_x, and σ_y may compensate each other and may interact with the calibrated correlation. A lower RMSRE under one instrument class does not automatically imply a more stable or more economically meaningful parameter path. Parameter regimes are therefore read together with convergence and boundary information, not in isolation.

In the swaption evidence set, the decisive regime distinction appears in the correlation parameter. Across the OIS swaption specifications, the calibrated correlations remain strongly negative. The central OIS 6M robust subset without one-year instruments is the most interior BOBYQA case, with median $\rho_{xy} = -0.9528$, but this is still close to perfect negative correlation. The broader OIS subset, the reduced grid, and the full OIS universe move closer to the lower boundary, with median correlations of -0.9901, -0.9875, and -0.9788, respectively. The full-universe run is therefore more interior than the liquid total-maturity subset but less interior than the central robust subset. Its remaining median parameter values in the raw full-universe output are approximately $a_x = 0.0698$, $a_y = 0.1318$, $\sigma_x = 0.0091$, and $\sigma_y = 0.0226$. The BBK comparison moves more visibly into the interior, but this comes with clearly weaker fit. The targeted Nelder–Mead check changes the solver reading: on the same OIS no-1Y subset, the median fit remains essentially unchanged, while the median correlation moves back to -0.9951.

The swaption evidence therefore reduces exact boundary contact in some specifications, but it does not remove the near-perfect negative correlation.

7.4.6 Boundary Behaviour of the Correlation Parameter

The key diagnostic object in the cap-versus-swaption comparison is the correlation parameter ρ_{xy}. The cap diagnostics show frequent lower-bound and near-boundary behaviour. The swaption comparison tests whether this pattern remains visible when the model is fitted to a richer swaption surface.

The analysis distinguishes exact lower-bound contact from a broader near-boundary range. Exact contact refers to values numerically indistinguishable from $\rho_{xy} = -1$ under the stated tolerance. The near-boundary range also includes values close to, but not exactly at, the lower boundary.

Table 7.7 and Figure 7.20 show that the severity of the correlation-boundary behaviour depends on the swaption design. Within the BOBYQA baseline, the strongest moderation occurs for the central OIS 6M robust subset without one-year instruments, where the share of dates with $\rho_{xy} < -0.99$ falls to 33.00%. The broader OIS subset, the reduced grid, and the ECB comparison remain closer to the lower boundary. The full OIS universe lies between these readings, with median $\rho_{xy} = -0.9788$, exact lower-bound contact on 19.47% of successful dates, and $\rho_{xy} < -0.99$ on 44.89% of successful dates. BBK is more interior in ρ_{xy}, but its fit is weaker. The Nelder–Mead check gives: on the same OIS no-1Y subset, the fit remains essentially unchanged, while the near-boundary share rises to 59.50%. Swaption calibration therefore reduces exact boundary contact in selected designs, but it does not remove the strongly negative correlation pattern.

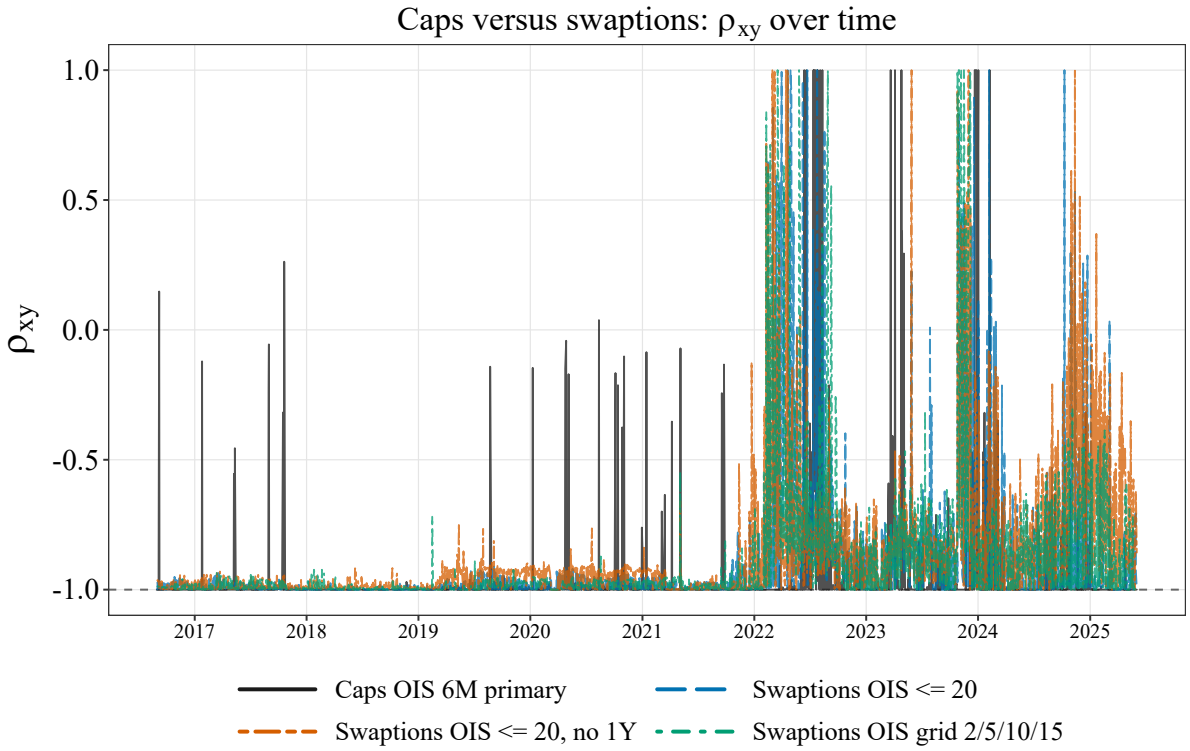


Figure 7.20: Cap-versus-swaption comparison of ρ_{xy} over time for the main OIS 6M BOBYQA swaption configurations.

7.4.7 Interpretation and Empirical Scope

The comparison is conditional on the investigated sample, curve conventions, swaption subsets, solver choices, and parameter box. It assesses whether swaption calibration reduces exact lower-bound contact in ρ_{xy} and whether strongly negative correlations remain visible.

Within the BOBYQA baseline, the OIS no-1Y subset is the most interior selected OIS swaption specification. The broader OIS subset, the reduced grid, the ECB AAA comparison, and the full OIS universe remain closer to the lower correlation boundary. The full universe adds the largest quoted surface considered here, but it does not overturn the hierarchy: its median RMSRE is weaker than the selected liquid OIS subsets, and ρ_{xy} remains close to the lower boundary on a substantial share of successful dates.

The targeted Nelder–Mead check is the main solver qualification. On the same OIS no-1Y subset, the fit remains essentially unchanged, while the correlation moves back towards the lower boundary. Swaption calibration therefore moderates exact lower-bound contact in selected designs, but it does not remove the strongly negative correlation pattern. Caps remain the analytical diagnostic reference; swaptions provide the numerical instrument-class extension under documented data filters and solver controls.

7.4.8 Representative Cross-Instrument Transfer Check

The preceding comparison studied own-instrument fit, parameter regimes, and boundary behaviour. This subsection adds a selected transfer check. Parameter vectors estimated on one liquid instrument class are used to price related instruments that were not used in that calibration. The two directions are cap-to-swaption and swaption-to-cap. In both directions, the fixed-strike cap/floor layer is added as a non-ATM target family.

The date design is representative and limited by common target availability. The ATM cap–swaption panel contains 144 dates from 5 September 2016 to 2 June 2025. The fixed-strike cap/floor layer is shorter and contributes 99 dates from 2 January 2018 to 2 January 2024. The sampling rule keeps the first common date in each month, the last common date in each quarter, and event-near dates only if the common target panel contains an observation within seven calendar days of the requested event. This avoids substituting a distant proxy for a missing stress date. In the common panel used here, the event additions that pass this rule are the COVID stress date (requested 18 March 2020, selected 18 March 2020), the ECB first hike (requested 21 July 2022, selected 20 July 2022), and the ECB first cut (requested 6 June 2024, selected 5 June 2024).

Table 7.8: Representative date design for the selected overlap panels used in the cross-instrument transfer check.

Panel	Window	Dates	Rule
ATM cap–swaption overlap panel	2016-09-05 to 2025-06-02	144	Monthly first common date, quarterly last common date, event dates only within seven days
Fixed-strike target overlap	2018-01-02 to 2024-01-02	99	Same representative dates, restricted to fixed-strike cap/floor availability

Table 7.9 gives the main pricing-error summary. The own-instrument median daily RMSREs are 0.64% for the ATM cap calibration on ATM caps and 3.58% for the swaption calibration on ATM swaptions. The cross-direction errors are significantly larger: swaptions priced with cap parameters have median daily RMSRE 15.65%, while ATM caps priced with swaption parameters have median daily RMSRE 9.59%. On the fixed-strike cap/floor target, the corresponding median daily RMSREs are 54.98% for the cap source and 101.13% for the swaption source. These errors measure out-of-sample transfer across instrument classes, since the target instruments are priced without recalibrating the G2++ parameters to that target family.

Table 7.9: Representative cross-instrument transfer check summary. Errors are relative to the observed target prices and aggregated datewise before the daily RMSRE columns are summarized.

Source	Target	Dates	Obs	Median daily RMSRE	Q90 daily RMSRE	Median abs. rel. error	Q90 abs. rel. error
ATM Caps calibration	ATM Caps	144	1728	0.64%	4.00%	0.43%	2.58%
ATM Swaptions calibration	ATM Caps	144	1728	9.59%	38.13%	5.14%	34.77%
ATM Caps calibration	ATM Swaptions	144	10143	15.65%	3463.97%	11.36%	100.00%
ATM Swaptions calibration	ATM Swaptions	144	10143	3.58%	6.21%	2.14%	7.24%
ATM Caps calibration	Fixed-strike caps/floors	99	3444	54.98%	93.15%	30.76%	99.54%
ATM Swaptions calibration	Fixed-strike caps/floors	99	3444	101.13%	659.26%	33.18%	120.06%

Figure 7.21 shows the daily RMSRE distributions. The violin-boxplot form reports both location and dispersion over representative dates. A calibration source that prices its own target well can still produce a broad and shifted error distribution on another target family.

The datewise view in Figure 7.22 adds the time dimension. The cross-instrument pricing error level is not constant over time, and the post-2022 period is visibly relevant for the transfer check. This matters for the later risk-premium and validation chapters, which inherit dated risk-neutral calibration inputs. The figure is based on the selected overlap panel and is not an exhaustive robustness panel over all available dates and instrument variants.

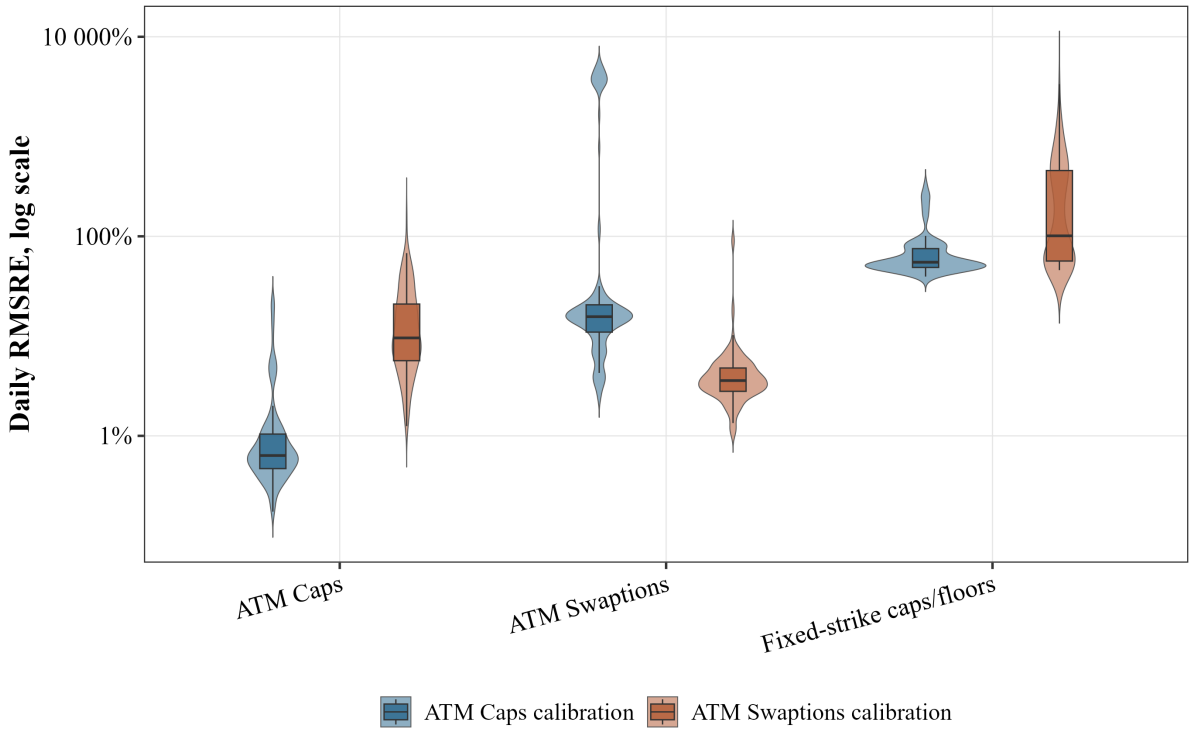


Figure 7.21: Daily RMSRE distributions in the representative cross-instrument transfer check. Violin layers show the distribution over representative dates on the overlap panels used in the comparison; embedded boxplots show the median and interquartile range.

The fixed-strike target extends the transfer check beyond ATM instruments. It tests how parameter vectors calibrated to ATM caps or swaptions price strikes away from the ATM level, including cap and floor sides. Figure 7.23 reports the median and interquartile range of absolute relative errors by maturity and strike index. The fixed-strike layer is read as an additional diagnostic stress test under the same source-calibration, date-sampling, and reporting protocol as the cap–swaption comparison.

The transfer check supports a bounded reading. Own-instrument fit does not imply comparable pricing accuracy on related target families. The chosen ATM cap–swaption panel and the shorter fixed-strike panel show target-dependent transfer errors, especially outside the calibration instrument class. This is relevant as a diagnostic stress test, but exact pricing of the full fixed-strike cap/floor surface is not the primary application target in the later PIA- and PRIIPs-related contexts. There, the calibrated interest-rate model is mainly used as a governed input

Table 7.10: Pairwise comparison of calibration sources on each target family within the selected overlap panels. The ratio is the datewise RMSRE of the swaption-calibrated source divided by the cap-calibrated source.

Target	Comparable dates	Cap source better	Swaption source better	Median ratio swpt/cap	Interpretation
ATM Caps	144	99.31%	0.69%	12.466	Cap source lower median RMSRE
ATM Swaptions	144	0.00%	100.00%	0.249	Swaption source lower median RMSRE
Fixed-strike caps/floors	99	88.89%	11.11%	1.786	Cap source lower median RMSRE



Figure 7.22: Datewise daily RMSRE in the representative cross-instrument transfer check. Points are shown on a logarithmic scale and faceted by target family.

for scenario generation, product classification, and risk reporting rather than as a stand-alone pricing engine for all out-of-the-money floor quotes. Caps remain the analytical reference for local diagnostics; the cross-instrument exercise documents the additional pricing layer on selected overlap panels.

7.5 The One-Factor Benchmark: Hull–White versus G2++

The preceding sections developed the risk-neutral calibration and diagnostic analysis for G2++. This is the model carried forward into the calibration under $\mathbb{P}$ and the actuarial application. The empirical findings above raise a one-factor benchmark question: whether the cap prices that lead G2++ to a near-boundary correlation regime can already be explained sufficiently well by an ordinary one-factor Hull–White model.

Hull–White is therefore used as a restricted benchmark, not as an alternative main model. It uses the same initial-curve convention, market price vector, and RMSRE objective where possible, but reduces the local tangent space from five G2++ directions to two Hull–White directions. The comparison asks whether the improved flexibility of G2++ justifies the additional

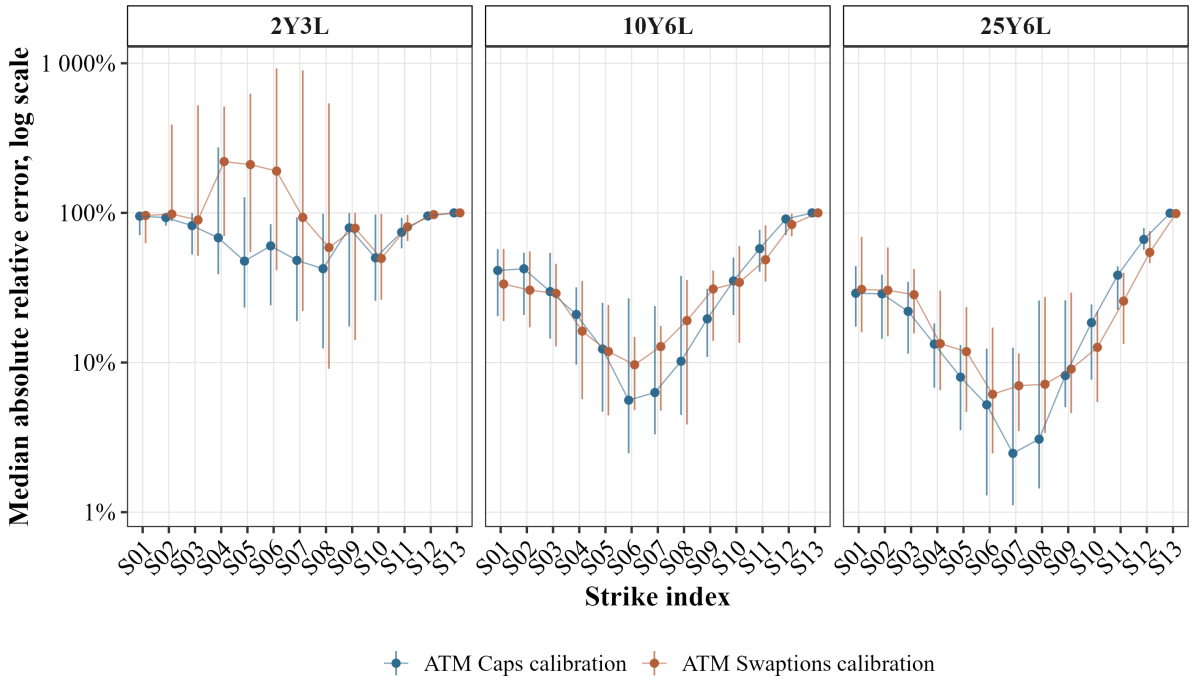


Figure 7.23: Fixed-strike cap/floor transfer errors on the shorter overlap panel. Points show the median absolute relative error across representative dates; vertical ranges show the interquartile range.

diagnostic burden: boundary-active correlation, possible rank loss, concentrated leverage, and larger parameter uncertainty. The cap side carries the empirical benchmark comparison, while the Jamshidian route records the analytical swaption contrast.

7.5.1 The Hull–White Model as a One-Factor Benchmark

The one-factor Hull–White model is the time-inhomogeneous Gaussian extension of the Vasicek model introduced by [62]. Under the risk-neutral measure, the short rate is described by

$$\mathrm{d}r_t = \big(\vartheta(t) - a r_t\big)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \tag{7.20}$$

where $a > 0$ is the mean-reversion speed and $\sigma > 0$ is the short-rate volatility. The deterministic function $\vartheta(t)$ is chosen so that the model reproduces the initial market term structure. The initial curve therefore acts as the deterministic anchor that makes the one-factor Gaussian dynamics compatible with the observed zero-coupon bond curve.

In the notation of Chapter 6, the standard choice of ϑ can be written in terms of the initial instantaneous forward curve $f^M(0,t)$ as

$$\vartheta(t) = \frac{\partial f^M(0,t)}{\partial t} + a f^M(0,t) + \frac{\sigma^2}{2a}\big(1 - e^{-2at}\big). \tag{7.21}$$

With this specification, zero-coupon bond prices retain the affine form

$$P(t,T) = A(t,T)\exp\big(-B_a(t,T)r_t\big), \tag{7.22}$$

where

$$B_a(t,T) = \frac{1 - e^{-a(T-t)}}{a}. \tag{7.23}$$

The exact expression for $A(t,T)$ depends only on the initial curve and the model parameters; see [16, 62]. For the comparison in this chapter, the relevant point is that the stochastic part of $P(t,T)$ is driven by only one Gaussian factor. This gives Hull–White its analytical tractability, but it also imposes a strong restriction on the local shape of the cap-price surface.

For calibration, the parameter vector is

$$\theta^{\mathsf{HW}} = (a,\sigma)^\top,$$

whereas the G2++ parameter vector is

$$\Pi = (a_x, a_y, \sigma_x, \sigma_y, \rho_{xy})^\top.$$

Both models can be calibrated with the same RMSRE objective and the same market price vector. Hence the comparison is not a comparison of two different loss functions. It is a comparison of two local pricing formulas under the same empirical design.

7.5.2 Hull–White Cap and Floor Prices

The cap pricing formula can be written in the same bond-option notation used for G2++. Consider a caplet with rate-setting date t_{i-1}, i.e. the date on which the underlying floating rate for the period is determined, payment date t_i, accrual period $\delta_i = t_i - t_{i-1}$, strike K, and notional N. Define

$$L_i = \log\left(\frac{P(0,t_{i-1})}{(1+K\delta_i)P(0,t_i)}\right). \tag{7.24}$$

As before, write

$$C(\alpha,u,v) = \frac{1-e^{-\alpha(v-u)}}{\alpha}. \tag{7.25}$$

The integrated Hull–White bond-option variance for the caplet period $[t_{i-1},t_i]$ is

$$\Sigma^2_{\mathsf{HW},i} = \sigma^2 C(a,t_{i-1},t_i)^2 C(2a,0,t_{i-1}). \tag{7.26}$$

Set $\Sigma_{\mathsf{HW},i} = \sqrt{\Sigma^2_{\mathsf{HW},i}}$ and

$$h^{\mathsf{HW}}_{+,i} = \frac{L_i}{\Sigma_{\mathsf{HW},i}} + \frac{\Sigma_{\mathsf{HW},i}}{2}, \qquad h^{\mathsf{HW}}_{-,i} = \frac{L_i}{\Sigma_{\mathsf{HW},i}} - \frac{\Sigma_{\mathsf{HW},i}}{2}. \tag{7.27}$$

Then the caplet price can be written as

$$\mathsf{Caplet}^{\mathsf{HW}}_i = N\left[P(0,t_{i-1})\Phi\left(h^{\mathsf{HW}}_{+,i}\right) - (1+K\delta_i)P(0,t_i)\Phi\left(h^{\mathsf{HW}}_{-,i}\right)\right]. \tag{7.28}$$

This is the one-factor analogue of the G2++ caplet expression used above. The initial curve, strike, accrual periods and discount factors are treated as fixed calibration inputs. The parameter dependence enters through $\Sigma^2_{\mathsf{HW},i}$. The price of a cap is obtained by summing the caplet prices over the caplet dates.

Floors do not add a new local diagnostic geometry under the same curve and strike convention. The caplet–floorlet parity relation is

$$\mathsf{Caplet}^{\mathsf{HW}}_i - \mathsf{Floorlet}^{\mathsf{HW}}_i = N\left[P(0,t_{i-1}) - (1+K\delta_i)P(0,t_i)\right]. \tag{7.29}$$

The right-hand side is fixed once the initial curve, strike, accrual period, and payment dates are fixed. It does not depend on the Hull–White stochastic parameters (a,σ). Consequently,

$$\frac{\partial\,\mathsf{Caplet}^{\mathsf{HW}}_i}{\partial p} = \frac{\partial\,\mathsf{Floorlet}^{\mathsf{HW}}_i}{\partial p}, \qquad p \in \{a,\sigma\}. \tag{7.30}$$

The same argument applies at the cap and floor level by summing over the corresponding caplets and floorlets. Hence a separate floor calibration would not create a different local WLS geometry if the same schedule, strike convention, and initial curve are used. This is why the one-factor benchmark can be developed on caps without losing the corresponding local information for floors.

7.5.3 Analytical Hull–White Cap Jacobian

The parameter dependence of the Hull–White caplet price enters through $\Sigma^2_{\mathsf{HW},i}$. Therefore the same chain-rule structure used in the G2++ cap diagnostics applies. For $p \in \{a, \sigma\}$,

$$\frac{\partial \,\mathsf{Caplet}_i^{\mathsf{HW}}}{\partial p} = \mathcal{M}_i^{\mathsf{HW}} \frac{\partial \Sigma^2_{\mathsf{HW},i}}{\partial p}, \tag{7.31}$$

with

$$\mathcal{M}_i^{\mathsf{HW}} = \frac{NP(0,t_{i-1})\varphi\big(h^{\mathsf{HW}}_{+,i}\big)}{2\Sigma_{\mathsf{HW},i}}. \tag{7.32}$$

For the mean-reversion derivatives, define

$$B_i(a) = C(a, t_{i-1}, t_i), \qquad G_i(a) = C(2a, 0, t_{i-1}). \tag{7.33}$$

The derivative with respect to the volatility parameter is immediate:

$$\frac{\partial \Sigma^2_{\mathsf{HW},i}}{\partial \sigma} = 2\sigma B_i(a)^2 G_i(a). \tag{7.34}$$

For the mean-reversion parameter,

$$\frac{\partial \Sigma^2_{\mathsf{HW},i}}{\partial a} = \sigma^2 \left(2B_i(a)G_i(a)\frac{\partial B_i(a)}{\partial a} + B_i(a)^2 \frac{\partial G_i(a)}{\partial a} \right), \tag{7.35}$$

where, for $\tau = v - u$,

$$\frac{\partial C(a,u,v)}{\partial a} = \frac{e^{-a\tau}(a\tau + 1) - 1}{a^2}, \tag{7.36}$$

and hence, for $T = t_{i-1}$,

$$\frac{\partial C(2a,0,T)}{\partial a} = \frac{e^{-2aT}(2aT + 1) - 1}{2a^2}. \tag{7.37}$$

For a cap k consisting of caplets $i \in \mathcal{I}_k$, the Hull–White Jacobian row is therefore

$$J^{\mathsf{HW}}_{k,p} = \frac{\partial \,\mathsf{Cap}_k^{\mathsf{HW}}}{\partial p} = \sum_{i \in \mathcal{I}_k} \mathcal{M}_i^{\mathsf{HW}} \frac{\partial \Sigma^2_{\mathsf{HW},i}}{\partial p}, \qquad p \in \{a, \sigma\}. \tag{7.38}$$

This formula is the central diagnostic bridge. It shows that no new diagnostic theory is needed for the one-factor case. Once the G2++ Jacobian is replaced by the two-column matrix J^{HW}, the same local WLS machinery applies.

7.5.4 Diagnostic Geometry of the One-Factor Benchmark

Let J^{HW} denote the cap-price Jacobian in (7.38). Using the same market prices y_k^{Mkt} and the same RMSRE weights

$$w_k = \frac{1}{(y_k^{\mathsf{Mkt}})^2},$$

the weighted design matrix is

$$X^{\mathrm{HW}} = W^{1/2} J^{\mathrm{HW}}.$$

The corresponding hat matrix is

$$H_w^{\mathrm{HW}} = W^{1/2} J^{\mathrm{HW}} \big((J^{\mathrm{HW}})^\top W J^{\mathrm{HW}}\big)^{+} (J^{\mathrm{HW}})^\top W^{1/2}. \tag{7.39}$$

Leverage, effective degrees of freedom, influence and covariance are then defined exactly as before:

$$h_{kk}^{\mathrm{HW}} = (H_w^{\mathrm{HW}})_{kk}, \tag{7.40}$$

$$\mathrm{EDoF}^{\mathrm{HW}} = \mathrm{tr}(H_w^{\mathrm{HW}}), \tag{7.41}$$

$$\mathrm{IF}_k^{\mathrm{HW}} = \big(y_k^{\mathrm{Mkt}} - g_k^{\mathrm{HW}}(\hat\theta^{\mathrm{HW}})\big)\big((J^{\mathrm{HW}})^\top W J^{\mathrm{HW}}\big)^{+} (J_k^{\mathrm{HW}})^\top w_k, \tag{7.42}$$

$$\widehat{\mathrm{Cov}}(\hat\theta^{\mathrm{HW}}) \approx \hat\sigma_{\mathrm{res}}^2 \big((J^{\mathrm{HW}})^\top W J^{\mathrm{HW}}\big)^{+}. \tag{7.43}$$

The formulas are identical in structure to the G2++ diagnostics. Their interpretation is not identical. The covariance approximation has the same Fisher-information-type form as in G2++, but in Hull–White it refers only to the two local directions a and σ.

In the Hull–White model, the rank of $W^{1/2} J^{\mathrm{HW}}$ is at most two. Locally, model prices can move only through the mean-reversion direction a and the volatility direction σ. If Hull–White fits the observed cap prices well, then these two directions are sufficient for this calibration problem. If the residuals remain large or show a clear maturity pattern, the one-factor variance surface is too rigid for the market data.

G2++ adds a second mean-reversion speed, a second volatility scale, and the correlation coordinate ρ_{xy}. This gives the model more ways to shape caplet variances across maturities, but it also introduces additional diagnostic issues: near-collinear parameter directions, active-boundary effects, reduced effective dimensionality, and concentrated leverage. The one-factor benchmark therefore asks whether the improved G2++ fit justifies this additional local diagnostic burden.

7.5.5 Endpoint Leverage in the Hull–White Cap Benchmark

The Hull–White benchmark explains why a U-shaped leverage profile can already arise in a one-factor cap calibration. Section 7.2 interpreted leverage as a measure of reconstruction difficulty: a maturity has high leverage when its local price-sensitivity row is hard to reproduce from the remaining rows. The exact identity behind this interpretation is recorded in Appendix D. The Hull–White case gives a simpler setting in which the mechanism can be seen explicitly. Its cap Jacobian has only two columns, yet it already produces an endpoint-leverage structure. This shows that endpoint leverage is already built into the cumulative cap design of the one-factor benchmark.

The reference geometry is the leverage formula for a simple-regression design. In an unweighted regression with intercept and one predictor,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad z_i = \begin{pmatrix} 1 \\ x_i \end{pmatrix},$$

the hat matrix projects the response vector onto the column space of the design matrix. Its diagonal entries are the leverage scores. Hoaglin and Welsch [60, Sec. 3, Eq. (3.1)] give

$$h_i = \frac{1}{N_{\mathrm{obs}}} + \frac{(x_i - \bar x)^2}{\sum_{k=1}^{N_{\mathrm{obs}}} (x_k - \bar x)^2},$$

where N_{obs} is the number of observations in the simple-regression design. Leverage therefore increases with the squared distance of the predictor value from the design centre. Points near the centre of the predictor range are easier to replace by neighbouring observations. Points near the endpoints of the predictor range contribute more strongly to the fitted slope and tend to have higher leverage. This is the regression-diagnostic reference used below. The local weighted cap Jacobian in the Hull–White benchmark is algebraically equivalent to the weighted analogue of this $(1, x)$-design geometry. This formulation embeds the nonlinear calibration within the classical regression-diagnostic framework of [60, 8] adapted to the nonlinear least-squares setting of [85].

The word design is used here in a local diagnostic sense. The cap maturities are fixed market instruments. Around the calibrated parameter vector, the nonlinear pricing function is read through its weighted Jacobian $W^{1/2}J$. The leverage calculation is therefore based on the local WLS projection geometry; compare the least-squares matrix background in [53]. Methodologically, this is the standard local-linear view of nonlinear least squares. Around the calibrated parameter vector, the nonlinear pricing map is represented by its Jacobian, as in the nonlinear-regression treatments of [85]. The resulting weighted Jacobian is then read as the local WLS design matrix. Its leverage scores are the diagonal entries of the associated weighted hat matrix, following the regression-diagnostic treatment of [60, 8].

For the six-month cap benchmark, the accrual tenor is fixed at $\delta = 0.5$. Write $T = t_{i-1}$ for the rate-setting date of the caplet period $[t_{i-1}, t_i]$. The Hull–White caplet variance can then be written as

$$\Sigma^2_{\text{HW}}(T) = A\big(1 - e^{-\alpha T}\big), \qquad \alpha = 2a, \qquad A = \frac{\sigma^2 C(a, 0, \delta)^2}{2a}. \tag{7.44}$$

This is a local reparametrisation of the one-factor Hull–White caplet variance. The parameter α determines a, and A then determines σ. The representation changes only the coordinates, not the variance entering the standard Hull–White caplet or bond-option formula; compare the one-factor short-rate-model treatment in Brigo and Mercurio [16, Ch. 3].

For $a > 0$, $\sigma > 0$, and fixed $\delta > 0$, this coordinate change is invertible. The inverse is given by $a = \alpha/2$ and

$$\sigma = \frac{\sqrt{\alpha A}}{C(\alpha/2, 0, \delta)}.$$

The reparametrisation changes the coordinates of the two-column Jacobian, but not its local column space. Since the weighted hat matrix depends on this column space, the leverage values are unchanged by this coordinate choice.

The two local parameter directions are the two partial derivatives of the scalar variance function in (7.44). Define the caplet variance-sensitivity vector

$$\Gamma_{A,\alpha}(T) := \begin{pmatrix} \partial_A \Sigma^2_{\text{HW}}(T) \\ \partial_\alpha \Sigma^2_{\text{HW}}(T) \end{pmatrix}.$$

Differentiation gives

$$\Gamma_{A,\alpha}(T) = \begin{pmatrix} 1 - e^{-\alpha T} \\ A T e^{-\alpha T} \end{pmatrix}. \tag{7.45}$$

In the local coordinates (A, α), combining these variance-sensitivity vectors with the caplet derivative factorisation in (7.31) gives the Hull–White cap Jacobian row for cap maturity index n:

$$j_n^{\text{HW}} = \sum_{i \in \mathcal{I}_n} \mathcal{M}^{\text{HW}}_{n,i} \Gamma_{A,\alpha}(t_{i-1}), \qquad \mathcal{M}^{\text{HW}}_{n,i} > 0. \tag{7.46}$$

The positive factors $\mathcal{M}_{n,i}^{\mathrm{HW}}$ are the caplet-specific price-to-variance derivatives. Thus a cap row is a positive cumulative combination of the caplet-level variance-sensitivity vectors belonging to that cap. Hence

$$j_n^{\mathrm{HW}} = \begin{pmatrix} U_n^{\mathrm{HW}} \\ AV_n^{\mathrm{HW}} \end{pmatrix}, \tag{7.47}$$

with

$$U_n^{\mathrm{HW}} = \sum_{i \in \mathcal{I}_n} \mathcal{M}_{n,i}^{\mathrm{HW}} \big(1 - e^{-\alpha t_{i-1}}\big), \qquad V_n^{\mathrm{HW}} = \sum_{i \in \mathcal{I}_n} \mathcal{M}_{n,i}^{\mathrm{HW}} t_{i-1} e^{-\alpha t_{i-1}}.$$

For the cap maturities used here, $U_n^{\mathrm{HW}} > 0$, so the first component can be used as a row-normalising scale. The weighted row entering the hat matrix (7.39) is

$$x_n^{\mathrm{HW}} = \sqrt{w_n} U_n^{\mathrm{HW}} \begin{pmatrix} 1 \\ A q_n^{\mathrm{HW}} \end{pmatrix}, \qquad q_n^{\mathrm{HW}} = \frac{V_n^{\mathrm{HW}}}{U_n^{\mathrm{HW}}}.$$

The factor $\sqrt{w_n} U_n^{\mathrm{HW}}$ is a positive row scale and enters as the row weight

$$\omega_n^{\mathrm{HW}} = w_n \big(U_n^{\mathrm{HW}}\big)^2.$$

The factor $A > 0$ only rescales the second column of the design matrix. This rescaling does not change the column space and therefore does not change the hat-matrix diagonal. For the leverage calculation, the Hull–White cap row can therefore be read as

$$z_n = \begin{pmatrix} 1 \\ q_n^{\mathrm{HW}} \end{pmatrix}.$$

The local one-factor Hull–White WLS geometry is therefore the weighted analogue of a $(1, x)$-design with predictor coordinate q_n^{HW}.

Let $Z = (z_1^\top, \dots, z_m^\top)^\top$ and $\Omega = \mathsf{diag}(\omega_1^{\mathrm{HW}}, \dots, \omega_m^{\mathrm{HW}})$. If the two columns are locally independent, the weighted hat matrix is

$$H^{\mathrm{HW}} = \Omega^{1/2} Z \big(Z^\top \Omega Z\big)^{-1} Z^\top \Omega^{1/2}.$$

Applying the same hat-matrix calculation as in the simple-regression leverage formula of Hoaglin and Welsch [60, Sec. 3, Eq. (3.1)], now with WLS weights ω_n^{HW}, gives

$$h_n^{\mathrm{HW}} = \omega_n^{\mathrm{HW}} \left[\frac{1}{S_0^{\mathrm{HW}}} + \frac{\big(q_n^{\mathrm{HW}} - \bar{q}_\omega^{\mathrm{HW}}\big)^2}{S_2^{\mathrm{HW}}} \right], \tag{7.48}$$

where

$$S_0^{\mathrm{HW}} = \sum_{\ell=1}^{m} \omega_\ell^{\mathrm{HW}}, \qquad \bar{q}_\omega^{\mathrm{HW}} = \frac{\sum_{\ell=1}^{m} \omega_\ell^{\mathrm{HW}} q_\ell^{\mathrm{HW}}}{S_0^{\mathrm{HW}}}, \qquad S_2^{\mathrm{HW}} = \sum_{\ell=1}^{m} \omega_\ell^{\mathrm{HW}} \big(q_\ell^{\mathrm{HW}} - \bar{q}_\omega^{\mathrm{HW}}\big)^2.$$

Formula (7.48) shows the two sources of one-factor leverage. A cap maturity can have high Hull–White leverage because its row weight ω_n^{HW} is large, or because its predictor coordinate q_n^{HW} is far from the weighted design centre $\bar{q}_\omega^{\mathrm{HW}}$. In the reconstruction interpretation of Appendix D, this means that an endpoint row is harder to rebuild from the remaining local sensitivity rows when the remaining maturities mainly provide interpolation information. The endpoint effect is therefore a property of the weighted local design.

It remains to identify the maturity coordinate behind q_n^{HW}. For a single caplet, remove the positive column scale A from (7.45). The relevant variance-sensitivity vector is then

$$\begin{pmatrix} 1 - e^{-\alpha T} \\ T e^{-\alpha T} \end{pmatrix}.$$

For $T > 0$, its first component is positive. Dividing by this first component gives the one-dimensional coordinate

$$\xi_\alpha(T) = \frac{Te^{-\alpha T}}{1 - e^{-\alpha T}} = \frac{T}{e^{\alpha T} - 1}.$$

Thus $\xi_\alpha(T)$ is the caplet-level ratio of the two Hull–White variance sensitivities. It is the coordinate that enters the simple-regression representation of the local cap design.

The cap-level predictor q_n^{HW} is the corresponding positive weighted average of these caplet-level coordinates. Since

$$t_{i-1}e^{-\alpha t_{i-1}} = \big(1 - e^{-\alpha t_{i-1}}\big)\xi_\alpha(t_{i-1}),$$

one obtains

$$q_n^{\mathsf{HW}} = \frac{\sum_{i\in\mathcal{I}_n} \mathcal{M}_{n,i}^{\mathsf{HW}}\big(1 - e^{-\alpha t_{i-1}}\big)\xi_\alpha(t_{i-1})}{\sum_{i\in\mathcal{I}_n} \mathcal{M}_{n,i}^{\mathsf{HW}}\big(1 - e^{-\alpha t_{i-1}}\big)}.$$

Thus q_n^{HW} is the cap-level average of the one-dimensional coordinates carried by the caplets inside the cap. This average still contains the strike-specific price-to-variance factors $\mathcal{M}_{n,i}^{\mathsf{HW}}$. Therefore the realised WLS rows are not forced to follow the monotone ordering of the caplet-level coordinate ξ_α.

To isolate the model-specific maturity geometry, separate the strike-specific price-to-variance factors and consider the cumulative variance-sensitivity benchmark

$$z_n(\alpha) = \sum_{i\in\mathcal{I}_n} \widetilde{\Gamma}_\alpha(t_{i-1}), \qquad \widetilde{\Gamma}_\alpha(T) = \begin{pmatrix} 1 - e^{-\alpha T} \\ Te^{-\alpha T} \end{pmatrix}. \tag{7.49}$$

Writing

$$z_n(\alpha) = \begin{pmatrix} \bar{U}_n(\alpha) \\ \bar{V}_n(\alpha) \end{pmatrix} = \bar{U}_n(\alpha) \begin{pmatrix} 1 \\ \bar{q}_n(\alpha) \end{pmatrix}, \qquad \bar{q}_n(\alpha) = \frac{\bar{V}_n(\alpha)}{\bar{U}_n(\alpha)},$$

the corresponding caplet-level coordinate is the ratio of the second component to the first component,

$$\xi_\alpha(T) = \frac{Te^{-\alpha T}}{1 - e^{-\alpha T}} = \frac{T}{e^{\alpha T} - 1}. \tag{7.50}$$

The coordinate ξ_α is strictly decreasing. With $x = \alpha T$,

$$\xi_\alpha(T) = \frac{1}{\alpha}\frac{x}{e^x - 1},$$

and

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x}{e^x - 1}\right) = \frac{e^x(1 - x) - 1}{(e^x - 1)^2} < 0, \qquad x > 0.$$

Later rate-setting dates therefore correspond to smaller caplet-level coordinates. A longer cap adds later rate-setting dates to the cumulative row and moves the row-normalised benchmark coordinate along the maturity grid. In this benchmark geometry, short and long caps form the endpoint design points, while intermediate caps are closer to interpolation positions.

This argument concerns the row-normalised benchmark, not the realised WLS row. The realised coordinate q_n^{HW} still contains strike-dependent price-to-variance factors and row weights. The benchmark therefore explains the endpoint mechanism after separating these effects. The empirical OIS 6M diagnostics then show that this endpoint geometry is realised in the maintained cap design.

The one-factor result also clarifies the difference to G2++. In Hull–White, the endpoint mechanism is explicit because the two-column Jacobian reduces to a weighted simple-regression design.

In G2++, the same WLS and reconstruction geometry applies, but the five-column Jacobian does not reduce to a one-dimensional predictor coordinate. Its boundary-dominated leverage is therefore assessed through the full weighted Jacobian and the empirical benchmark verification.

7.5.6 Hull–White Swaptions and the Jamshidian Route

Swaptions make the difference between one-factor and two-factor Gaussian models visible in a second way. In the one-factor Hull–White model, the value of the fixed-leg coupon bond at option maturity is a monotone function of the single short-rate state. This monotonicity is the key condition behind Jamshidian's decomposition; see [63, 65]. It allows a swaption to be represented as a portfolio of zero-coupon bond options.

More precisely, let

$$\mathcal{T} = \{t_1, \dots, t_n\}, \qquad T < t_1 < \cdots < t_n,$$

write $\delta_i = t_i - t_{i-1}$, and define the fixed-leg cash-flow coefficients by

$$c_i = \begin{cases} K\delta_i, & i = 1, \dots, n-1, \\ 1 + K\delta_n, & i = n. \end{cases}$$

Let r^* be the unique short rate at option maturity T for which the coupon bond is at par,

$$\sum_{i=1}^{n} c_i P(T, t_i; r^*) = 1. \tag{7.51}$$

Set $X_i = P(T, t_i; r^*)$. Then Jamshidian's decomposition gives the Hull–White swaption price in the form

$$\mathsf{Swaption}^{\mathsf{HW}}(0, T, \mathcal{T}, N, K, \chi) = N \sum_{i=1}^{n} c_i \; \mathsf{ZBO}^{\mathsf{HW}}\big(0, T, t_i, X_i, \omega\big), \qquad \omega = -\chi, \tag{7.52}$$

where $\chi = 1$ denotes a payer swaption and $\chi = -1$ a receiver swaption. With this sign convention, the payer swaption is represented by bond puts and the receiver swaption by bond calls. The Hull–White zero-coupon-bond option price can be written as

$$\mathsf{ZBO}^{\mathsf{HW}}(0, T, S, X, \omega) = \omega \begin{bmatrix} P(0,S)\Phi\big(\omega d(0;T,S,X)\big) \\ \quad - XP(0,T)\Phi\big(\omega(d(0;T,S,X) - v(0;T,S))\big) \end{bmatrix}, \tag{7.53}$$

with

$$\begin{aligned} v(0;T,S) &= \sigma B_a(T,S)\sqrt{\frac{1 - e^{-2aT}}{2a}}, \\ d(0;T,S,X) &= \frac{\log\big(P(0,S)/(XP(0,T))\big)}{v(0;T,S)} + \frac{v(0;T,S)}{2}. \end{aligned} \tag{7.54}$$

Thus, after one scalar root r^* has been found, the Hull–White swaption reduces to a finite sum of bond-option prices. This is the precise sense in which the one-factor swaption problem is analytically more accessible than the corresponding G2++ problem.

The same representation also gives a formal route to a Hull–White swaption Jacobian. For $p \in \{a, \sigma\}$, define

$$F(r, p) = \sum_{i=1}^{n} c_i P(T, t_i; r, p) - 1.$$

The boundary $r^* = r^*(p)$ is determined by $F(r^*(p), p) = 0$, and hence

$$\frac{\partial r^*}{\partial p} = -\frac{\partial F(r^*, p)/\partial p}{\partial F(r^*, p)/\partial r}, \tag{7.55}$$

whenever the denominator is non-zero. The strike $X_i = P(T, t_i; r^*(p), p)$ then satisfies

$$\frac{\partial X_i}{\partial p} = \frac{\partial P(T, t_i; r^*, p)}{\partial p} + \frac{\partial P(T, t_i; r^*, p)}{\partial r}\frac{\partial r^*}{\partial p}. \tag{7.56}$$

Differentiating (7.52) therefore yields the formal structure

$$\frac{\partial \,\mathsf{Swaption}^{\mathsf{HW}}}{\partial p} = N\sum_{i=1}^{n} c_i \left[\frac{\partial \,\mathsf{ZBO}_i^{\mathsf{HW}}}{\partial p} + \frac{\partial \,\mathsf{ZBO}_i^{\mathsf{HW}}}{\partial X_i}\frac{\partial X_i}{\partial p}\right], \qquad p \in \{a, \sigma\}. \tag{7.57}$$

The formula documents the analytical route. A full swaption-based Hull–White diagnostic study would require the same safeguards as any numerical swaption analysis: model-consistent price reconstruction, stable root solving, tolerance checks, derivative checks, and the same empirical sample design. The empirical one-factor-versus-two-factor benchmark in this chapter therefore remains cap-based.

In G2++, the corresponding coupon-bond value depends on two Gaussian factors and does not yield the same one-dimensional Jamshidian decomposition. The G2++ swaption price remains tractable through a semi-analytical representation with an implicit exercise boundary and a one-dimensional numerical integral; compare [64, 84, 82]. From the diagnostic perspective, this means that the cap comparison and the swaption comparison have different roles. Caps provide the analytical diagnostic benchmark; swaptions provide the instrument-class extension with explicit numerical controls discussed in the previous section.

7.5.7 Empirical Evidence from the Paired Cap Sample

The benchmark compares Hull–White and G2++ on the same OIS 6M cap calibration sample, using the same reconstructed market prices, the same RMSRE objective, and the same daily window from 1 September 2016 to 2 June 2025. It is not intended to exhaust every possible curve family or instrument subset. It asks, on a controlled full-window cap sample, whether the one-factor tangent space is sufficient or whether the additional G2++ directions carry material pricing information.

Across 2157 common calibration days, G2++ attains a lower RMSRE on 2141 days, that is, on 99.26% of the paired sample. The median ratio $\mathsf{RMSRE}^{\mathsf{G2++}}/\mathsf{RMSRE}^{\mathsf{HW}}$ is approximately 0.127. Hull–White achieves the lower aggregate RMSRE on only 16 days. Fifteen of these 16 cases are numerically negligible, with a median absolute RMSRE advantage of about $5.2 \cdot 10^{-9}$. One date, 11 September 2024, stands out as a genuine one-factor outlier. Figure 7.24 shows the full paired comparison.

The fit result is therefore clear on this sample. The one-factor model does not reproduce the cap price vector with the same accuracy as G2++. The improvement is visible on almost all paired dates and is large in relative terms. At the same time, the comparison does not say that G2++ is simply better in every respect. The diagnostic summaries show the cost of the additional flexibility.

To isolate the role of the lower correlation boundary, Figure 7.25 repeats the paired RMSRE comparison only for those dates on which the G2++ calibration places the correlation parameter at the lower boundary, using the numerical criterion $|\hat{\rho}_{xy} + 1| \leq 10^{-8}$. This restriction leaves

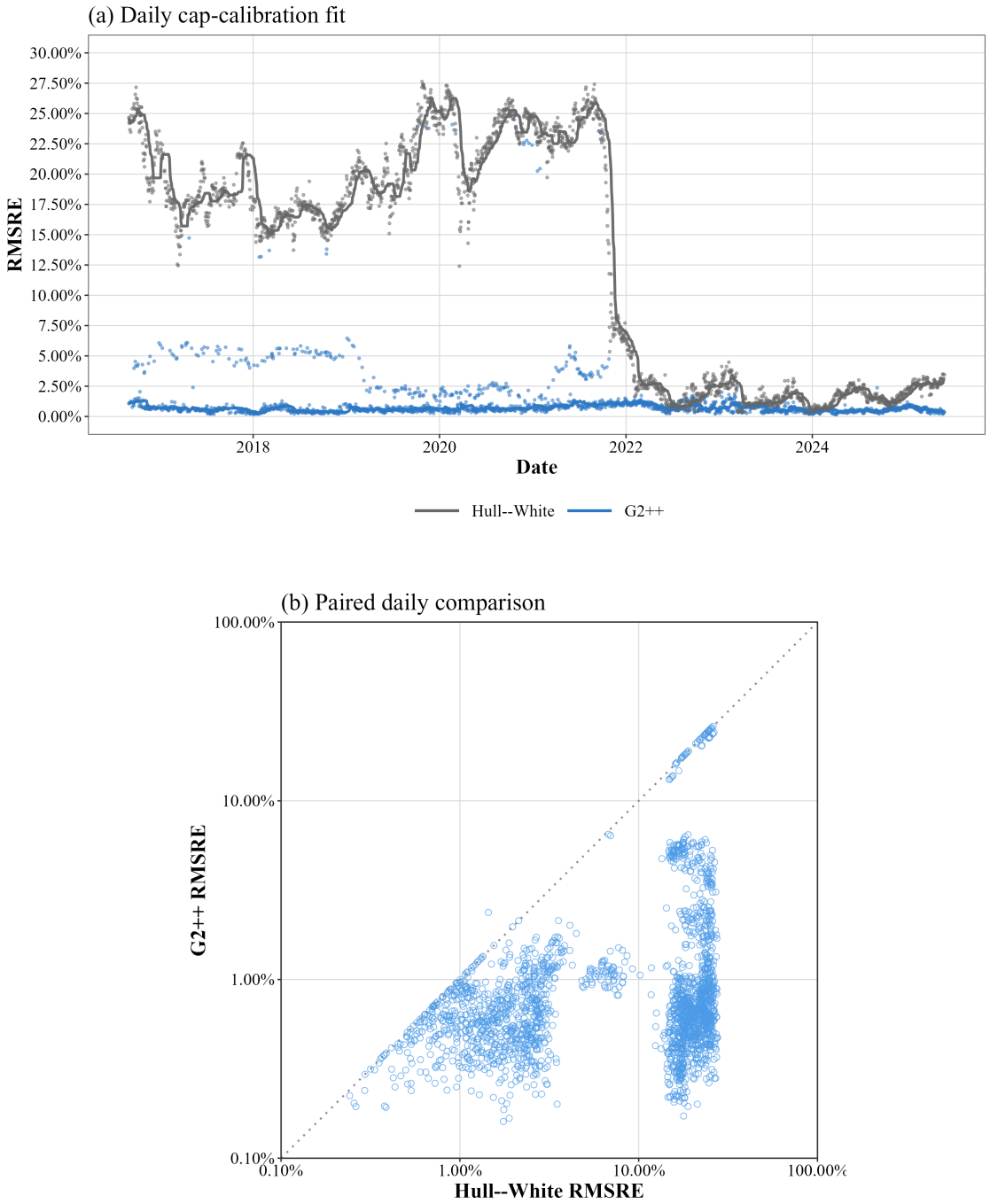


Figure 7.24: Daily paired RMSRE comparison for Hull–White and G2++ on the OIS 6M cap sample from 2016-09-01 to 2025-06-02. Panel (a) reports the aggregate daily RMSRE values, and Panel (b) compares the paired daily RMSRE pairs. Points below the diagonal indicate a lower G2++ RMSRE. The figure summarises this controlled sample comparison and does not by itself imply universal model dominance.

1432 paired calibration days, that is, 66.39% of the full OIS 6M cap sample. On this boundary subsample, G2++ attains a lower RMSRE on 1428 days, or 99.72% of the restricted sample, and the median ratio $\mathsf{RMSRE}^{\mathsf{G2++}}/\mathsf{RMSRE}^{\mathsf{HW}}$ is approximately 0.096. The same qualitative conclusion remains under the broader near-boundary cutoffs $\hat{\rho}_{xy} < -0.999$ and $\hat{\rho}_{xy} < -0.99$, for which the median ratios are approximately 0.098 and 0.095, respectively.

In the investigated OIS 6M cap sample, even the boundary-correlation subsample does not indicate an empirical reduction of G2++ to ordinary Hull–White. The lower correlation boundary reduces the effective Brownian rank, while the calibrated pricing map can still differ from the Hull–White pricing map. This is consistent with the two-kernel structure in Subsection 7.5.8: at $\rho_{xy} = -1$, G2++ has one Brownian driver, but unless $a_x = a_y$, its stochastic component still contains two distinct exponential mean-reversion kernels.

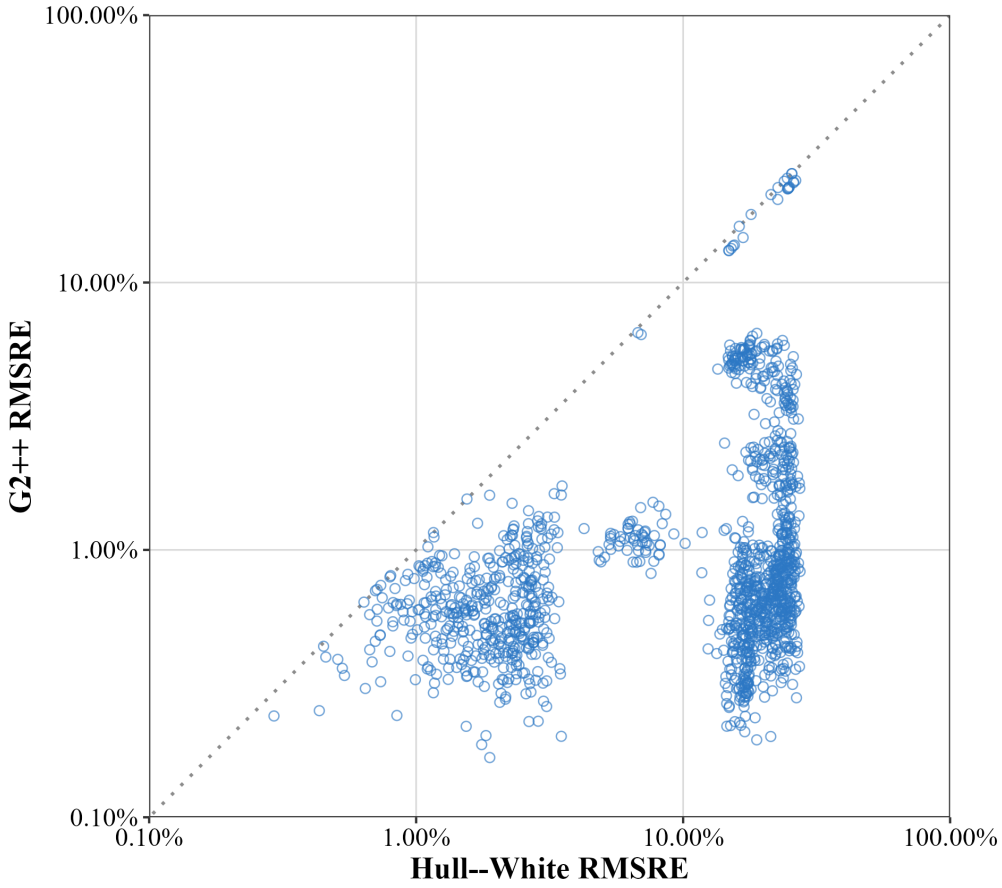


Figure 7.25: Daily paired RMSRE comparison for Hull–White and G2++ on the exact-boundary OIS 6M cap subsample defined by $|\hat{\rho}_{xy} + 1| \leq 10^{-8}$. Both axes are shown on logarithmic scales, and points below the diagonal indicate a lower G2++ RMSRE. The boundary-restricted comparison examines whether the G2++ pricing map empirically collapses to the ordinary Hull–White fit when the correlation parameter is at the lower boundary within this sample.

Table 7.11 condenses the full-sample and boundary-restricted RMSRE evidence in compact form.

The full-window comparison also makes the fit-versus-diagnostics trade-off visible at the level of daywise summary statistics. The mean effective degrees of freedom is approximately 4.33 for G2++ and 2.00 for Hull–White. This is consistent with the higher-dimensional tangent space of G2++. The average daily maximum leverage rises from about 0.403 in Hull–White to 0.927 in G2++, and the average daily maximum influence rises from about 0.0177 to 0.502. The additional G2++ directions reduce the objective value and concentrate local calibration power

Table 7.11: Boundary-restricted Hull–White versus G2++ RMSRE comparison on the paired OIS 6M cap sample. The table reports the number of paired dates, the share of the full sample, the number of dates with lower G2++ RMSRE, and the median ratio $\mathsf{RMSRE}^{\mathsf{G2++}}/\mathsf{RMSRE}^{\mathsf{HW}}$.

Subsample	Paired dates	Share of sample	G2++ lower RMSRE days/share	Median ratio
Full sample	2157	100.00%	2141 (99.26%)	0.127
Exact boundary	1432	66.39%	1428 (99.72%)	0.096
$\hat{\rho}_{xy} < -0.999$	1695	78.58%	1689 (99.65%)	0.098
$\hat{\rho}_{xy} < -0.99$	1865	86.46%	1858 (99.62%)	0.095

more strongly in particular instruments and dates.

Figure 7.26 isolates the one-factor diagnostic geometry. The upper panel confirms that the Hull–White calibration remains close to its two-dimensional local tangent space throughout the sample. The dashed reference line marks the full two-parameter dimension, in analogy to the G2++ EDoF plots. The lower violin-boxplots show the maturity-wise leverage and influence distributions. The Hull–White leverage profile has a clear endpoint structure: leverage is high at the short and long ends and substantially lower in the middle maturities. This pattern arises already in the RMSRE-weighted two-column cap Jacobian and is therefore distinct from the G2++ boundary-dominance pattern. Influence inherits this endpoint structure, while also depending on the realised residuals.

The endpoint pattern is also visible in the daily maxima. In Hull–White, the maximum leverage is attained at one of the two endpoint maturities on every day of the paired sample: on about 91.5% of days at the 30Y cap and on about 8.5% at the 3Y cap. The median leverage is about 0.424 at 30Y, 0.326 at 3Y, and only 0.054 at 10Y. Interior caps are therefore comparatively easier to reconstruct from neighbouring maturities, while the endpoints carry more local design information. In G2++, leverage is also concentrated near the maturity boundary, but with a different asymmetry: the maximum leverage falls on the 3Y cap on about 97.8% of days and on the 30Y cap on only about 2.1% of days. For influence, the Hull–White maxima are again endpoint-driven on about 99.3% of days, with about 82.2% led by the 3Y cap and about 17.1% by the 30Y cap. In G2++, endpoint-maximum influence occurs on about 42.0% of days. The influence profile therefore reflects both the local design geometry and the realised residual pattern.

The paired maturity-wise residual summaries show that the one-factor fit gap extends beyond a single short-end instrument. For Hull–White, the median absolute relative residual is about 40.5% at 3Y and 16.8% at 4Y, drops to about 1.9%–3.2% around 5Y–6Y, and then rises into a broad medium-to-long-maturity hump of about 9.2%–14.7% from 8Y to 20Y. G2++ compresses this profile into single-digit percentages, with median absolute relative residuals between about 0.6% and 1.5% across all maturities. The largest proportional improvement occurs at the short end, but the two-factor gain also removes the medium-to-long-maturity residual hump left by the one-factor variance surface.

Table 7.12 summarises this trade-off.

The benchmark therefore supports a cost–benefit interpretation. Hull–White is both a simpler model and diagnostically simpler. Its local geometry is easier to read because the model can move prices in only two parameter directions. G2++ fits the cap market much better on the paired OIS 6M sample, but this fit gain comes with a more demanding local diagnostic structure. The second factor is useful because it captures maturity-profile information that the one-factor

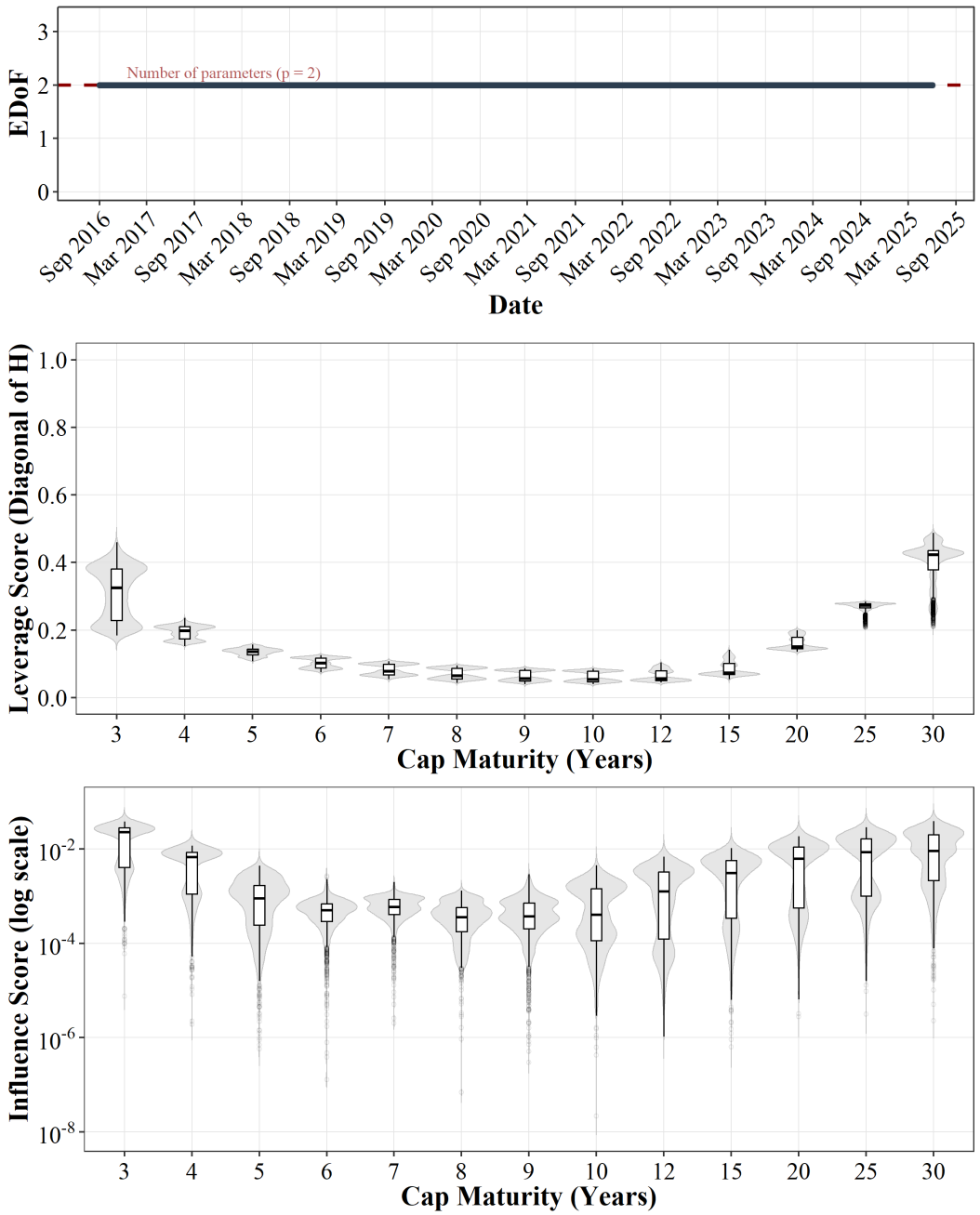


Figure 7.26: Diagnostic summaries for the Hull–White cap calibration on the paired OIS 6M sample from 2016-09-01 to 2025-06-02. The upper panel shows that the effective degrees of freedom stay equal to the two model parameters, indicating full local rank of the two-column Hull–White Jacobian. The middle panel reports leverage by cap maturity and shows the endpoint pattern of the one-factor cap design. The lower panel reports influence scores on a logarithmic scale; unlike leverage, influence also reflects the realised residuals.

Table 7.12: One-factor versus two-factor cap-calibration benchmark on the paired OIS 6M sample.

Object	Hull–White	G2++	Interpretation
Parameter dimension	2	5	G2++ has a richer tangent space.
Mean EDoF	2.00	4.33	Additional directions are empirically active.
Average max. leverage	0.403	0.927	G2++ concentrates local calibration power more strongly.
Average max. influence	0.0177	0.502	G2++ is locally more sensitive to selected observations.
Lower RMSRE days	16 of 2157	2141 of 2157	G2++ dominates fit on the paired sample.
Median RMSRE ratio	0.127		$\mathsf{RMSRE}^{\mathsf{G2++}}/\mathsf{RMSRE}^{\mathsf{HW}}$ with a large relative fit gain.

variance surface cannot reproduce, and it must be accompanied by leverage, influence, EDoF, covariance, and boundary diagnostics.

The rare Hull–White win dates do not overturn this conclusion. Their median RMSRE advantage is only about $5.2 \cdot 10^{-9}$, and the largest advantage, about $9.28 \cdot 10^{-3}$, occurs on 11 September 2024. Fifteen of the 16 Hull–White win dates are therefore practically numerical ties. This single date is about 2954 times larger than the second-largest Hull–White win and is therefore an outlier, not representative one-factor evidence. Even on that date the advantage is mainly a short-end event: the absolute relative residual is lower in Hull–White by about 0.644 at 3Y and 0.360 at 4Y, while G2++ remains slightly better at 30Y. The rare Hull–White wins therefore do not identify a general one-factor dominance. They identify a small set of atypical dates, of which only one is materially different from numerical near-ties.

The benchmark therefore supports a specific empirical conclusion. On the paired OIS 6M cap window, it establishes the one-factor-versus-two-factor fit difference, the leverage contrast, the maturity-wise residual structure, and the interpretation of the rare Hull–White win dates, without extending the claim to broader multi-curve or multi-subset comparisons.

Table 7.13 condenses the empirical comparison layer into one map of the chapter's main conclusions.

The table summarises the comparison layers without repeating the detailed numerical evidence from the preceding figures and tables. It records the question addressed by each layer, the conclusion carried forward, and the aspects left outside the empirical scope.

7.5.8 Why Boundary Correlation Does Not Imply Hull–White

The boundary-restricted comparison above raises a more precise theoretical question. The empirical G2++ cap calibrations often place the correlation parameter close to the lower boundary. This observation should not be read as a direct reduction to Hull–White. At the lower correlation boundary, G2++ has one Brownian driver. It does not, in general, have the dynamics or variance surface of an ordinary one-factor Hull–White model.

To see this, consider the stochastic part of the G2++ short rate under the risk-neutral measure,

$$r_t = x_t + y_t + \psi(t),$$

where

$$\mathrm{d}x_t = -a_x x_t \,\mathrm{d}t + \sigma_x \,\mathrm{d}W_t^x, \qquad \mathrm{d}y_t = -a_y y_t \,\mathrm{d}t + \sigma_y \,\mathrm{d}W_t^y,$$

Comparison layer	Question	Main evidence	Interpretation	Limitation
Analytical cap benchmark	Which instruments and directions dominate the local fit?	The analytical cap Jacobian reveals leverage concentration, reduced effective dimensionality on stressed dates, realised influence differences, and active-boundary effects even when RMSRE is small.	A low calibration error does not by itself imply a stable local geometry.	The full local diagnostic machinery is available only for the cap pricing map.
Swaption subset extension	Does a richer maturity-tenor surface change the fitted parameter regime?	Fit and boundary activity vary materially across the selected swaption subsets, especially between the OIS no-1Y core and the broader surfaces.	Instrument choice changes the severity of the observed regime.	Swaption diagnostics require finite-difference Jacobians and numerical-governance checks; they are therefore not analytically equivalent to the cap diagnostics.
Cap-versus-swaption boundary comparison	Is near-boundary correlation merely a cap artefact?	Swaptions reduce exact lower-bound contact under the BOBYQA baseline, but a clear near-boundary regime remains and the targeted Nelder–Mead check restores stronger boundary mass.	The phenomenon is not purely cap-specific.	The conclusion is empirical and conditioned on the sample, subset, solver, and parameter box.
Cross-instrument transfer check	Does the qualitative calibration reading remain plausible when parameters are transported to related derivative instruments on selected overlap panels?	On the representative OIS overlap panels, own-instrument errors remain much smaller than cross-direction errors; fixed-strike cap/floor pricing is visibly harder than ATM pricing.	The exercise documents bounded and target-dependent transfer behaviour in the sample, while instrument-class fit, local diagnostics, and transfer pricing remain separate evidence layers.	The exercise is a representative ATM cap–swaption and fixed-strike transfer check, not a full derivative-universe validation.
Hull–White benchmark	What is gained and what is paid for by the second factor?	On the paired OIS 6M cap sample, G2++ fits materially better and removes the broad one-factor residual hump, while also raising leverage, influence, and diagnostic burden.	The second factor adds explanatory power, not a free improvement.	This is a paired cap benchmark, not a universal model ranking.

Table 7.13: Synthesis of the Chapter-7 empirical comparison layer. The cap and swaption rows refer to the 2016-09-01 to 2025-06-02 evidence, with the shorter BBK swaption sample qualified in the main text; the cross-instrument row refers to the 144-date ATM overlap and the shorter 99-date fixed-strike overlap panels; the final row refers to the paired OIS 6M cap benchmark on the same long window.

and suppose for the moment that

$$\rho_{xy} = -1.$$

Then one may write, after choosing a common Brownian motion W,

$$\mathrm{d}W_t^x = \mathrm{d}W_t, \qquad \mathrm{d}W_t^y = -\mathrm{d}W_t.$$

Assuming zero initial stochastic factors for notational simplicity, the stochastic deviation from the deterministic shift is

$$z_t := x_t + y_t = \int_0^t \left[\sigma_x e^{-a_x(t-s)} - \sigma_y e^{-a_y(t-s)}\right] \mathrm{d}W_s. \tag{7.58}$$

Thus the stochastic part is indeed driven by only one Brownian motion, but its deterministic kernel is not generally a single exponential. It is the difference of two exponential kernels.

The distinction is structural. In the ordinary one-factor Hull–White model, the stochastic deviation from the fitted initial curve has the form

$$z_t^{\mathrm{HW}} = \int_0^t \sigma e^{-a(t-s)} \,\mathrm{d}W_s, \tag{7.59}$$

with one mean-reversion speed a and one volatility scale σ. Comparing (7.58) and (7.59) shows that equality would require

$$\sigma_x e^{-a_x u} - \sigma_y e^{-a_y u} = \sigma e^{-au} \qquad \text{for all } u \geq 0. \tag{7.60}$$

If $a_x \neq a_y$ and both volatility components are non-zero, this identity cannot hold. Indeed, exponentials with distinct exponents are linearly independent. Equivalently, the logarithmic derivative of the left-hand side of (7.60),

$$\frac{-a_x\sigma_x e^{-a_x u} + a_y\sigma_y e^{-a_y u}}{\sigma_x e^{-a_x u} - \sigma_y e^{-a_y u}},$$

is not constant in u in the generic case, whereas the logarithmic derivative of a single exponential kernel is the constant $-a$. Hence the rank-one boundary case of G2++ is not, in general, an ordinary Hull–White model.

The same conclusion can be read directly from the state dynamics. The sum $z_t = x_t + y_t$ satisfies

$$dz_t = -\big(a_x x_t + a_y y_t\big)\,\mathrm{d}t + (\sigma_x - \sigma_y)\,\mathrm{d}W_t. \tag{7.61}$$

Unless $a_x = a_y$, the drift term cannot be written as a function of z_t alone. The process therefore lacks the one-dimensional short-rate-factor form of ordinary Hull–White. The model has one Brownian driver at $\rho_{xy} = -1$, but its stochastic component is still generated by two mean-reversion speeds.

The same distinction extends to a rank-one n-factor construction. Let

$$r_t = \psi(t) + \sum_{k=1}^{n} z_{k,t}, \qquad \mathrm{d}z_{k,t} = -a_k z_{k,t}\,\mathrm{d}t + s_k\sigma_k\,\mathrm{d}W_t, \qquad s_k \in \{-1, 1\},$$

so that all stochastic factors are driven by the same Brownian motion and differ only by sign, volatility, and mean reversion. The Brownian rank of this model is one. Nevertheless, the stochastic short-rate deviation is

$$\sum_{k=1}^{n} z_{k,t} = \int_0^t \left(\sum_{k=1}^{n} s_k\sigma_k e^{-a_k(t-u)}\right) \mathrm{d}W_u. \tag{7.62}$$

Unless all active a_k coincide or all but one effective volatility component vanishes, the kernel in (7.62) is not a single exponential. Thus Brownian dimension and mean-reversion spectrum are different concepts. A rank-one Brownian driver can still generate several deterministic exponential decay components. The two-factor $\rho_{xy} = -1$ case is only the smallest non-trivial example of this distinction.

The distinction disappears in degenerate cases. If $a_x = a_y = a$, then

$$z_t = \int_0^t (\sigma_x - \sigma_y) e^{-a(t-s)} \, \mathrm{d}W_s,$$

which is a one-factor Hull–White stochastic component with mean-reversion speed a and effective volatility $|\sigma_x - \sigma_y|$. Similarly, if one volatility component vanishes, only one exponential kernel remains. These cases show that the ordinary Hull–White model is contained in the boundary part of the G2++ family as a special or limiting case. They do not imply that the generic boundary case is identical to Hull–White.

This distinction is also visible in the caplet-variance formulas. In the one-factor Hull–White model, the integrated bond-option variance has the form

$$\Sigma^2_{\mathrm{HW},i} = \sigma^2 C(a, t_{i-1}, t_i)^2 C(2a, 0, t_{i-1}), \tag{7.63}$$

whereas the G2++ variance at $\rho_{xy} = -1$ is

$$\begin{aligned} \Sigma^2_{-1,i} = {} & \sigma_x^2 C(a_x, t_{i-1}, t_i)^2 C(2a_x, 0, t_{i-1}) \\ & + \sigma_y^2 C(a_y, t_{i-1}, t_i)^2 C(2a_y, 0, t_{i-1}) \\ & - 2\sigma_x \sigma_y C(a_x, t_{i-1}, t_i) C(a_y, t_{i-1}, t_i) C(a_x + a_y, 0, t_{i-1}). \end{aligned} \tag{7.64}$$

If $a_x = a_y = a$, this expression reduces to

$$\Sigma^2_{-1,i} = (\sigma_x - \sigma_y)^2 C(a, t_{i-1}, t_i)^2 C(2a, 0, t_{i-1}),$$

which is the Hull–White variance shape. If $a_x \neq a_y$, however, the variance profile contains two different mean-reversion kernels and their cross term. It therefore belongs to a richer family of maturity profiles than (7.63).

This explains why a near-boundary G2++ correlation estimate is not, by itself, evidence that the two-factor model has become redundant. The boundary indicates that the Brownian factor structure is weakly identified by the cap portfolio. It does not imply that the model price surface has collapsed to the Hull–White surface. At the level of caplet variance shapes, Hull–White is embedded as a special case of the $\rho_{xy} = -1$ boundary family through the restriction $a_x = a_y$ and $|\sigma_x - \sigma_y| = \sigma$. Allowing $a_x \neq a_y$ adds maturity-profile flexibility through two mean-reversion components and their cross term.

The empirical superiority of G2++ in RMSRE is therefore not contradicted by near-boundary correlation estimates. It is consistent with the fact that the observed cap-price vector may contain maturity-shape information beyond the ordinary Hull–White variance family. The boundary behaviour is a diagnostic warning about identification and parameter stability, not evidence that the calibrated G2++ surface has reduced to Hull–White.

7.5.9 Conclusion of the One-Factor Benchmark

The one-factor comparison gives three conclusions. First, Hull–White is a transparent lower-dimensional benchmark. Its cap and floor Jacobians are analytically available, and the same local

WLS diagnostics apply. Second, the empirical fit comparison on the paired OIS 6M cap sample favours G2++. This also holds on the boundary-restricted subsample with $|\hat{\rho}_{xy} + 1| \leq 10^{-8}$. The result is consistent with the theoretical distinction above: $\rho_{xy} = -1$ reduces the Brownian rank, but the generic boundary G2++ model still contains two mean-reversion components and is not identical to ordinary Hull–White dynamics. Third, the improved fit comes with a higher diagnostic burden. The G2++ tangent space has five parameter directions instead of two, and its local leverage and influence structure can be more concentrated.

This benchmark supports the use of G2++ as the main model and justifies the diagnostic analysis developed above. A better RMSRE is insufficient as a stand-alone argument for the richer model. The fit gain has to be read together with local rank, leverage, influence, boundary activity, and parameter stability.

The subsequent calibration under $\mathbb{P}$ therefore starts from a diagnosed risk-neutral parameter history. The relevant information includes the fit advantage over Hull–White, the local sensitivity structure, the maturity-endpoint behaviour, and the boundary behaviour of the G2++ correlation parameter.

8 Calibration Step III: The Risk Premium

Having calibrated the G2++ model under the risk-neutral measure $\mathbb{Q}$, we now turn to the physical measure $\mathbb{P}$ used for real-world actuarial simulation. Risk-neutral dynamics are appropriate for market-consistent pricing, but they do not by themselves specify the drift structure required for projected real-world interest-rate scenarios.

We treat the shift from $\mathbb{Q}$ to $\mathbb{P}$ as a separate calibration problem. The risk-neutral G2++ parameters are kept fixed, and the real-world layer is introduced through deterministic factor-specific drift adjustments λ_t^x and λ_t^y. Once the forecast targets are specified, the calibration of these premium functions reduces to a linear identification problem.

The chapter proceeds in four steps. It first restates the real-world factor dynamics. It then discusses the available forecast families, derives constant, two-level, and n-level risk premium specifications together with their calibration systems, and finally interprets the resulting scenario families through their ergodic long-run laws. The chapter is connected to the term-premium and yield-curve-forecasting literature because the real-world calibration has to specify which expected future rates and which risk premia the scenario generator should encode [36, 71, 2, 22, 10, 31]. The detailed derivations of the real-world factor dynamics, the step-function integrals, and the expected spot-rate calibration equations are collected in Appendix B. The present chapter keeps the formulas needed for the calibration argument and their interpretation.

8.1 Constructing the risk premium

The G2++ short rate keeps the form

$$r_t = x_t + y_t + \psi_t,$$

where the deterministic shift ψ_t is fixed by the initial term structure as in Chapter 4. The change from $\mathbb{Q}$ to $\mathbb{P}$ is introduced through the drift of the two Gaussian factors, not through an additional independent short-rate factor. Under the physical measure, we write

$$\begin{aligned} \mathrm{d}x_t &= a_x\left(\lambda_t^x - x_t\right)\mathrm{d}t + \sigma_x \mathrm{d}W_t^{x,\mathbb{P}}, \qquad x_0 = 0, \\ \mathrm{d}y_t &= a_y\left(\lambda_t^y - y_t\right)\mathrm{d}t + \sigma_y \mathrm{d}W_t^{y,\mathbb{P}}, \qquad y_0 = 0. \end{aligned}$$

The deterministic functions λ_t^x and λ_t^y are the physical mean-reversion levels of the two Gaussian factors. Relative to the risk-neutral specification, they represent the real-world drift adjustment, while the risk-neutral parameters $\left(a_x, a_y, \sigma_x, \sigma_y, \rho_{xy}\right)$ remain fixed.

Taking expectations gives

$$\mathbb{E}_{\mathbb{P}}[r_t] = \psi_t + \int_0^t e^{-a_x(t-u)} a_x \lambda_u^x \,\mathrm{d}u + \int_0^t e^{-a_y(t-u)} a_y \lambda_u^y \,\mathrm{d}u.$$

The two mean-reversion levels enter through factors with different mean-reversion speeds. The faster factor has stronger short-horizon effects, while the more persistent factor affects the longer part of the simulated dynamics. The interpretation is factorwise. It should not be read as a split in which one parameter is a short-end premium and the other one is a long-end premium; forecast anchors generally depend on both factors.

8.1.1 Constant mean-reversion levels

The baseline specification assumes constant premium levels,

$$\lambda_t^x := d_x, \qquad \lambda_t^y := d_y, \qquad t \geq 0.$$

For the first factor,

$$\int_0^t e^{-a_x(t-u)} a_x d_x \, \mathrm{d}u = d_x \left(1 - e^{-a_x t}\right),$$

and the same calculation applies to the second factor. Hence,

$$\mathbb{E}_{\mathbb{P}}^{const}[r_t] = \psi_t + d_x(1 - e^{-a_x t}) + d_y(1 - e^{-a_y t}).$$

For the long-run analysis, assume that the market forward curve has a finite long-end limit. We define

$$\psi_\infty := \lim_{t\to\infty} \psi_t = f^M(0,\infty) + \frac{\sigma_x^2}{2a_x^2} + \frac{\sigma_y^2}{2a_y^2} + \rho_{xy}\frac{\sigma_x\sigma_y}{a_x a_y}.$$

Then

$$\mathbb{E}_{\mathbb{P}}^{const}[r_t] \xrightarrow{t\to\infty} \psi_\infty + d_x + d_y.$$

Under the constant specification, d_x and d_y determine the expected short-rate path at every horizon. For finite t, they enter through the adjustment

$$d_x(1 - e^{-a_x t}) + d_y(1 - e^{-a_y t}).$$

In the limit, they determine the long-run mean $\psi_\infty + d_x + d_y$. Thus the constant premium pair fixes both the finite-horizon adjustment and the long-run centre of the physical short-rate distribution. The constant specification therefore uses only two parameters, but it does not separate the finite-horizon drift adjustment from the long-run centre.

8.1.2 Step-function premium families

Empirical and regulatory forecasts rarely justify an unrestricted premium path. Step-function specifications provide an intermediate approach. They allow the premium to change over selected time intervals, while keeping the number of unknowns tied to the available forecast information. The two-level case is linked to the G2++ measure-change specifications studied in [11]; the general n-level family below is the finite-grid extension used in the present calibration framework. Each additional time block introduces two further premium levels. It should therefore be introduced only if the forecast side supplies additional independent identifying conditions.

Two-level premium family

The two-level specification distinguishes an initial level from a long-run level. For a switching time $\tau > 0$, set

$$\lambda_t^x = \mathbb{1}_{\{t\leq\tau\}} d_x + \mathbb{1}_{\{t>\tau\}} l_x,$$
$$\lambda_t^y = \mathbb{1}_{\{t\leq\tau\}} d_y + \mathbb{1}_{\{t>\tau\}} l_y.$$

The parameters (d_x, d_y) govern the first block and (l_x, l_y) govern the second block. Integrating the deterministic drift contribution gives, for the first factor,

$$\int_0^t e^{-a_x(t-u)} a_x \lambda_u^x \, \mathrm{d}u = d_x \left(e^{-a_x(t-\min\{t,\tau\})} - e^{-a_x t} \right) + l_x \left(1 - e^{-a_x(t-\min\{t,\tau\})} \right).$$

The expected short rate is therefore

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}^{step}[r_t] = \psi_t &+ d_x \left(e^{-a_x(t-\min\{t,\tau\})} - e^{-a_x t} \right) + l_x \left(1 - e^{-a_x(t-\min\{t,\tau\})} \right) \\ &+ d_y \left(e^{-a_y(t-\min\{t,\tau\})} - e^{-a_y t} \right) + l_y \left(1 - e^{-a_y(t-\min\{t,\tau\})} \right). \end{aligned}$$

For $t \to \infty$ and fixed τ, the first-block contributions vanish and

$$\mathbb{E}_{\mathbb{P}}^{step}[r_t] \xrightarrow{t\to\infty} \psi_\infty + l_x + l_y.$$

The two-level family separates these two roles. The parameters d_x and d_y determine the transition path up to the switching time τ, while l_x and l_y determine the long-run centre $\psi_\infty + l_x + l_y$. This is the smallest extension that separates the finite-horizon drift adjustment from the stationary mean.

General n-level premium family

The general step-function family uses a grid

$$0 = \tau_0 < \tau_1 < \cdots < \tau_{n-1} < \tau_n = \infty$$

and premium levels that are constant on each interval,

$$\begin{aligned} \lambda_t^x &= \sum_{i=1}^{n} \mathbb{1}_{\{\tau_{i-1} < t \le \tau_i\}} \lambda_{x,i}, \\ \lambda_t^y &= \sum_{i=1}^{n} \mathbb{1}_{\{\tau_{i-1} < t \le \tau_i\}} \lambda_{y,i}. \end{aligned}$$

Summing the blockwise contributions gives the closed form below; the derivation is given in Appendix B. For the first factor, the deterministic drift contribution can be written as

$$\int_0^t e^{-a_x(t-u)} a_x \lambda_u^x \, \mathrm{d}u = -\lambda_{x,1} e^{-a_x t} + \sum_{i=1}^{n-1} (\lambda_{x,i} - \lambda_{x,i+1}) e^{-a_x(t-\min\{t,\tau_i\})} + \lambda_{x,n}.$$

The expression for the second factor is analogous. Hence,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}^{steps}[r_t] = \psi_t &- \lambda_{x,1} e^{-a_x t} + \sum_{i=1}^{n-1} (\lambda_{x,i} - \lambda_{x,i+1}) e^{-a_x(t-\min\{t,\tau_i\})} + \lambda_{x,n} \\ &- \lambda_{y,1} e^{-a_y t} + \sum_{i=1}^{n-1} (\lambda_{y,i} - \lambda_{y,i+1}) e^{-a_y(t-\min\{t,\tau_i\})} + \lambda_{y,n}. \end{aligned}$$

As the exponential terms vanish,

$$\mathbb{E}_{\mathbb{P}}^{steps}[r_t] \xrightarrow{t\to\infty} \psi_\infty + \lambda_{x,n} + \lambda_{y,n}.$$

Under the general n-level specification, the last premium pair $(\lambda_{x,n}, \lambda_{y,n})$ determines the long-run mean. The earlier blocks determine how the model moves towards this long-run centre, but they vanish from the limiting mean as $t \to \infty$. Thus the n-level family separates transition dynamics from the long-run mean.

8.2 Forecast information and target construction

Physical-measure calibration is constrained by the availability of real-world interest-rate forecasts. While liquid derivatives provide a dense cross-section of market prices for risk-neutral calibration, forecast information under $\mathbb{P}$ is typically sparse. Real-world forecasts often consist of isolated single-point projections rather than continuous forecast curves. Institutions such as the OECD provide short-rate projections over limited horizons, while EIOPA specifies regulatory long-end quantities such as the Ultimate Forward Rate (UFR). These objects are useful anchors, but they do not by themselves provide the continuous target curves required to calibrate general step-function premium families.

To obtain such targets, we use today's forward-rate structure as the starting point. For the two maturities used in the calibration, namely $S \in \left\{\frac{1}{4}, 10\right\}$, representing durations of 3 months ($S = \frac{1}{4}$) and 10 years ($S = 10$), the observable forward-rate curves are extended from their last liquid point (LLP) to maturity-specific long-run levels. First, using the methodology employed by EIOPA, we calculate distinct UFRs for the 3-month and 10-year forward rates. Next, we interpolate today's observable forward-rate curves from their LLP to the respective UFR using natural splines. This yields two smooth forward-rate curves:

$$f(0, t, t+S), \qquad S \in \left\{\frac{1}{4}, 10\right\},$$

Here $f(0, t, t+S) := \frac{1}{S}\log\left(\frac{P(0,t)}{P(0,t+S)}\right)$ denotes the annualised continuously compounded forward rate over $[t, t+S]$ observed at time 0, which can be evaluated at the forecast horizons needed for the premium calibration.

The forward curve is not used as an unconditional forecast without adjustment. Empirical evidence rejects the pure Expectation Hypothesis, which assumes forward rates are unbiased predictors of future spot rates, and a constant term premium is generally too restrictive to capture observed rate dynamics. We therefore distinguish the observed forward rate from the expected future spot rate by writing:

$$f(0, t, t+S) = \mathbb{E}[i(t, t+S) \mid \mathcal{F}_0] + \pi(0, t, t+S), \tag{8.1}$$

where $\pi(0, t, t+S)$ denotes the time- and maturity-dependent term premium for horizon t and maturity S. With this sign convention, the forecast target for the expected future spot rate is:

$$\theta_S^F(t) := f(0, t, t+S) - \pi(0, t, t+S).$$

Equivalently, if the adjustment is defined with the opposite sign as a time- and duration-dependent forecast adjustment $\iota(t, S) = -\pi(0, t, t+S)$, the target can be written in the form:

$$\theta_S^F(t) := f(0, t, t+S) + \iota(t, S).$$

The adjusted forward-rate curves provide the forecast targets used in the linear systems for the premium levels. The target curve is fixed before the physical-measure calibration starts; it is not generated by the G2++ premium family itself.

The premium family then translates these targets into factor-specific physical mean-reversion levels. A constant premium requires two independent forecast conditions, a two-level family requires four, and an n-level family requires $2n$. The complexity of the premium family is therefore tied to the information content of the target construction. Additional levels are meaningful only if the forecast side supplies additional independent target information.

We distinguish three broad forecast families. The first uses isolated external anchors, such as regulatory long-run levels. The second uses forward-based curves adjusted by explicit term premia and is the primary family used in this chapter. The third uses filtered or model-based decompositions, such as macro-finance or Kalman-filter term-premium estimates. Richer forecast families can justify richer premium paths, but the additional flexibility must come from additional target information rather than from the premium specification alone.

8.3 Calibrating the risk premium

The numerical calibration matches model-implied expected spot rates to the selected forecast targets. Let $\theta^F_S(t)$ denote the target for the expected spot rate of maturity S at future time t. For fixed forecast targets and fixed risk-neutral G2++ parameters, the expected spot-rate equations are affine in the unknown premium levels. After the deterministic terms are moved to the right-hand side, the calibration problem becomes a linear system for these levels.

For maturity $S > 0$, define

$$R_x(S) := \frac{1 - e^{-a_x S}}{a_x}, \qquad R_y(S) := \frac{1 - e^{-a_y S}}{a_y}.$$

The model-implied expected spot rate can be written in the form

$$S\,\mathbb{E}_{\mathbb{P}}[i^S_t] = D_S(t) + R_x(S)\,\mathbb{E}_{\mathbb{P}}[x_t] + R_y(S)\,\mathbb{E}_{\mathbb{P}}[y_t],$$

where $D_S(t)$ collects the deterministic curve and variance terms. In particular,

$$D_S(t) = \int_t^{t+S} \psi_u \,\mathrm{d}u - \frac{V(S)}{2},$$

Equivalently, using the deterministic-shift convention of Chapter 4, this term can be written as

$$D_S(t) = \log\left(\frac{P^M(0,t)}{P^M(0,t+S)}\right) + \frac{V(t+S) - V(t) - V(S)}{2}.$$

This is the form used in the calibration right-hand sides below.

The precise form of $D_S(t)$ is known once the initial curve and the risk-neutral G2++ parameters have been fixed.

8.3.1 Calibration with constant mean-reversion levels

For the constant premium family, the two unknowns are d_x and d_y. Using two forecast anchors (t_1, S_1) and (t_2, S_2), the identifying equations are

$$\theta^F_{S_1}(t_1) = \mathbb{E}_{\mathbb{P}}[i^{S_1}_{t_1}], \qquad \theta^F_{S_2}(t_2) = \mathbb{E}_{\mathbb{P}}[i^{S_2}_{t_2}].$$

Since

$$\mathbb{E}_{\mathbb{P}}[x_t] = d_x(1 - e^{-a_x t}), \qquad \mathbb{E}_{\mathbb{P}}[y_t] = d_y(1 - e^{-a_y t}),$$

define

$$A_S(t) := R_x(S)(1 - e^{-a_x t}), \qquad B_S(t) := R_y(S)(1 - e^{-a_y t}).$$

The two constant physical mean-reversion levels solve

$$\begin{pmatrix} A_{S_1}(t_1) & B_{S_1}(t_1) \\ A_{S_2}(t_2) & B_{S_2}(t_2) \end{pmatrix} \begin{pmatrix} d_x \\ d_y \end{pmatrix} = \begin{pmatrix} S_1\theta^F_{S_1}(t_1) - D_{S_1}(t_1) \\ S_2\theta^F_{S_2}(t_2) - D_{S_2}(t_2) \end{pmatrix}.$$

Thus, in the constant case, the two forecast conditions determine d_x and d_y through a 2×2 linear system. A unique pair (d_x, d_y) exists if and only if the corresponding coefficient matrix is non-singular.

The determinant condition can be expressed directly in terms of the mean-reversion speeds and the selected forecast horizons. Let

$$q(u) := \frac{1 - \exp(-a_x u)}{1 - \exp(-a_y u)}, \qquad u > 0.$$

For one forecast anchor (t, S), the coefficients of d_x and d_y are

$$A_S(t) = R_x(S)(1 - \exp(-a_x t)), \qquad B_S(t) = R_y(S)(1 - \exp(-a_y t)).$$

The coefficient matrix is therefore

$$\begin{pmatrix} A_{S_1}(t_1) & B_{S_1}(t_1) \\ A_{S_2}(t_2) & B_{S_2}(t_2) \end{pmatrix}.$$

This matrix is singular exactly when its two rows are proportional, that is, when

$$\frac{A_{S_1}(t_1)}{B_{S_1}(t_1)} = \frac{A_{S_2}(t_2)}{B_{S_2}(t_2)}.$$

Using the definitions of R_x and R_y, the row ratio satisfies

$$\frac{A_S(t)}{B_S(t)} = \frac{a_y}{a_x} \frac{1 - \exp(-a_x S)}{1 - \exp(-a_y S)} \frac{1 - \exp(-a_x t)}{1 - \exp(-a_y t)} = \frac{a_y}{a_x} q(S) q(t).$$

The constant factor a_y/a_x cancels between the two rows. Hence the proportional-row condition is equivalently

$$q(S_1)q(t_1) = q(S_2)q(t_2).$$

This condition depends only on the risk-neutral mean-reversion speeds and the selected target horizons. It does not depend on the numerical forecast values themselves. If the coefficient matrix is singular, the two targets do not determine a unique allocation between d_x and d_y. If it is nearly singular, small changes in forecasts or curve inputs can produce large changes in the calibrated constant levels.

8.3.2 Calibration with two-level premium functions

For the two-level family, four premium levels have to be identified:

$$(d_x, l_x, d_y, l_y).$$

Define the block coefficients

$$H_a^d(t;\tau) := e^{-a(t - \min\{t,\tau\})} - e^{-at}, \qquad H_a^l(t;\tau) := 1 - e^{-a(t-\min\{t,\tau\})}.$$

For a forecast anchor (t, S), the corresponding x-factor coefficients are

$$A_S^d(t) := R_x(S) H_{a_x}^d(t;\tau), \qquad A_S^l(t) := R_x(S) H_{a_x}^l(t;\tau),$$

and analogously $B_S^d(t)$ and $B_S^l(t)$ for the y-factor. With four forecast anchors, the calibration system is

$$\begin{pmatrix} A_{S_1}^d(t_1) & A_{S_1}^l(t_1) & B_{S_1}^d(t_1) & B_{S_1}^l(t_1) \\ A_{S_2}^d(t_2) & A_{S_2}^l(t_2) & B_{S_2}^d(t_2) & B_{S_2}^l(t_2) \\ A_{S_3}^d(t_3) & A_{S_3}^l(t_3) & B_{S_3}^d(t_3) & B_{S_3}^l(t_3) \\ A_{S_4}^d(t_4) & A_{S_4}^l(t_4) & B_{S_4}^d(t_4) & B_{S_4}^l(t_4) \end{pmatrix} \begin{pmatrix} d_x \\ l_x \\ d_y \\ l_y \end{pmatrix} = \begin{pmatrix} S_1\theta_{S_1}^F(t_1) - D_{S_1}(t_1) \\ S_2\theta_{S_2}^F(t_2) - D_{S_2}(t_2) \\ S_3\theta_{S_3}^F(t_3) - D_{S_3}(t_3) \\ S_4\theta_{S_4}^F(t_4) - D_{S_4}(t_4) \end{pmatrix}.$$

The four levels are uniquely identified if and only if the coefficient matrix has rank four. The rank condition has a direct interpretation. The forecast anchors must span the early and late premium blocks. If all targets lie before the switching time τ, the late-block columns vanish. If the rows of the coefficient matrix are proportional in the relevant factor channels, the allocation between x- and y-risk premia is not uniquely determined.

8.3.3 Calibration with n-level premium functions

For the n-level family, there are $2n$ unknown premium levels. For a forecast anchor r with horizon t_r and maturity S_r, define the block coefficient

$$K_{a,i}(t_r) := \mathbb{1}_{\{t_r > \tau_{i-1}\}} \left[\exp\{-a(t_r - \min(t_r, \tau_i))\} - \exp\{-a(t_r - \tau_{i-1})\}\right].$$

Then

$$C^x_{r,i} := R_x(S_r) K_{a_x,i}(t_r), \qquad C^y_{r,i} := R_y(S_r) K_{a_y,i}(t_r).$$

The $2n$-column coefficient matrix is

$$\bar{\mathbf{A}} = \left(C^x_{\cdot,1}, \ldots, C^x_{\cdot,n}, C^y_{\cdot,1}, \ldots, C^y_{\cdot,n}\right).$$

With $2n$ forecast anchors, the premium path is uniquely identified if and only if

$$\operatorname{rank}(\bar{\mathbf{A}}) = 2n.$$

Equivalently, there must be no non-zero vector $u = (u^x_1, \ldots, u^x_n, u^y_1, \ldots, u^y_n)^\top$ such that

$$\sum_{i=1}^{n} C^x_{r,i} u^x_i + \sum_{i=1}^{n} C^y_{r,i} u^y_i = 0 \qquad \text{for all forecast anchors } r.$$

This is the higher-dimensional analogue of the determinant condition in the constant case. For $n > 2$, identification is governed by the full column rank of $\bar{\mathbf{A}}$.

A blockwise design gives a tractable sufficient structure. If two new forecast anchors are assigned to each block and the rows are ordered by horizon, the system becomes block lower triangular after subtracting the contributions of earlier blocks. The i-th block is then identified by a residual 2×2 system. With the block-relative horizon $\widetilde{h}_r := t_r - \tau_{i-1}$, the local diagonal block is

$$D_i = \begin{pmatrix} R_x(S_{2i-1})(1 - \exp(-a_x \widetilde{h}_{2i-1})) & R_y(S_{2i-1})(1 - \exp(-a_y \widetilde{h}_{2i-1})) \\ R_x(S_{2i})(1 - \exp(-a_x \widetilde{h}_{2i})) & R_y(S_{2i})(1 - \exp(-a_y \widetilde{h}_{2i})) \end{pmatrix}.$$

In this blockwise design, each new block is identified if

$$\det(D_i) \neq 0.$$

The constant-premium determinant criterion reappears locally with t_r replaced by the block-relative horizon $\widetilde{h}_r$. The rank analysis identifies when the chosen forecast design determines the premium levels. It does not validate the forecast targets themselves.

8.4 Comparing the risk-premium specifications

The constant, two-level, and n-level step-function specifications differ in their flexibility, data requirements, and long-run scenario behaviour. The relevant question is what each family can express, which forecast information it requires, and which long-run centre it implies.

The comparison is formulated at the level of premium families under the fixed curve convention established in Chapter 6. It connects term-premium modelling with actuarial scenario requirements, where real-world projections must be calibrated for their specific application and documented in accordance with ESG guidelines [31].

8.4.1 Forecast families and target construction

Any comparison of premium specifications starts with the target side. Under $\mathbb{Q}$, calibration can rely on liquid derivative prices. Under $\mathbb{P}$, identifying information is more limited. The available calibration material includes forward-based targets after one year, longer-horizon forward targets, annualised-forward constructions, and regulatory variants based on EIOPA-related long-end conventions.

When the forecast family changes, the risk premium path may change even if the risk-neutral G2++ parameters are held fixed. A richer risk premium specification should therefore not be interpreted as superior merely because it moves more flexibly. It has to be assessed together with the forecast information used to identify it.

The calibrated mean-reversion levels are sensitive to the target construction. Changing the forecast family affects the magnitude of the real-world drift adjustment and the allocation between the two Gaussian factors. A risk premium specification therefore cannot be assessed independently of the forecast information used to identify it. Representative calibration slices illustrating this dependence are reported in Appendix B.5.

8.4.2 Constant premium as the baseline

The constant premium is the baseline specification. It requires two identifying conditions and produces the simplest real-world drift adjustment. It isolates the change from $\mathbb{Q}$ to $\mathbb{P}$ and the role of the two factor-specific drift levels.

The restriction is that the same premium pair determines the finite-horizon adjustment and the long-run mean. If the chosen values are large enough to raise the long-run mean substantially, the simulated short end can overtake the long end too early, producing too many inverse curves or implausible long-run short-rate levels. This motivates the step-function extensions.

This relation is consistent with the measure-change construction in Berninger and Pfeiffer [11]. A constant real-world premium is a useful baseline, but it becomes restrictive when the same two numbers have to govern both the transition path and the stationary mean.

8.4.3 Two-level step functions as the first extension

The two-level step function is the first controlled extension. It separates a short-run risk premium from a long-run risk premium and allows the model to distinguish current market conditions from later normalisation. This changes both the finite-horizon expected rate paths and the asymptotic mean.

The interpretation is direct. The parameters d_x and d_y govern the first block, where the current market state remains influential. The parameters l_x and l_y determine the long-run mean. The two-level family is therefore the smallest premium family that can adjust the transition path without forcing the long-run mean to be determined by the same pair of parameters.

8.4.4 General n-level step functions and iterative calibration

The n-level step function allows the premium path to follow a sequence of forecast points rather than a single transition and a final long-run level. It offers more flexibility, but it also increases the risk of over-parametrisation. Each additional block requires additional forecast information.

The iterative construction is useful because it imposes order on the high-dimensional calibration problem. The premium path is built block by block. After the contribution of earlier blocks

has been fixed, the next block is calibrated to the remaining target information. The relevant question is therefore not how many levels can be estimated algebraically, but how much structure is supported by the available forecast targets and by the intended simulation horizon.

The n-level family is appropriate when the richer structure improves the scenario generator in a documented way: for example by preserving forecast-curve information that would otherwise be lost, by reducing implausible finite-horizon dynamics, or by producing a more credible long-run distribution. Without such information, added levels mainly transmit target noise into the simulated dynamics.

Family	**Identifying forecast information**	**Finite-horizon flexibility**	**Long-run mean**	**Governance reading**
Constant premium	Two identifying conditions and a sparse forecast family.	One common drift adjustment over all horizons.	Set by one constant premium pair.	Baseline for sparse information and short-horizon or tightly governed applications.
Two-level step function	Short-run and long-run anchors that separate transition and limit behaviour.	Distinguishes early-year adjustment from the long-run mean.	Set by the long-run premium pair.	Smallest controlled extension when current stress and later normalisation must be separated.
n-level step function	A forecast curve rich enough to identify several horizon blocks.	Can follow a richer premium path block by block.	Set by the final premium pair on the last block.	Appropriate only when the target family is sufficiently informative and the added governance burden is accepted explicitly.

Table 8.1: Operational comparison of the constant, two-level, and n-level risk-premium families.

The table summarises already-defined specifications. It does not introduce a new model. Its role is to make the trade-off explicit: flexibility, identifying information, and governance burden increase together.

8.4.5 A controlled common illustration

Figure 8.1 gives a controlled comparison. All three risk premium families are combined with the same risk-neutral G2++ calibration and the same forward-based forecast setup on 2 December 2024. This holds the $\mathbb{Q}$-model fixed and changes only the risk premium family.

The comparison shows that the risk premium family affects both the transition path and the long-run behaviour. In this example, the constant specification moves the expected rates towards a substantially lower long-run level. The two-level and n-level specifications produce higher long-run means and different transition shapes. The richer families also change the finite-horizon ordering of short and long rates and therefore the tendency toward normal or inverse yield-curve shapes.

The figure should not be read as a ranking of the three specifications. A richer family is preferable only if the implied path is economically interpretable and supported by the forecast information used to identify it. The figure illustrates that the risk premium family affects the scenario dynamics and should therefore be treated as a modelling decision.

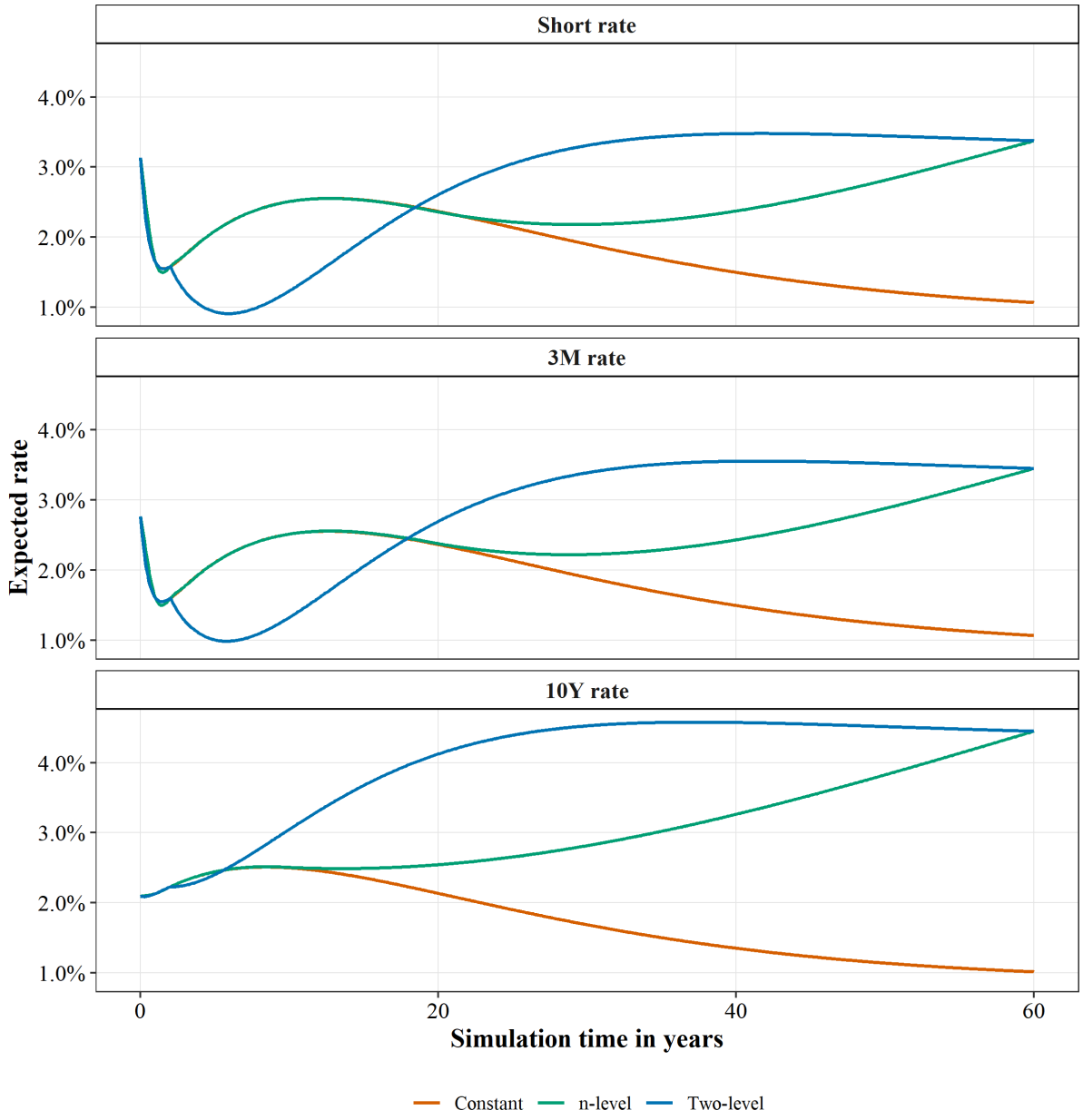


Figure 8.1: Expected short-rate, 3-month, and 10-year paths under constant, two-level, and n-level risk premiums for a common calibration date and common risk-neutral specification.

8.5 Ergodic stationary distributions

For long-horizon actuarial projections, the finite-horizon forecast fit can be complemented by the implied long-run distribution. The relevant object is the limiting stationary law obtained after the transient premium blocks have ceased to matter. The stationary distribution should place its main probability mass in a historically and economically plausible region of the rate space. This check belongs together with term-premium evidence and actuarial ESG guidance, because the long-run mean is determined by the chosen premium specification rather than by the short-horizon calibration equations alone [22, 10, 31].

8.5.1 The short rate

The risk premia change the stationary mean but not the stationary variance, because the shifts are deterministic. The expected short-rate limits are

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}^{const}[r_t] &\xrightarrow{t\to\infty} \psi_\infty + d_x + d_y,\\ \mathbb{E}_{\mathbb{P}}^{step}[r_t] &\xrightarrow{t\to\infty} \psi_\infty + l_x + l_y,\\ \mathbb{E}_{\mathbb{P}}^{steps}[r_t] &\xrightarrow{t\to\infty} \psi_\infty + \lambda_{x,n} + \lambda_{y,n}.\end{aligned}$$

The variance satisfies

$$\mathsf{Var}[r_t] = \frac{\sigma_x^2}{2a_x}\left(1-e^{-2a_x t}\right) + \frac{\sigma_y^2}{2a_y}\left(1-e^{-2a_y t}\right) + 2\rho_{xy}\frac{\sigma_x\sigma_y}{a_x+a_y}\left(1-e^{-(a_x+a_y)t}\right),$$

and therefore

$$\mathsf{Var}[r_t] \xrightarrow{t\to\infty} \sigma_r^2 := \frac{\sigma_x^2}{2a_x} + \frac{\sigma_y^2}{2a_y} + 2\rho_{xy}\frac{\sigma_x\sigma_y}{a_x+a_y}.$$

The short-rate ergodic laws are Gaussian:

$$\begin{aligned}r_t^{const} &\Rightarrow \mathcal{N}\left(\psi_\infty + d_x + d_y, \sigma_r^2\right),\\ r_t^{step} &\Rightarrow \mathcal{N}\left(\psi_\infty + l_x + l_y, \sigma_r^2\right),\\ r_t^{steps} &\Rightarrow \mathcal{N}\left(\psi_\infty + \lambda_{x,n} + \lambda_{y,n}, \sigma_r^2\right).\end{aligned}$$

The final premium pair determines the ergodic mean: (d_x, d_y) in the constant case, (l_x, l_y) in the two-level case, and $(\lambda_{x,n}, \lambda_{y,n})$ in the n-level case.

8.5.2 Spot rates of maturity S

For the continuously compounded spot rate with maturity $S > 0$, define

$$\alpha_x(S) := \frac{1-e^{-a_x S}}{a_x S}, \qquad \alpha_y(S) := \frac{1-e^{-a_y S}}{a_y S}.$$

The expected spot-rate limits are

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}^{const}[i_t^S] &\xrightarrow{t\to\infty} \psi_\infty - \frac{V(S)}{2S} + \alpha_x(S)d_x + \alpha_y(S)d_y,\\ \mathbb{E}_{\mathbb{P}}^{step}[i_t^S] &\xrightarrow{t\to\infty} \psi_\infty - \frac{V(S)}{2S} + \alpha_x(S)l_x + \alpha_y(S)l_y,\\ \mathbb{E}_{\mathbb{P}}^{steps}[i_t^S] &\xrightarrow{t\to\infty} \psi_\infty - \frac{V(S)}{2S} + \alpha_x(S)\lambda_{x,n} + \alpha_y(S)\lambda_{y,n}.\end{aligned}$$

The common stationary variance is

$$\sigma_{i(S)}^2 := \alpha_x(S)^2\frac{\sigma_x^2}{2a_x} + \alpha_y(S)^2\frac{\sigma_y^2}{2a_y} + 2\rho_{xy}\frac{\sigma_x\sigma_y}{a_x+a_y}\alpha_x(S)\alpha_y(S).$$

Therefore the spot-rate ergodic laws are

$$\begin{aligned}i_t^{S,const} &\Rightarrow \mathcal{N}\left(\psi_\infty - \frac{V(S)}{2S} + \alpha_x(S)d_x + \alpha_y(S)d_y, \sigma_{i(S)}^2\right),\\ i_t^{S,step} &\Rightarrow \mathcal{N}\left(\psi_\infty - \frac{V(S)}{2S} + \alpha_x(S)l_x + \alpha_y(S)l_y, \sigma_{i(S)}^2\right),\\ i_t^{S,steps} &\Rightarrow \mathcal{N}\left(\psi_\infty - \frac{V(S)}{2S} + \alpha_x(S)\lambda_{x,n} + \alpha_y(S)\lambda_{y,n}, \sigma_{i(S)}^2\right).\end{aligned}$$

The premium family changes the long-run mean, while the stationary variance is fixed by the risk-neutral G2++ volatility parameters and the maturity S.

8.5.3 Numerical verification and interpretation

The analytical limits were checked against the simulation implementation. The simulated long-run means and variances agree with the derived Gaussian limits within numerical tolerance.

Figure 8.2 illustrates the result on 2 December 2024. In this date-specific illustration, the historical densities are visibly multi-modal, whereas the model-implied ergodic distributions are normal and therefore have a single centre. The constant family remains closer to the lower part of the historical density, the two-level family lies near the upper historical density region, and the n-level family shifts the centre further upward. This ordering reflects the selected calibration date, the identifying forecast targets, and the resulting final premium pairs. It is therefore a conditional long-run plausibility reading of this calibration slice. The example separates two modelling effects: the premium family affects the finite-horizon path and, through its final premium pair, the long-run mean.

The ergodic analysis adds a long-run plausibility check to the finite-horizon calibration equations. A premium family should match the identifying forecast conditions and imply a limiting distribution that is credible for the intended actuarial use. The final premium pair is therefore economically relevant: in the constant, two-level, and n-level cases it determines the long-run mean of the simulated rate distribution.

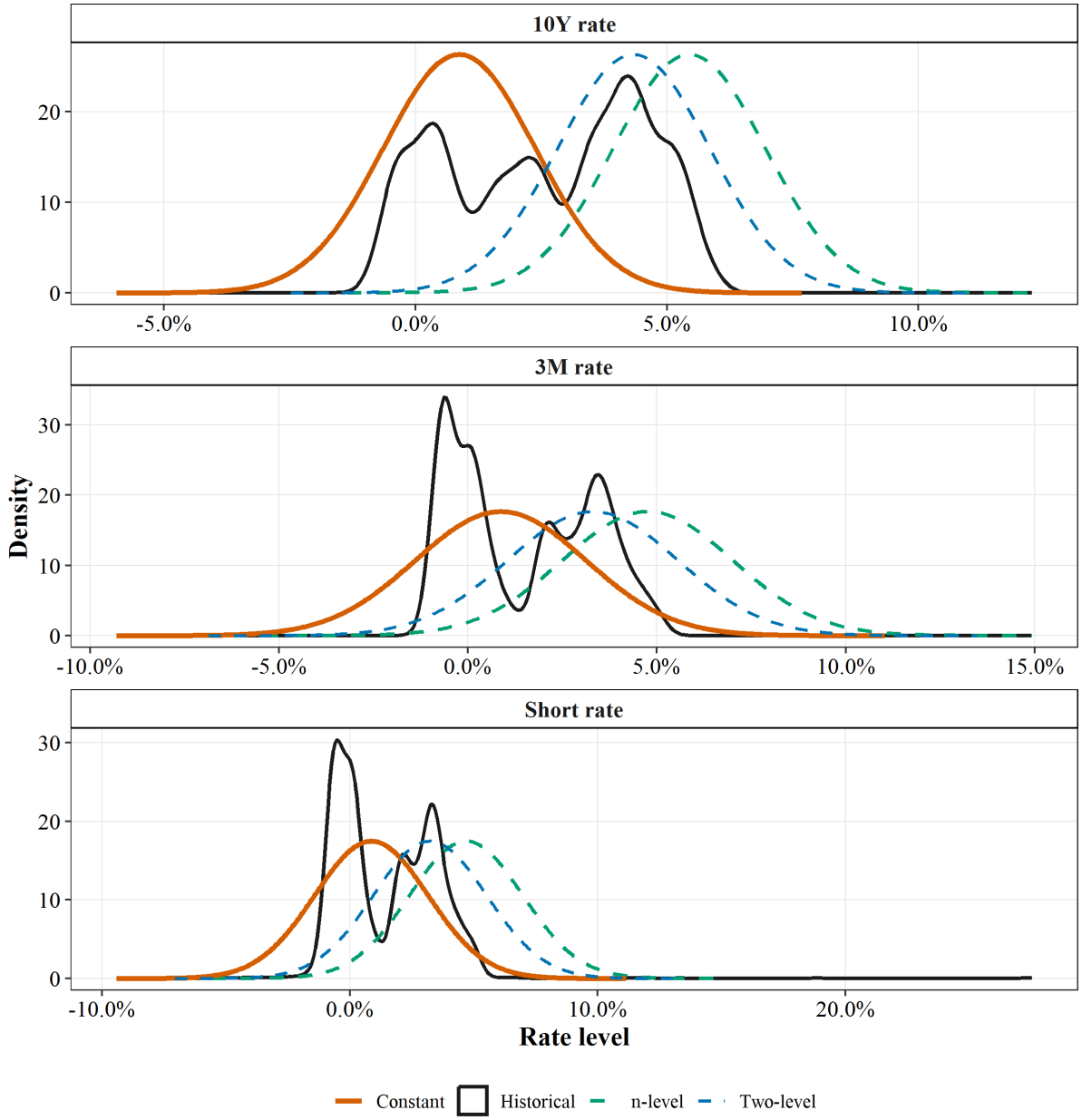


Figure 8.2: Historical rate densities and implied ergodic stationary distributions for the short rate, the 3-month rate, and the 10-year rate under constant, two-level, and n-level risk premiums on a common calibration date.

9 Case Study: openIRM Calibration

This chapter applies the calibration chain developed in the preceding chapters to the published *openIRM* benchmark [90]. The *openIRM* framework is a public internal-risk-model benchmark for a stylised life insurer in a Solvency-II setting. It combines an economic scenario generator, an asset-liability model, available-capital calculations, and a nested simulation problem. This chapter uses the framework as an application case for the calibration chain and concentrates on the calibration of the capital-market model.

The relevant link to this dissertation is the use of the same G2++ interest-rate model under two measures. The model is calibrated to market instruments under $\mathbb{Q}$ and then equipped with a real-world drift adjustment for one-year scenarios under $\mathbb{P}$. In the *openIRM* calibration, this also involves a deliberate curve split: the risk-neutral cap calibration is based on OIS market data, while the regulatory real-world simulation is initialised with the EIOPA risk-free term structure. This choice is part of the calibration design as a whole, because the curve used for derivative calibration and the curve used for solvency simulation serve different purposes. The resulting real-world scenarios enter the balance-sheet projection and affect available-capital distributions and solvency-relevant quantities.

The chapter reconstructs this calibration logic in detail. It explains the curve choices, the risk-neutral G2++ calibration, the construction of the real-world risk premium, and the role of these choices in the *openIRM* application. The broader Solvency-II internal-risk-model construction, the available-capital calculation, the nested-simulation setup, and the numerical benchmark design are summarised only where they are needed to understand how the calibrated market model is used. The chapter therefore reuses published *openIRM* material as an application case, while the general calibration methodology is developed in the preceding chapters.

9.1 Regulatory context: PRIIPs and Solvency II

Economic scenario generators are used in two regulatory environments discussed in this dissertation. In the PIA- and PRIIPs-related product-information setting, the capital-market model affects chance-risk classifications and standardised product comparisons. In the Solvency-II setting, it affects available-capital distributions, capital requirements, and internal-model validation. The present case study belongs primarily to the second setting. Its immediate regulatory object is the one-year internal-model problem of a life insurer.

The central output of *openIRM* is the distribution of available capital and the resulting *Solvency Capital Requirement* (SCR). The regulatory horizon is one year, the loss is measured by the change in available capital, and the SCR is defined through the 99.5% value-at-risk of the one-year loss distribution. This creates the nested simulation structure used in internal risk models. Outer scenarios describe the real-world development over the first year. Conditional on the outer risk factors, inner simulations value the future balance sheet under a risk-neutral measure.

The validation and governance requirements behind this structure are based on the Solvency-II internal-model framework [46, 40].

This setting requires a clear separation between $\mathbb{Q}$- and $\mathbb{P}$-calibration. Market consistency in the inner valuation step requires a risk-neutral model. The outer risk step relies on physical-measure scenarios to describe the one-year development of the risk factors. The transition between the two measures must therefore be documented and compatible with the initial term structure and the instruments used for risk-neutral calibration.

The case study also addresses a practical limitation of internal-model research. Non-public internal models make methodological comparison difficult. Nested-simulation methods, regression-based proxy models, explainability techniques, and variance-reduction methods are often tested on benchmark problems whose economic structure is only partially accessible. *openIRM* provides a public benchmark problem. In the calibration context, this means that numerical results can be reproduced and interpreted against a documented market-model specification.

9.2 Calibrating for European market regulation

The calibration proceeds in three steps. For each business day from September 2016 to the end of December 2023, the relevant term structures are constructed, the risk-neutral G2++ parameters are calibrated, and the real-world risk premium is determined. In the *openIRM* implementation, curve construction does not mean choosing one curve for all purposes. The risk-neutral cap calibration is based on OIS market data, whereas the regulatory real-world simulation is initialised with the EIOPA risk-free term structure. The curve layer therefore fixes the market and regulatory environments in which the later calibration steps are interpreted. The risk-neutral parameters fix derivative prices under $\mathbb{Q}$. The risk-premium parameters convert the model for the one-year real-world scenarios under $\mathbb{P}$.

The calibration design follows three requirements. First, the risk-neutral calibration must be tied to liquid market instruments and remain tractable on a daily sample over several years. Second, the real-world calibration must match the one-year Solvency-II risk horizon rather than an unrestricted long-horizon forecasting problem. Third, the setup has to remain interpretable for reproducible methodological comparison. These requirements favour a transparent calibration design over additional model layers whose effect on the solvency outputs would be difficult to separate.

In the notation of the preceding chapters, the case study implements the following chain:

curve construction $\longrightarrow$ $\mathbb{Q}$-calibration $\longrightarrow$ $\mathbb{P}$-calibration $\longrightarrow$ actuarial simulation and valuation.

9.2.1 Data preparation

The first data layer consists of the term structures used in the *openIRM* calibration. As in Chapter 4, the deterministic shift ψ in the G2++ short rate is used to fit the initial term structure at the calibration date. The initial term structure is represented through the forward-rate curve $f^M(0,t)$ derived from market data.

Two curve inputs enter here. For calibration under $\mathbb{Q}$, the benchmark uses observable OIS rates. This is consistent with the interest-rate derivatives used in the calibration step, because the caps are priced with the OIS 6-month curve as the underlying yield curve and a matching 6-month tenor. For the real-world simulation under $\mathbb{P}$, the benchmark uses the EIOPA risk-free interest-rate term structure. This is natural in a Solvency-II setting, where EIOPA curves are

part of the liability-valuation and capital-requirement environment [42]. The OIS/EIOPA curve split ensures that derivative calibration and regulatory balance-sheet simulation are performed in their respective consistent environments.

For simulation and risk-premium calibration, the curve must be available as a continuous object in maturity. The benchmark therefore uses the Nelson–Siegel–Svensson representation of the forward-rate curve $f^M(0,t)$, together with the corresponding spot-rate curve $i(0,t)$. This gives smooth term-structure inputs for both the simulation and the construction of forward-rate targets. The NSS fit is carried out before the risk-premium calibration, so that the targets are derived from a smooth term structure rather than from a discontinuous maturity grid.

The second data layer consists of the interest-rate caps used to identify the risk-neutral G2++ parameters. The benchmark calibrates to at-the-money caps with maturities from 3 years to 30 years. The strike of each cap is the corresponding forward swap rate at initiation. These instruments are chosen because the G2++ model admits explicit formulas for cap prices, as discussed in Section 4.3.1. The risk-neutral calibration can therefore be performed without an inner simulation loop.

Swaptions would also be relevant calibration instruments, but their G2++ pricing is only semi-analytical and therefore computationally more expensive than cap pricing. Since the benchmark requires daily calibration over several years, it uses at-the-money caps with explicit G2++ pricing formulas and comparatively liquid market quotes. The restriction to ATM instruments also avoids an additional strike dimension from in-the-money and out-of-the-money quotes.

The third data layer consists of the forward-rate forecasts used to determine the real-world risk premium. These forecasts are forward-rate targets derived from the EIOPA risk-free term structure used to initialise the real-world simulation, not option prices. Since *openIRM* simulates one year into the future under $\mathbb{P}$ before the inner valuation starts, the start time of the forward contract is set to $t = 1$. A short and a long maturity are used, namely three months and ten years. The two target rates are

$$f\left(0, 1, 1 + \frac{1}{4}\right) \quad \text{and} \quad f(0, 1, 1 + 10),$$

where again,

$$f(0, t, t + S) = -\frac{\log P(0, t + S) - \log P(0, t)}{S}.$$

These two targets determine the constant risk premium parameters λ_x and λ_y. Hence, the one-year 3-month forward and the one-year 10-year forward provide a short-end/long-end forecast pair for the one-year Solvency-II horizon. The construction does not imply a mechanical interpretation in which λ_x is the short-end premium and λ_y is the long-end premium. Both forward targets depend on both G2++ factors.

Other specifications would be possible, including different forward-rate pairs, different transition horizons or alternative curve bases. The benchmark design uses the one-year short-end/long-end pair because the outer simulation runs for one year and two constant risk premium parameters require two identifying conditions. The selected three-month and ten-year forwards give a direct target pair compatible with the two-factor model and the one-year Solvency-II risk horizon.

9.2.2 Model specifications

The capital-market model of *openIRM* consists of the already discussed G2++ interest-rate model and a generalised Black–Scholes stock model.

The choice of the G2++ framework is a key design choice of the case study. The earlier asset-liability model on which *openIRM* builds used a Vasiček short-rate model. The move to G2++ increases the flexibility of the short-rate dynamics while retaining an analytically manageable Gaussian structure. This trade-off is important for a regulatory benchmark. A one-factor model would be easier to calibrate but more restrictive in its term-structure behaviour. A richer model family could improve local fit but would reduce transparency and make daily calibration across a long sample less robust. Recent actuarial work on G2++ simulation and valuation supports the relevance of this tractable two-factor family for insurance applications [54, 3].

The risk-neutral parameter vector is

$$\theta = (a_x, a_y, \sigma_x, \sigma_y, \rho_{xy}).$$

These are the parameters calibrated to cap prices. The real-world risk-premium parameters are the factor-specific mean-reversion levels λ_x and λ_y. In the notation of Chapter 8, *openIRM* uses the baseline constant premium family on the interval $[0, 1]$. The application does not estimate an entire premium path. It determines which constant drift levels for the two factors are required so that the one-year outer simulation starts from the selected regulatory curve and reproduces one short-end and one long-end real-world target at $t = 1$. Since the outer *openIRM* simulation runs only one year under $\mathbb{P}$, these levels are taken to be constant over that horizon,

$$\lambda_x(t) := \lambda_x, \qquad \lambda_y(t) := \lambda_y.$$

This assumption is local to the one-year horizon. Constant mean-reversion levels are suitable for this benchmark horizon but may be too rigid for longer-horizon applications, where they can lead to inverse term-structure behaviour. In the present case, λ_x and λ_y are application-specific drifts, not universal long-run term-premium parameters.

Richer real-world specifications from Chapter 8 are not used in the benchmark calibration. Additional levels, transition dates, or smoother premium paths would introduce extra degrees of freedom on the $\mathbb{P}$-side. Such specifications can be valuable in long-horizon pension applications, but the one-year Solvency-II benchmark supplies two premium coordinates and two identifying forward targets. The constant specification therefore keeps the real-world premium no more flexible than the application can identify.

The stock process is calibrated in a simpler way. The stock excess return is set to 7% and the stock volatility to 25%, with the calibration motivated by the desired one-year equity stress behaviour. This is sufficient for the purpose of *openIRM*, because the main methodological role of the chapter lies in the interest-rate model and its use in the nested valuation problem. The equity component remains deliberately simple. Richer equity-rate coupling, stochastic equity volatility, or Heston-type dynamics with interest-rate-dependent drift could affect long-horizon product distributions and CRK stability. Such extensions are outside the G2++ interest-rate calibration chain used here.

9.2.3 Calibration results

The risk-neutral calibration is formulated as a cap-pricing least-squares problem. Let C_i^M denote the observed market price of cap i and let $C_i(0, T_i^C, N_i^C, K_i^C; \theta)$ denote the corresponding G2++ model price. With ν caps in the calibration set, the calibration minimises the relative root mean squared error

$$\min_{\theta=(a_x,a_y,\sigma_x,\sigma_y,\rho_{xy})} \sqrt{\frac{1}{\nu}\sum_{i=1}^{\nu}\left(\frac{C_i^M - C_i(0, T_i^C, N_i^C, K_i^C; \theta)}{C_i^M}\right)^2}$$

subject to

$$a_x > 0, \qquad a_y > 0, \qquad \sigma_x > 0, \qquad \sigma_y > 0, \qquad -1 \leq \rho_{xy} \leq 1.$$

The numerical optimisation algorithm used is Nelder–Mead.

The relative objective is important because cap prices across maturities are not on a common scale. An absolute-price objective, for example RMSE, would give more weight to expensive long-maturity contracts (see Section 7.3.2). The relative criterion treats errors proportionally and avoids reducing the benchmark calibration to a long-maturity price-matching exercise. This objective is not presented as universally optimal. It is aligned with the benchmark purpose of obtaining a balanced daily calibration across the cap universe.

Nelder–Mead is defensible in this setting. The parameter space is low-dimensional, the objective is nonlinear, and the calibration has to run repeatedly across many dates. Alternative solvers, parameter bounds, and cap subsets remain meaningful robustness checks. The final case-study design keeps the optimisation method simple enough to remain transparent and flexible enough to solve the daily calibration problem without a derivative or Hessian infrastructure at the benchmark stage.

The final calibration step determines λ_x and λ_y. With constant mean-reversion levels, the expected value of the first factor is

$$\mathbb{E}^{\mathbb{P}}[x(t)] = \lambda_x \left(1 - e^{-a_x t}\right),$$

and analogously for y. The expected spot rate for maturity $t + S$ is therefore linear in λ_x and λ_y. After subtracting the deterministic curve and variance terms, the two forward-rate targets define a two-by-two linear system of the form

$$\begin{pmatrix} A_{S_1}(t_1) & B_{S_1}(t_1) \\ A_{S_2}(t_2) & B_{S_2}(t_2) \end{pmatrix} \begin{pmatrix} \lambda_x \\ \lambda_y \end{pmatrix} = \begin{pmatrix} \Gamma_1 \\ \Gamma_2 \end{pmatrix}.$$

Here the coefficient matrix is the constant-premium matrix from Section 8.1.1, evaluated at the two target pairs (t_1, S_1) and (t_2, S_2), up to the notation used in the present chapter. The right-hand side contains the two forecast targets after subtracting the known curve and variance contributions. Invertibility of this matrix is the identification condition for a unique pair (λ_x, λ_y).

The system gives a direct interpretation of the real-world adjustment. Once the $\mathbb{Q}$-parameters are fixed, the two unknowns are the constant mean-reversion levels λ_x and λ_y. They are not separate short-end and long-end premia. Both targets depend on both G2++ factors, and the two maturities provide different factor sensitivities. The solution therefore specifies the two factor drifts needed to match the selected one-year short-end and long-end targets.

Other specifications would also be possible. One could choose different forward-rate targets, a different transition horizon, a different curve basis, or a different cap universe. These choices become relevant in broader actuarial applications. For the present benchmark, the specification is kept deliberately simple: the curve environment is fixed, the G2++ model is calibrated under $\mathbb{Q}$ to caps, and the $\mathbb{P}$ drift adjustment is determined from two forward-rate targets.

9.3 From calibration to available-capital outputs

The calibrated economic scenario generator is embedded into a stylised life insurer whose balance sheet is projected over time. The company holds bonds, stocks, cash, actuarial reserves, bonus reserves, equity, and free reserves. To avoid an arbitrary starting balance sheet, the case study constructs historical balance-sheet paths from 2016 to 2023 using observed market data. Bond

values are driven by historical spot-rate curves, while the stock component is linked to the MSCI World index. This produces the starting values for the solvency simulations.

Available capital is defined through two valuation routes. The direct method values policyholder cash flows under the risk-neutral measure, while the indirect method values shareholder cash flows. A central theoretical result of the *openIRM* publication is that, under the assumptions of the model, both methods converge to the same available-capital quantity. This result is important for the benchmark role of *openIRM*: numerical comparisons of direct and indirect estimators can then be interpreted as comparisons of simulation methods rather than as comparisons of different economic quantities.

The SCR estimation problem is written as a nested simulation problem. The outer simulation runs one year under $\mathbb{P}$ and generates real-world risk factors. Conditional on these risk factors, the inner simulation values available capital under $\mathbb{Q}$. The one-year loss is the difference between available capital at time zero and discounted available capital at time one, and the SCR is estimated as the 99.5% value-at-risk of this loss distribution. At this stage, the calibration chain enters the solvency calculation: the $\mathbb{Q}$-calibration affects the inner valuation, and the $\mathbb{P}$-calibration affects the outer loss distribution.

The risk model is simplified but structured. It covers market risks and life-insurance risks. Market risks include interest-rate risk, equity risk, and a simplified spread/default risk. Interest-rate risk is represented by stressing principal components of yield curves fitted with the Nelson–Siegel–Svensson method. Equity level and volatility risks are represented through the stock process and volatility shocks. Life-insurance risks include mortality, longevity, and lapse risks. Dependencies are introduced through a Gaussian copula. Despite its stylised construction, the model generates the features required of a benchmark nested-simulation problem: non-linearity, heteroscedasticity, asymmetry, heavy lower tails, and time variation.

The calibrated market model is therefore used in an actuarial environment in which management rules, liabilities, risk factors, and dependence assumptions interact. Calibration choices made at the curve, $\mathbb{Q}$, and $\mathbb{P}$ layers can propagate through the asset-liability model and affect loss distributions, short-rate scenarios, and solvency estimates. The benchmark is relevant for the calibration chain because it shows how curve, $\mathbb{Q}$-, and $\mathbb{P}$-choices affect solvency-relevant outputs.

The numerical study uses selected dates from 2016 to 2023, typically the last trading day up to October 15 of each year, and compares available-capital distributions under the direct and indirect methods. Both methods give very similar conditional means, as predicted by the theoretical consistency result, while their conditional distributions differ. The indirect method often has a smaller variance, although it may produce more extreme values in some conditional settings. The resulting available-capital distributions are generally robust but respond visibly to market conditions, especially to the interest-rate and yield-curve environment.

The calibration design has clear limits. The risk-neutral calibration is based on an RMSRE objective and Nelder–Mead optimisation. The local-diagnostic layer of Chapter 7 and therefore residual structure, leverage, influence, effective degrees of freedom, and identifiability are not part of the benchmark's own calibration protocol. The real-world risk premium is constant over the one-year outer horizon. This is appropriate for the Solvency-II benchmark, but it is not sufficient by itself for long-horizon pension applications. The split between OIS data for risk-neutral calibration and EIOPA curves for regulatory simulation is justified by purpose, but it must remain documented because the curve choice affects initial term structures, forward targets, and the interpretation of the resulting scenarios [42, 43, 31]. Finally, the benchmark uses simplified representations of the individual risk modules and of the dependence structure

between them.

The *openIRM* case study therefore provides a public application in which the $\mathbb{Q}$- and $\mathbb{P}$-parts of the calibration chain are both required and in which interest-rate calibration affects solvency-relevant outputs. At the same time, it does not replace the diagnostic and validation layers developed elsewhere in the dissertation. The next chapter turns this perspective into an explicit ex-post validation catalogue for simulated outputs.

10 Ex-Post Validation

Calibration does not end with the measure change. Once the model has been fitted and the real-world drift has been specified, the practical question is whether the resulting simulated paths are plausible for the intended actuarial application. This chapter formulates validation criteria for this output layer. The distinction between calibration, validation, intended use, and governance is also consistent with supervisory, actuarial and model-risk-management standards [31, 46, 40, 80, 1, 14].

The preceding chapters have already fixed the initial term structure, the risk-neutral calibration, the real-world risk premium construction, and the openIRM benchmark application. Ex-post validation is applied after these steps have produced simulated quantities that can be inspected. It is therefore conditional on the modelling conventions fixed earlier in the manuscript. In the main calibration chain, this includes the documented single-curve core setup; in the openIRM benchmark application it also includes the corresponding EIOPA-based overlay. Within these documented conventions, the question is whether the simulated outputs remain plausible for the actuarial purpose at hand.

10.1 Ex-Post Validation versus Calibration Diagnostics

The distinction from the diagnostics of Chapter 7 is important. The diagnostics under $\mathbb{Q}$ inspect the local geometry of the calibration map: residual structure, leverage, influence, effective degrees of freedom, identifiability, and boundary activity. These are questions about the *fit* and its local stability near the estimated parameter vector.

Ex-post validation asks a different class of questions. After the model has been carried through calibration and simulation, the relevant objects are the simulated outputs themselves: expected values of the stochastic processes, empirical quantiles, tail-sensitive summaries, inverse yield-curve frequencies, reference portfolios, and classification results. The question is whether these objects remain plausible and interpretable for the intended actuarial use. This output-oriented view is close to actuarial model validation practice, where model purpose, limitations, documentation, calibration quality, and output testing have to be considered together [31, 1, 3, 14].

This distinction can be stated schematically:

$$\begin{aligned} &\text{Diagnostics under } \mathbb{Q} : \text{residuals, leverage, influence, identifiability,} \\ &\text{Ex-post validation under } \mathbb{P} : \text{expected values, quantiles, tail measures, benchmark stability.} \end{aligned}$$

The two layers address different questions. Local diagnostics assess the fitted pricing map, whereas ex-post validation assesses the simulated outputs generated from the accepted calibration and premium specification.

The same distinction is embedded in the validation taxonomy in Table 10.1. The table places

calibration diagnostics and output validation within the broader language of model design, implementation, intended use, and governance.

Validation layer	**Supervisory or actuarial concept**	**Object in the calibration chain**
Conceptual soundness	Model design, intended use, and limitations.	G2++ as workhorse model, single-curve convention, $\mathbb{Q}$–$\mathbb{P}$ separation, and explicit scope limits.
Implementation validation	Reproducibility, numerical correctness, and controlled approximations.	Pricing formulas, curve construction, analytical cap Jacobian, finite-difference governance for swaptions, and solver checks.
Calibration validation	Fit, stability, sensitivity, and data adequacy.	Residuals, leverage, influence, EDoF, FIM-type uncertainty, parameter histories, and active constraints.
Output validation	Plausibility of simulated distributions and benchmark quantities.	Expected rates, quantiles, tail measures, inverse-curve shares, reference portfolios, and CRK-related outputs.
Governance validation	Documentation, review triggers, and acceptance decisions.	Validation catalogue, follow-up actions, and explicit limitations of the accepted calibration.

Table 10.1: Validation taxonomy connecting supervisory language to the calibration chain.

The later numerical examples are therefore not repetitions of Chapter 7. They concern the output-validation and governance layers rather than the local geometry of the pricing calibration. The chapter keeps this distinction fixed throughout.

10.2 A Documented Actuarial Validation Perspective

In actuarial practice, ex-post validation is usually documented as a review of how the current market environment affects the model and its simulated outputs. This framing matters because validation is not another optimisation exercise. It examines the consequences of the chosen calibration for the simulation and for the quantities that matter in the intended application.

Such validation documentation is naturally organised around concrete output objects. These include the current market situation at the simulation start, the simulated interest-rate development, the share of inverse yield curves, and expected equity returns over longer horizons. Depending on the actuarial application, more specific quantities are added. In a PIA-specific setting, validation documentation may also include the behaviour of reference portfolios and Chance-Risk Classes, as well as the reaction of example products. This shows why the validation layer has to be separated from the calibration objective. None of these objects appears directly in the calibration objective under $\mathbb{Q}$. The supervisory language of validation standards points in the same direction: the model has to be monitored for appropriateness, performance, stability, sensitivity, and use in decision processes, not merely for local parameter fit [46, 40].

10.2.1 Validation starts at the simulation origin

The first validation question concerns the market state at time zero. A simulation study starts from a concrete market environment, which fixes the initial level and shape of the yield curve and the relative position of risky and less risky assets. Ex-post validation therefore first checks whether the chosen calibration date is translated into a plausible simulation origin. This matters because a calibration may fit derivative prices while still producing an initial simulation state that is difficult to reconcile with the economic interpretation of the current market regime.

10.2.2 Expected values, quantiles, and tail-sensitive summaries

Expected values alone are not sufficient for ex-post validation. Simulated paths have to be summarised by several statistics. Let $Z^{(1)}, \ldots, Z^{(N)}$ denote simulated values of an output quantity, for example the short rate, an equity return, or the value of a pension contract. Relevant validation objects include

$$\widehat{\mu}_Z = \frac{1}{N}\sum_{i=1}^{N} Z^{(i)}, \qquad \widehat{q}_\alpha(Z), \qquad \widehat{ES}_\alpha(Z),$$

that is, empirical means, empirical quantiles, and tail-sensitive functionals such as Expected Shortfall.

These summaries are typical for actuarial validation because they describe not only the central path, but also the dispersion and the tail behaviour of simulated outcomes. The PIA setting gives a useful example. There the classification is based on distributional summaries of simulated contract outcomes. The chance measure is the annualised return derived from the mean scenario outcome over all scenarios, whereas the risk measure is the corresponding annualised return derived from the mean of the worst 20% of scenario outcomes. Thus the risk measure is based on a lower-tail average, or Conditional Tail Expectation, rather than on a single quantile. This illustrates why ex-post validation has to look beyond the residuals of the pricing calibration.

In the numerical part of this chapter, these summaries are used for the reconstructed ECB-hike-date case. They provide the validation layer for the simulated short-rate and equity-return distributions without relying only on expected values.

10.2.3 Yield-curve shape and the share of inverse curves

One of the most concrete validation criteria is the share of inverse yield curves in the simulation. The calibration objective does not contain such a quantity directly, yet the quantity is economically important because it summarises whether the model produces too many states in which the short end overtakes the long end.

This criterion is closely connected to the discussion in Chapter 8. Constant risk premiums can produce undesirable long-run ordering effects. The step-function extensions were introduced because the expected short rate may otherwise overtake the longer maturities too early or too often. Ex-post validation turns this modelling concern into a practical validation statistic. Instead of relying on visual plausibility alone, one asks how frequently inverse curves occur and whether that frequency is acceptable for the intended actuarial use.

The corresponding validation statistic can be written in a simple pathwise form. If $i_t^{3M,(n)}$ and $i_t^{10Y,(n)}$ denote the simulated 3-month and 10-year spot rates on path n at horizon t, then the empirical share of inverse curves is

$$\widehat{p}_t^{\,\mathrm{inv}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}\left\{ i_t^{3M,(n)} > i_t^{10Y,(n)} \right\}.$$

The criterion is hence the observed share of simulated scenarios in which the short end overtakes the long end at a given horizon. This endpoint comparison should not be read as a full shape classification of the simulated yield curve: a path counted as inverse by this statistic may still have a humped maturity profile between the selected short and long maturities.

The same logic extends to the ergodic distributions derived in Chapter 8. The stationary law is not itself a calibration diagnostic. It is a long-run validation object. It describes the asymp-

totic centre and dispersion of the simulation. Comparing that long-run law to historical rate experience provides another plausibility benchmark for the simulated term structure.

The controlled comparison in Chapter 8 sharpens the same point further by holding the $\mathbb{Q}$ specification fixed and changing only the premium family. Thus even after the risk-neutral layer has been fixed, significantly different long-run curve geometries remain possible under $\mathbb{P}$. Ex-post validation is precisely the stage at which one decides whether these implied long-run scenario properties are acceptable for the intended actuarial use.

10.2.4 Reference portfolios, CRK, and example products

Actuarial validation documentation in this setting does not stop at rates. It also covers benchmark objects that are built from those rates. In the PIA setting the central intermediate objects are the *Musterportfolios* and the resulting Chance-Risk Classes. These objects are crucial because they define the relative coordinate system by which products are judged.

The argument from Chapter 2 can now be restated as a validation rule. If the calibrated capital-market model causes the benchmark portfolios to move erratically, then the resulting CRK boundaries become unstable. A product may then jump between classes not because its economic nature changed, but because the underlying benchmark system became too sensitive to the calibration. That is a validation failure, even if the market fit itself looked good.

This is why actuarial validation documentation also looks at example products. Product outputs need not remain numerically fixed over time, but their ordering and interpretation must remain coherent. A new calibration may shift the entire chance-risk plane, which can be acceptable. If benchmark objects become unstable, classifications may become difficult to interpret, even when the product itself has not changed economically in any direct sense.

10.2.5 Broader actuarial validation style

Even when the surrounding application differs, actuarial validation documentation tends to ask closely related questions. It looks at summary statistics of economically relevant simulated outputs, at tail-sensitive quantities rather than only means, at the stability of benchmark or classification systems, and at monitoring across recalibration dates. In this sense, the validation objects used here are consistent with a broader actuarial validation style.

Recent actuarial scenario-generator literature reinforces this point from a different angle. Work on ensemble ESGs emphasizes parameter uncertainty and the difficulty of relying on a single supposedly optimal scenario generator; work on machine-learning-based scenario generators adds validation tasks such as checking dependence structures and detecting memorization effects [7, 68]. These papers are not part of the core G2++ methodology developed here, but they show that ESG validation remains an active research area and that the validation layers used here are consistent with current actuarial validation concerns.

10.3 A reconstructed ECB-hike-date validation case

To make the validation catalogue more concrete, it is useful to reconstruct one historical case using the same type of validation objects introduced above. The date chosen here is 21 July 2022, that is, the first ECB rate hike that already serves as a common event marker in Section 7.2 for the post-2022 transition. The reconstruction starts from the OIS-6M cap-based calibration under $\mathbb{Q}$ on that date and then applies the same forecast-to-premium logic as in the validation application: OIS-based short-term targets at $\tau = 25/12$ years, long-term targets at 60 years,

and a two-step premium specification under $\mathbb{P}$. In the notation of Chapter 8, this yields

$$d_x \approx -0.01\%, \qquad d_y \approx 5.68\%, \qquad l_x \approx -1.58\%, \qquad l_y \approx 2.39\%.$$

The corresponding forecast targets are a 3-month rate of about 1.90% and a 10-year rate of about 2.40% at τ, together with long-run targets of about 3.88% for the 3-month rate and 4.85% for the 10-year rate at 60 years. This example is used to show how the calibrated $\mathbb{Q}/\mathbb{P}$ chain behaves when the initial market state has already left the negative-rate regime but still produces a challenging validation profile for simulated outputs.

For this exact event date, the reconstruction is restricted to the rate, inverse-curve, cap-fit, and equity-return layers; it does not include a full CRK, sample-portfolio, or example-product panel. The interpretation is therefore limited to the validation objects that are actually carried through. In particular, the ECB-hike-date reconstruction covers the relative cap-price deviations, the expected-rate and forecast plot, the short-rate quantile fan, the decomposition of the expected short-rate path, the share of inverse yield curves, and the four equity-return plots used for this event-date validation layer. This is enough to show how the validation criteria of this chapter behave on a reconstructed policy-transition event date.

Table 10.2 records the evidence boundary between the general validation criteria introduced in this chapter and the reconstructed event-date artefacts. The manuscript does not claim a full historical CRK or product-validation panel where the corresponding output family is not part of the evidence used here.

Evidence object	**Source status**	**Role in this chapter**	**Claim allowed**
General validation criteria	Methodological validation layer introduced in this chapter.	Defines which simulated outputs are inspected.	Conceptual validation criteria, not numerical acceptance evidence.
ECB-hike-date reconstruction	Reconstructed event-date evidence.	Provides the numerical case for cap fit, expected rates, quantiles, inverse curves, and equity-return summaries.	Qualified event-date reading, not production-benchmark acceptance.
Historical CRK and product outputs for 2022-07-21	Not part of the event-date evidence used here.	Excluded from the reconstructed case rather than approximated.	No claim of full CRK, sample-portfolio, or example-product validation for this date.

Table 10.2: Evidence boundary for the ex-post validation examples.

Figure 10.1 starts with the cap-price deviation plot. On the ECB-hike date all 15 quoted maturities are available, and the relative pricing deviations remain in a narrow corridor of roughly -5.3% to $+5.7\%$. The largest positive deviation occurs at 18 months, whereas the largest negative deviation occurs around 15 years. The validation question is not whether the model misses the cap surface globally. It is whether a calibration with a reasonably tight initial fit still produces plausible long-run simulation outputs on a monetary-policy transition date. The central interest-rate validation layer is summarised in Figure 10.2. The relevant point is not only that the forecast targets are met, but also how far these targets already move the simulation away from the still-low starting environment. At the short forecast horizon τ, the expected 3-month rate is already about 1.90%, the expected short rate is close to that level, and the expected 10-year rate is about 2.40%. Thus the model moves rather quickly from the low-rate starting point into a clearly positive interest-rate regime. The same effect becomes stronger at long horizons: by year 60 the expected short rate is about 3.85%, the expected 3-month rate about 3.88%, and the expected 10-year rate about 4.85%. The decomposition of the expected short-rate path shows where this movement comes from. The deterministic ψ-component remains the

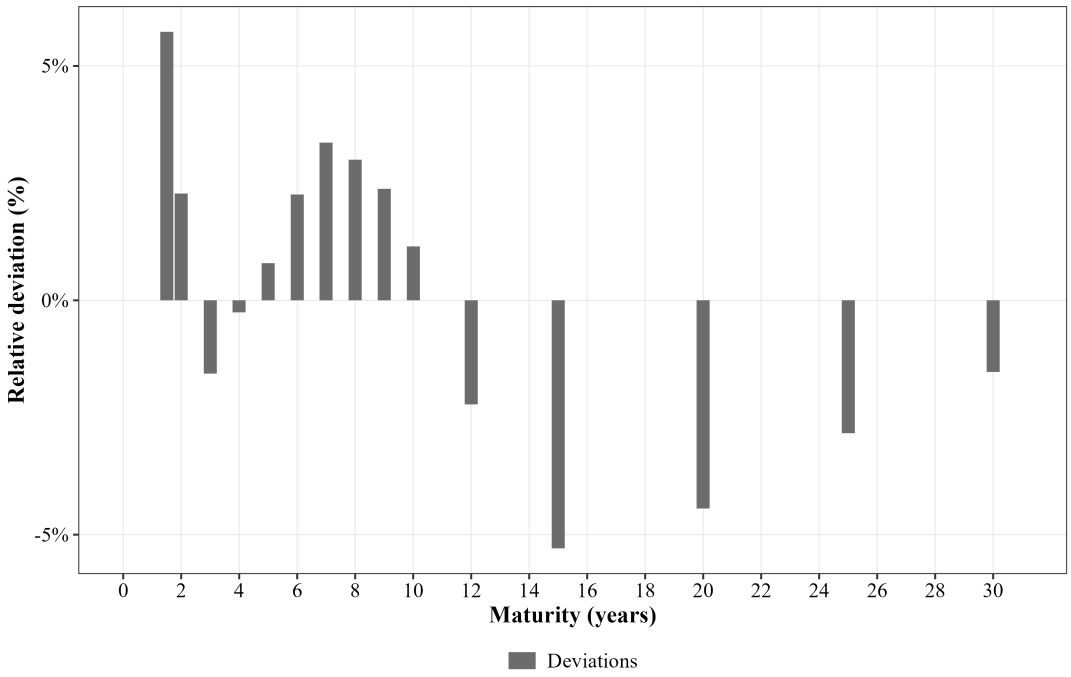


Figure 10.1: Relative cap-price deviations for the reconstructed ECB-hike-date validation case on 21 July 2022. The fitted cap surface remains comparatively tight across the full maturity set, so the validation discussion does not rest on a visibly broken price fit.

main long-run anchor, the y-factor adds a positive premium contribution, and the x-factor stays mildly negative. Hence the high long-run level is not a numerical artefact of the simulation, but follows from the chosen forecast targets and the corresponding premium specification.

The quantile fan and the inverse-curve share show the same transition from a different angle. At 60 years the 1%–99% short-rate band is approximately $[-8.3\%, 16.0\%]$, so the simulated rate distribution is much wider than the central mean path alone would suggest. The inverse-curve criterion summarises this information by counting how often the simulated 3-month spot rate exceeds the simulated 10-year spot rate. For this ECB-hike reconstruction, the share of inverse curves rises to a peak of about 58.8% after roughly 12 years and still ends around 28.8% in the long run, which is relatively high.

Figure 10.3 extends the same logic to the equity-return layer. The first row reports cumulative log-return summaries, first through the mean–median comparison and then through the corresponding quantile fan. The second row repeats the exercise for annual log returns. The equity mean path approaches the 4% reference, but the surrounding dispersion remains substantial. In the cumulative-return view the expected log return reaches about 3.93% by year 60, while the interquartile range at that horizon remains broad at roughly $[1.9\%, 5.9\%]$. In the yearly-return view the long-run mean is higher, reaching about 5.77% by year 60, but the interquartile range remains very wide at approximately $[-7.7\%, 19.3\%]$.

For validation, the important point is that the equity layer cannot be summarised by its long-run mean alone. Although the central tendency moves close to the reference level, the surrounding dispersion remains substantial. An actuarial reading must therefore distinguish between the mean path and the full output distribution. In this respect, the equity plots perform the intended role of ex-post validation: they prevent a single expected-value statement from standing in for the simulated distribution.

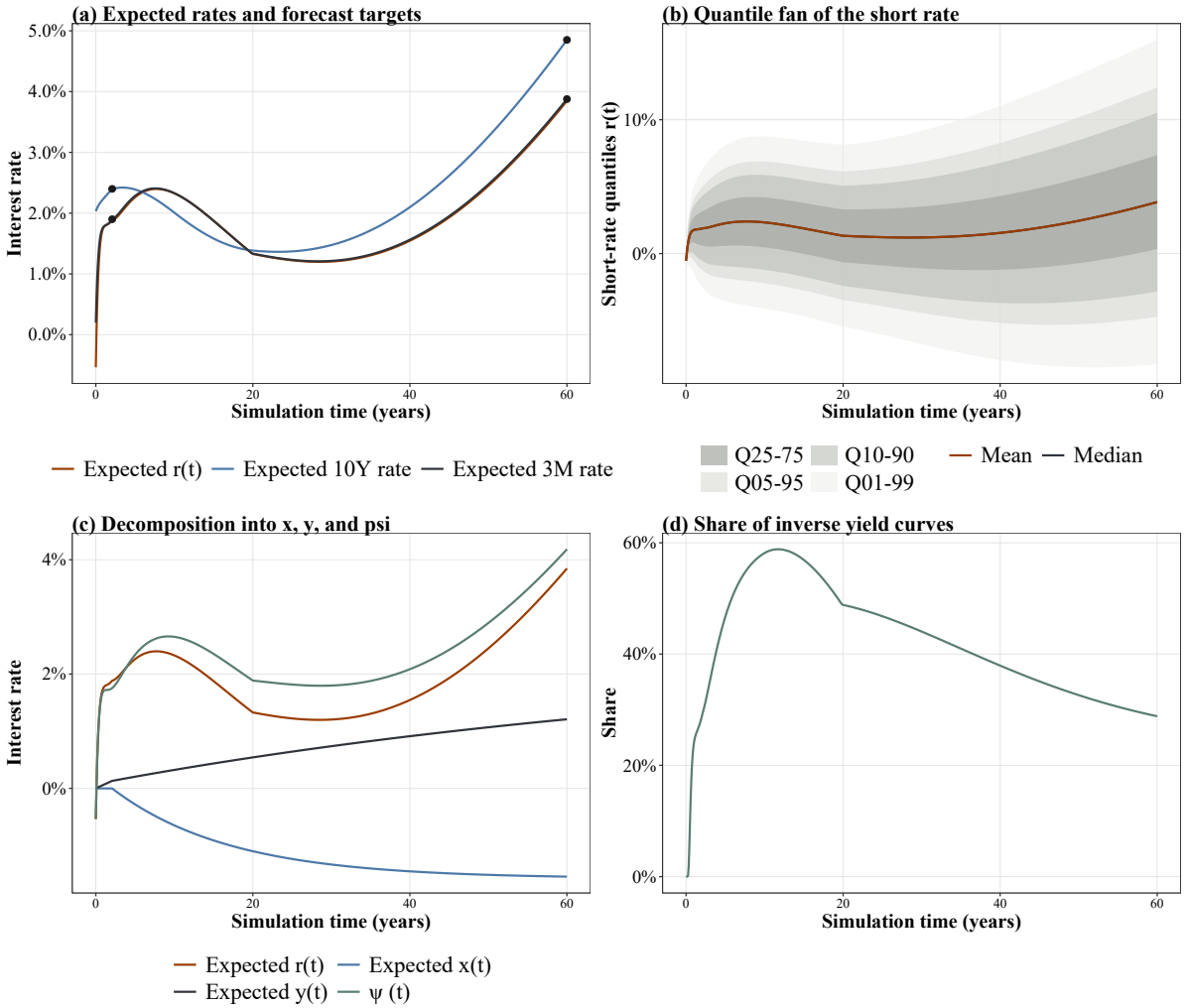


Figure 10.2: Interest-rate validation layer for the reconstructed ECB-hike-date case. The figure reproduces the validation objects introduced earlier for expected rates, quantiles, factor decomposition, and inverse-curve frequencies.

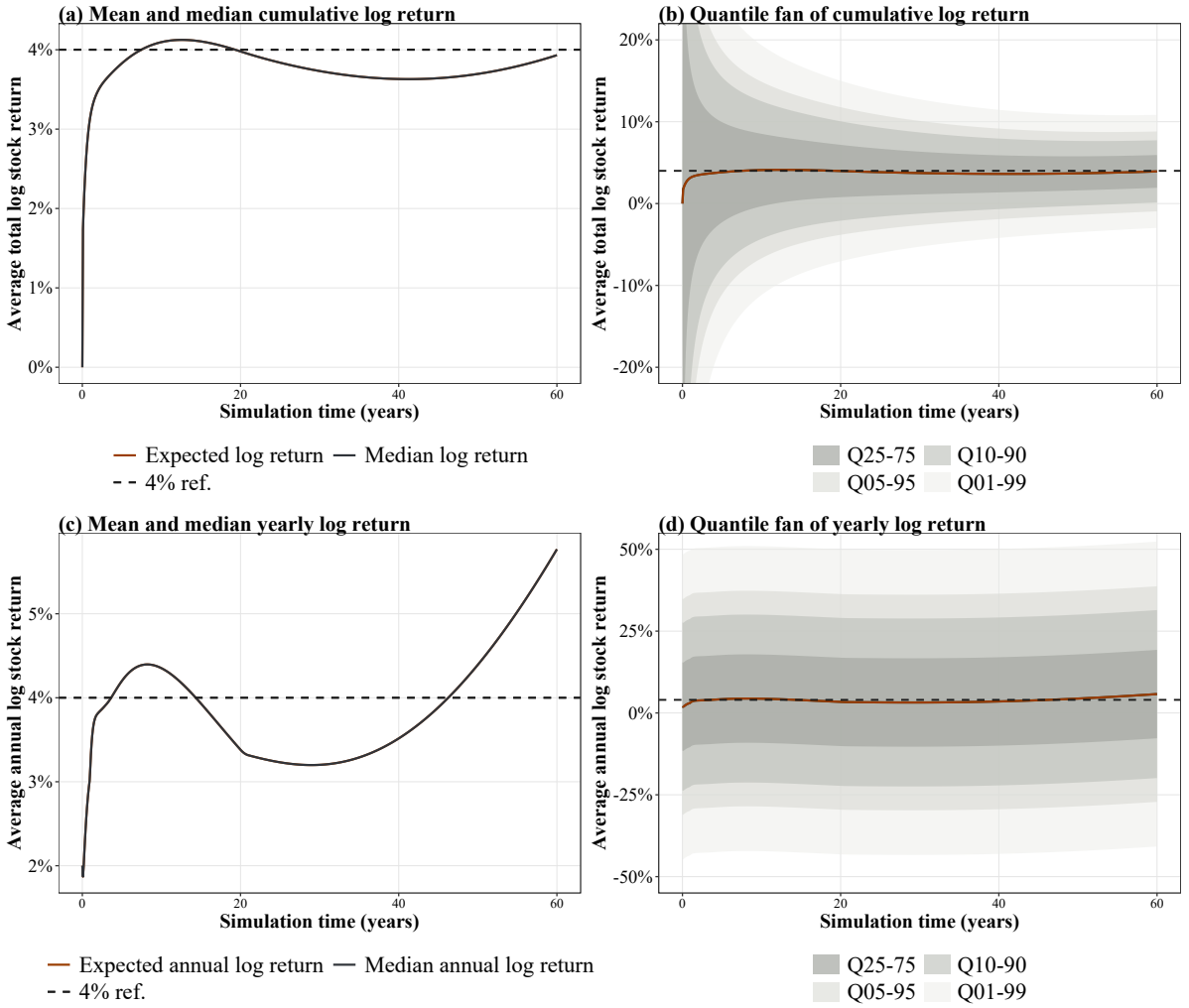


Figure 10.3: Equity-return validation layer for the reconstructed ECB-hike-date case. These equity plots form the equity layer of the event-date evidence; CRK and example-product outputs are outside the claim for this date.

Taken together, the ECB-hike reconstruction gives a mixed picture. The cap fit is comparatively tight, and the equity mean path approaches the 4% reference. At the same time, the share of inverse curves remains relatively high over long horizons. The example therefore shows why the validation step is useful even when the pricing fit does not look problematic. For this date, the model produces reasonable features in some outputs and less convincing behaviour in others. The case should therefore be read as an event-date example for this particular monetary-policy transition, not as a general statement about all hiking phases.

10.4 A structured validation catalogue

The validation logic developed in this chapter and the broader actuarial context together suggest a validation catalogue with five layers. The layers are ordered from the simulation starting point to the final governance decision. This order matters because a later output problem should not be interpreted before the market state, path distribution, and benchmark objects have been checked.

Plausibility of the initial simulation state

Before analysing long horizons, one checks whether the simulated world starts in an economically plausible way. This includes the level and shape of the initial yield curve, the relation between current market rates and benchmark portfolios, and the qualitative regime interpretation of the chosen calibration date.

Plausibility of simulated path distributions

The second layer inspects expected values, empirical quantiles, and tail summaries of key simulated quantities. This includes simulated short rates, selected longer maturities, and where relevant equity-related return objects. Equity-related expected values and quantiles can also become validation objects when equity dynamics enter the actuarial application.

Shape criteria for the term structure

The third layer studies the geometry of the simulated curve. The share of inverse curves is the most explicit example, but the same logic applies to persistent humps, implausibly high short ends, or long-run stationary laws that drift away from historical experience. This layer is where the practical effect of the chosen premium family also becomes visible.

Stability of benchmark objects

The fourth layer examines the objects that mediate between the capital-market model and the final decision output: reference portfolios, CRK boundaries, sample customers, and comparable internal benchmark constructions. These objects must remain interpretable if the later classification and communication are to remain trustworthy.

Stability of final application outputs and governance decisions

At this layer, validation results are translated into an acceptance, qualification, revision, or rejection decision. The output of validation is not just a descriptive report. If the simulated expected values, quantiles, inverse-curve shares, or benchmark objects show unacceptable behaviour, then the calibration has to be revised, accepted with qualification, or used only with explicit limitations for the intended application.

10.5 From validation findings to the final workflow

Ex-post validation completes the methodological chain. After the initial curve has been chosen, the model has been calibrated under $\mathbb{Q}$, and the real-world drift has been specified under $\mathbb{P}$, the resulting simulation is subjected to an actuarial test. This final validation step is required before the calibration can be treated as part of a practical calibration-and-validation framework rather than as an isolated estimation result.

The final chapter summarises the resulting calibration-and-validation workflow.

11 Conclusion and Calibration Workflow

11.1 Answer to the research question

The central question of this dissertation was how a stochastic interest-rate model can be calibrated for financial-market and actuarial applications without reducing the problem to a single numerical fit. The answer developed here is conditional on the chosen model framework and on the intended application. A calibration should be regarded as reliable only if the market fit, the local identifiability of the fitted parameters, the real-world dynamics, and the resulting simulation outputs are jointly plausible. A small pricing error is therefore an entry condition. It is not, by itself, evidence that the fitted parameter vector is stable, that the real-world risk premium specification is economically meaningful, or that the simulated application outputs are acceptable.

This leads to two validation layers. The first layer is local and risk-neutral. After calibration under $\mathbb{Q}$, the fitted pricing map is read through residuals, leverage, influence, effective degrees of freedom, conditioning, active constraints, and local covariance approximations. The second layer is downstream and actuarial. After the transition to $\mathbb{P}$, the simulated outputs are assessed through the quantities that matter for the intended use, such as rate distributions, inverse-curve frequencies, reference portfolios, product classifications, or solvency-relevant quantities. These two layers answer different questions and neither one replaces the other. A calibration that fits market prices but fails local identifiability remains fragile. A calibration that passes local diagnostics but produces implausible application outputs is not acceptable for the actuarial use case.

The result is therefore not a universal model ranking, but a calibration-and-validation criterion for the chosen G2++ framework. In this framework, calibration quality has to be assessed along the whole chain from data convention to governance decision.

11.2 The calibration-and-validation workflow

The workflow begins with the empirical market-data layer. Raw market data are inspected, cleaned, and translated into the market objects used for calibration. Market conventions and the initial term structure then fix discounting, forwarding, interpolation, extrapolation, and quote translation, and thereby determine the price targets to which the model is fitted. The workflow continues through risk-neutral calibration and local diagnostics, where the fitted $\mathbb{Q}$ parameter vector is judged beyond the residual norm. Only after this layer is diagnostically acceptable does the analysis move to the $\mathbb{P}$-measure risk premium construction. The workflow ends only after the resulting simulation outputs have been validated in the application environment.

Table 11.1 condenses this workflow into the main modelling decision, diagnostic or validation object, the main result, and the remaining limitation at each stage.

Stage	**Main modelling decision**	**Diagnostic or validation object**	**Main result**	**Remaining limitation or extension**
Initial term structure and quote reconstruction	Choose curve convention, interpolation, extrapolation, and quote-to-price translation.	Strike reconstruction, price reconstruction, and convention consistency.	The initial term structure and the quote translator are part of calibration rather than neutral preprocessing.	A full multi-curve trading reconstruction remains outside the core scope considered here.
Risk-neutral calibration	Fit G2++ under $\mathbb{Q}$ to reconstructed monetary targets.	RMSRE, residual structure, and parameter histories.	Market fit is necessary but cannot be read as the whole calibration result.	Any conclusion remains conditional on the chosen curve and instrument universe.
Local diagnostics	Read the weighted Jacobian geometry after fit.	Leverage, influence, EDoF, covariance, conditioning, and active constraints.	Low price error can coexist with weak local identification and concentrated calibration power.	The strongest analytical diagnostic layer is available for caps, while swaptions are treated as an instrument-class extension with explicit numerical controls.
Instrument-class transfer checks	Compare caps with swaptions under a common protocol.	Boundary activity, fit ranking, parameter-regime comparison, and cross-instrument transfer pricing on common date samples.	Swaptions do not remove the broader near-boundary regime; own-instrument fit does not guarantee transfer fit.	Treat these results as checks on the common date samples, not as a substitute for instrument-specific diagnostics or as a full derivative-universe validation.
One-factor benchmark	Compare Hull–White with G2++ on the paired cap sample.	Paired RMSRE, maturity-wise residuals, leverage, and influence.	G2++ fits the paired cap sample better, but it also creates a heavier diagnostic burden.	This is a benchmark comparison, not a universal ranking of all Gaussian rate models.
Real-world calibration	Choose constant, two-level, or n-level risk premium family under $\mathbb{P}$.	Forecast identification, transition paths, and ergodic centres.	Risk premium families serve different forecast-information sets and different application horizons.	Richer premium families require richer forecast support and stronger governance.
Ex-post validation	Judge simulated outputs in the application environment.	Quantiles, inverse curves, benchmark stability, CRK or solvency outputs, and acceptance decisions.	Ex-post validation is distinct from local $\mathbb{Q}$ diagnostics and is the final acceptance gate.	The evidence base can always be widened, but the core validation chain is complete.

Table 11.1: Compact summary of the calibration-and-validation workflow.

The table is not a substitute for the detailed chapters. Its role is to keep the final workflow summary compact without repeating the preceding evidence. Each row names the decision that must be fixed, the object used to judge it, and the corresponding claim boundary.

11.3 Main contributions

The contributions of the dissertation have both a methodological and an empirical focus. The first conceptual contribution is to treat the initial term structure as the first calibration decision. The curve chapter shows that strike reconstruction, shifted-Black translation, discounting conventions, and institutional curve sources determine the market object to which the model is later fitted. They therefore have to be documented before any parameter comparison can be interpreted.

The second contribution is the WLS and local diagnostic framework for risk-neutral G2++ calibration. The empirical evidence shows that a small RMSRE is not enough. Residual structure, local rank, leverage, influence, effective degrees of freedom, active bounds, Fisher-information-type uncertainty, and regime sensitivity all carry information about the credibility of a calibration. These quantities are local readings of the same weighted price geometry induced by the RMSRE objective and the Jacobian of the pricing map.

The third contribution is the controlled comparison layer. Caps provide the analytical benchmark for the diagnostic framework. Swaptions provide an instrument-class extension with explicit numerical controls and a robustness check across calibration instruments. The paired Hull–White comparison separates the gain in fit from the diagnostic burden created by the second factor. Together these comparisons control the scope of the G2++ evidence without turning the dissertation into a universal model ranking.

The fourth contribution is the transition from the risk-neutral measure $\mathbb{Q}$ to the real-world measure $\mathbb{P}$. This step is not treated as a purely formal change of measure, but as a separate calibration problem. In the G2++ model, the real-world drift is controlled by the two components λ^x and λ^y. Since these components affect the expected spot rates differently across maturities, the choice of forecast targets and of the premium specification matters for the simulated paths.

This motivates the analysis of constant, two-level, and more general step-function specifications in Chapter 8. These specifications differ in how much forecast information they require, how flexible the transition path becomes, and which long-run levels and stationary distributions they imply. A finer step structure is useful only if it is supported by the available forecasts and by the intended application horizon. Otherwise it adds complexity without making the real-world calibration more reliable.

The fifth contribution is the actuarial validation chain. In the PIA and PRIIPs context, calibration quality affects reference portfolios, chance-risk measures, and product classifications. In the *openIRM* context, it affects one-year real-world scenarios, nested valuation, available-capital distributions, and solvency-relevant quantities. These applications have different regulatory surfaces, but they share the same methodological requirement: calibration is complete only when the simulated outputs remain plausible, stable, and acceptable for the intended use.

11.4 Boundaries of the evidence

The risk-neutral evidence is limited to the G2++ setup used in this dissertation, the analytical cap diagnostic framework, the paired Hull–White versus G2++ cap benchmark, and the swaption instrument classes considered in Chapter 7. On the swaption side, even the full-window

OIS numerical-Jacobian panel remains a documented numerical result rather than a closed-form cap-style benchmark. Numerical Jacobians support diagnostics only under documented step-size, tolerance, conditioning, and boundary checks for the empirical sample under consideration.

The real-world evidence is limited by the forecast families and premium specifications studied in Chapter 8. The constant, two-level, and n-level specifications should not be read as automatic improvements of the same premium model. They use different amounts of forecast information and give different degrees of freedom to the transition from the current market state to the long-run behaviour. A richer risk premium path is useful only if the available forecasts support this additional flexibility and if the resulting transition and ergodic behaviour remain interpretable for the intended application.

The validation evidence in Chapter 10 is limited to the reconstructed ECB-hike event case on 2022-07-21. This case is used to illustrate the output-validation logic after the calibrated model has been carried from $\mathbb{Q}$ to $\mathbb{P}$. For that event date, the evidence covers the rate, inverse-curve, cap-fit, and equity-return layers. It does not include a full CRK, sample-portfolio, or example-product panel. The case should therefore be read as a controlled reconstruction of selected validation objects, not as a complete application-level validation exercise.

11.5 Extensions and future research

Future work could extend the chosen framework through richer model and instrument universes. The post-2022 evidence could motivate modified parametrisations, state-dependent parameters, regime-switching Gaussian short-rate models, stochastic-volatility extensions, or richer factor structures. The diagnostic logic could also be rebuilt for other interest-rate models, including affine multi-factor models, LIBOR-market-model-type specifications, or regime-dependent Gaussian models. Wider instrument coverage would add full multi-curve G2++ calibrations, fixed-strike caps and floors over longer windows, and additional curve or tenor conventions. In each case, the additional flexibility would only be useful if the resulting model could still be calibrated, diagnosed, transferred to $\mathbb{P}$, and validated in the final application. In the actuarial settings discussed here, such extensions would ultimately have to be assessed at the level of the full capital-market model. A concrete example is the PIA basic model described in [72], which combines a real-world interest-rate block with a generalized Black–Scholes equity block in which the simulated short rate enters the drift.

Robustness work would complement the local information reading rather than replace it. Bootstrap resampling, leave-one-out refits, robust refits, and wider multi-start or global-optimisation studies would answer questions that local Jacobian geometry does not answer by itself. Objective-function and weighting design also belongs here. The RMSE-versus-RMSRE check shows that the loss function is a genuine calibration-design choice; volume weights, liquidity weights, bid–ask information, or quote-frequency weights would define different calibration targets if the required market information is available and reliable.

The real-world and validation layers could also be widened. Deterministic step-function risk premia could be replaced by stochastic or state-dependent specifications, forecast uncertainty could be propagated directly into the risk premium path, and $\mathbb{P}$ calibration could be studied jointly over time instead of date by date. Longer monitoring windows, Monte-Carlo error bands for validation statistics, additional actuarial products, richer equity-rate interactions, further benchmark objects, and additional regime dates would make the validation layer more informative. However, the main structure would remain the same: the initial curve, the $\mathbb{Q}$-calibration, the $\mathbb{P}$-calibration, the actuarial use case, and the ex-post validation have to be assessed together.

References

[1] Actuarial Standards Board. ASOP no. 56: Modeling. Actuarial Standard of Practice, 2020.

[2] Tobias Adrian, Richard K. Crump, and Emanuel Moench. Pricing the term structure with linear regressions. *Journal of Financial Economics*, 110(1):110–138, 2013.

[3] Rohana Ambagaspitiya and Charles Ford. Calibrating interest rate models. Society of Actuaries Research Institute Research Report, October 2023.

[4] Leif B. G. Andersen and Vladimir V. Piterbarg. *Interest Rate Modeling*, volume 1–3. Atlantic Financial Press, 2010.

[5] Jan Annaert, Anouk G. P. Claes, Marc J. K. De Ceuster, and Hairui Zhang. Estimating the spot rate curve using the Nelson–Siegel model: A ridge regression approach. *International Review of Economics & Finance*, 27:482–496, 2013.

[6] A. C. Atkinson. Masking unmasked. *Biometrika*, 73(3):533–541, 1986.

[7] Jean-François Bégin. Ensemble economic scenario generators: Unity makes strength. *North American Actuarial Journal*, 27(3):444–471, 2023.

[8] David A. Belsley, Edwin Kuh, and Roy E. Welsch. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley, New York, 1980.

[9] Adi Ben-Israel and Thomas N. E. Greville. *Generalized Inverses: Theory and Applications*. CMS Books in Mathematics. Springer, New York, 2 edition, 2003.

[10] Andrea Berardi. Term premia and short rate expectations in the euro area. *Journal of Empirical Finance*, 74:101424, 2023.

[11] Christoph Berninger and Julian Pfeiffer. The Gauss2++ model: A comparison of different measure change specifications for a consistent risk neutral and real world calibration. *European Actuarial Journal*, 11:677–705, 2021.

[12] Ingo Beyna. *Interest Rate Derivatives: Valuation, Calibration and Sensitivity Analysis*. Springer, 2013.

[13] Fischer Black. The pricing of commodity contracts. *Journal of Financial Economics*, 3(1–2):167–179, 1976.

[14] Board of Governors of the Federal Reserve System and Office of the Comptroller of the Currency. Supervisory guidance on model risk management. Supervisory Letter SR 11-7 / OCC Bulletin 2011-12, 2011.

[15] Alan Brace, Dariusz Gatarek, and Marek Musiela. The market model of interest rate dynamics. *Mathematical Finance*, 7(2):127–155, 1997.

[16] Damiano Brigo and Fabio Mercurio. *Interest Rate Models: Theory and Practice*. Springer Finance. Springer, Berlin, Heidelberg, 2001.

[17] Bundesgesetzblatt. Gesetz zur Reform der gesetzlichen Rentenversicherung und zur Förderung eines kapitalgedeckten Altersvorsorgevermögens (Altersvermögensgesetz – AVmG). Bundesgesetzblatt Jahrgang 2001 Teil I Nr. 31, 2001.

[18] Bundesministerium der Justiz. Gesetz über die Zertifizierung von Altersvorsorge- und Basisrentenverträgen (Altersvorsorgeverträge-Zertifizierungsgesetz – AltZertG). Gesetze im Internet, 2001.

[19] Bundesministerium der Justiz. Altersvorsorge-Produktinformationsblattverordnung (AltvPIBV). Gesetze im Internet, 2015. Zuletzt geändert durch Art. 2 Abs. 1 V v. 12.11.2021.

[20] John Y. Campbell and Robert J. Shiller. Yield spreads and interest rate movements: A bird's eye view. *The Review of Economic Studies*, 58(3):495–514, 1991.

[21] Peter Christoffersen and Kris Jacobs. The importance of the loss function in option valuation. *Journal of Financial Economics*, 72(2):291–318, 2004.

[22] Benjamin H. Cohen, Peter Hördahl, and Dora Xia. Term premia: Models and some stylised facts. *BIS Quarterly Review*, pages 79–91, September 2018.

[23] R. Dennis Cook. Detection of influential observation in linear regression. *Technometrics*, 19(1):15–18, 1977.

[24] R. Dennis Cook and Sanford Weisberg. *Residuals and Influence in Regression*, volume 18 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1982.

[25] R. Dennis Cook and Sanford Weisberg. Regression diagnostics with dynamic graphics. *Technometrics*, 31(3):277–291, 1989.

[26] Christophe Croux, Geert Dhaene, and Dirk Hoorelbeke. Robust standard errors for robust estimators. CES Discussion Paper DPS 03.16, CES, KU Leuven, 2003.

[27] Sascha Desmettre and Ralf Korn. *Moderne Finanzmathematik – Theorie und praktische Anwendung Band 2: Erweiterungen des Black–Scholes-Modells, Zins, Kreditrisiko und Statistik*. Studienbücher Wirtschaftsmathematik. Springer Spektrum, Wiesbaden, 2018.

[28] Holger Dette, Viatcheslav B. Melas, and Andrey Pepelyshev. Optimal designs for estimating the slope of a regression. *Statistics*, 44(6):617–628, 2010.

[29] Deutsche Aktuarvereinigung. Ein Standardverfahren für PRIIP der Kategorie 4. Fachinformation, 01.07.2025, 2025.

[30] Deutsche Aktuarvereinigung. Ergebnisbericht des Ausschusses Lebensversicherung: Ein Standardverfahren für PRIIP der Kategorie 4. Ergebnisbericht, verabschiedet am 01.07.2025, 2025.

[31] Deutsche Aktuarvereinigung e.V. Anforderungen an einen ökonomischen Szenariengenerator. DAV-Hinweis, adopted by the DAV Executive Board on 27 November 2023, 2023. Non-binding professional guidance on the use, calibration and validation of capital-market models in insurance companies.

[32] Deutsche Bundesbank. Schätzung von Zinsstrukturkurven. Monatsbericht, Deutsche Bundesbank, October 1997.

[33] Deutsche Bundesbank. Term structure of interest rates in the debt securities market: Estimated values, listed federal securities, 2026. Daily Svensson parameter and curve series, accessed 27 May 2026.

[34] Deutsche Rentenversicherung. Die drei Säulen der Altersvorsorge. Online information page, 2026.

[35] Deutsche Rentenversicherung. Riester-Rente. Glossary entry, 2026.

[36] Francis X. Diebold and Canlin Li. Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364, 2006.

[37] Francis X. Diebold, Glenn D. Rudebusch, and S. Boragan Aruoba. The macroeconomy and the yield curve: A dynamic latent factor approach. *Journal of Econometrics*, 131(1–2):309–338, 2006.

[38] Digitale Rentenübersicht. Fonds-Sparplan ohne Garantie (Basis-Rente). Glossary entry, 2026.

[39] Darrell Duffie and Rui Kan. A yield-factor model of interest rates. *Mathematical Finance*, 6(4):379–406, 1996.

[40] European Insurance and Occupational Pensions Authority. Guidelines on the use of internal models. Guidelines, 2015.

[41] European Insurance and Occupational Pensions Authority. Report on the calculation of the Ultimate Forward Rate for 2026. Technical report, European Insurance and Occupational Pensions Authority, 2025.

[42] European Insurance and Occupational Pensions Authority. Risk-Free Interest Rate Term Structures. Technical documentation and monthly Solvency II RFR publications, 2025.

[43] European Insurance and Occupational Pensions Authority. Risk-Free Interest Rate Term Structures. EIOPA tools and data page, monthly technical information for Solvency II relevant risk-free interest rate term structures, 2026. Accessed in May 2026.

[44] European Insurance and Occupational Pensions Authority (EIOPA). Technical documentation of the methodology to derive EIOPA's risk-free interest rate term structures. Technical report, European Insurance and Occupational Pensions Authority (EIOPA), 2017.

[45] European Insurance and Occupational Pensions Authority (EIOPA). Risk-free interest rate term structures. Technical report, European Insurance and Occupational Pensions Authority (EIOPA), 2018.

[46] European Parliament and Council of the European Union. Directive 2009/138/EC (Solvency II Directive), Article 124: Validation Standards. Official Journal of the European Union; accessed via the EIOPA Single Rulebook, 2009.

[47] Fraunhofer ITWM. Klassifizierung privater Altersvorsorgeprodukte. Project and application description, 2026.

[48] Kenneth A. Froot. New hope for the expectations hypothesis of the term structure of interest rates. *The Journal of Finance*, 44(2):283–305, 1989.

[49] Helyette Geman, Nicole El Karoui, and Jean-Charles Rochet. Changes of numeraire, changes of probability measure and option pricing. *Journal of Applied Probability*, 32(2):443–458, 1995.

[50] Stefan Gerlach and Frank Smets. The term structure of euro-rates: Some evidence in support of the expectations hypothesis. *Journal of International Money and Finance*, 16(2):305–321, 1997.

[51] Manfred Gilli, Stefan Große, and Enrico Schumann. Calibrating the Nelson–Siegel–Svensson model, 2010.

[52] Tilmann Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

[53] Gene H. Golub and Charles F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, 4 edition, 2013.

[54] Sascha Günther and Peter Hieber. Efficient simulation and valuation of equity-indexed annuities under a two-factor G2++ model. *European Actuarial Journal*, 14:905–928, 2024.

[55] Patrick S. Hagan and Graeme West. Interpolation methods for curve construction. *Applied Mathematical Finance*, 13(2):89–129, 2006.

[56] Edmond Halley. An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslaw; with an attempt to ascertain the price of annuities upon lives. *Philosophical Transactions of the Royal Society of London*, 17:596–610, 1693.

[57] Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel. *Robust Statistics: The Approach Based on Influence Functions.* John Wiley & Sons, New York, 1986.

[58] David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, 60(1):77–105, 1992.

[59] Marc Henrard. *Interest Rate Modelling in the Multi-Curve Framework: Foundations, Evolution and Implementation.* Palgrave Macmillan, Basingstoke, 2014.

[60] David C. Hoaglin and Roy E. Welsch. The hat matrix in regression and ANOVA. *The American Statistician*, 32(1):17–22, 1978.

[61] Peter J. Huber. *Robust Statistics.* John Wiley & Sons, New York, 1981.

[62] John Hull and Alan White. Pricing interest-rate-derivative securities. *The Review of Financial Studies*, 3(4):573–592, 1990.

[63] John Hull and Alan White. Numerical procedures for implementing term structure models I: Single-factor models. *The Journal of Derivatives*, pages 7–16, 1994.

[64] John Hull and Alan White. Numerical procedures for implementing term structure models II: Two-factor models. *The Journal of Derivatives*, pages 37–48, 1994.

[65] John Hull and Alan White. A note on the models of Hull and White for pricing options on the term structure. *The Journal of Fixed Income*, 5(2):97–102, 1995.

[66] Farshid Jamshidian. An exact bond option formula. *The Journal of Finance*, 44(1):205–209, 1989.

[67] Farshid Jamshidian. LIBOR and swap market models and measures. *Finance and Stochastics*, 1(4):293–330, 1997.

[68] Gero Junike, Solveig Flaig, and Ralf Werner. Validation of machine learning based scenario generators. *Journal of Risk and Insurance*, 93(2):508–533, 2026.

[69] Samuel Karlin and William J. Studden. *Tchebycheff Systems: With Applications in Analysis and Statistics*. Interscience Publishers, New York, 1966.

[70] Jack Kiefer and Jacob Wolfowitz. Optimum designs in regression problems. *The Annals of Mathematical Statistics*, 30(2):271–294, 1959.

[71] Don H. Kim and Jonathan H. Wright. An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates. Finance and Economics Discussion Series 2005-33, Board of Governors of the Federal Reserve System, 2005.

[72] Ralf Korn and Andreas Wagner. Chance-Risk Classification of Pension Products: Scientific concepts and challenges. *Innovations in Insurance, Risk- and Asset Management*, pages 381–398, 2018.

[73] Ralf Korn and Andreas Wagner, editors. *Praxishandbuch Lebensversicherungsmathematik: Simulation und Klassifikation von Produkten*. VVW GmbH, Karlsruhe, 1 edition, 2019. In German.

[74] Asif Lakhany, Andrej Pintar, and Amber Zhang. Calibrating the Nelson–Siegel–Svensson model by genetic algorithm, 2021.

[75] Francis A. Longstaff, Pedro Santa-Clara, and Eduardo S. Schwartz. The relative valuation of caps and swaptions: Theory and empirical evidence. *The Journal of Finance*, 56(6):2067–2109, 2001.

[76] Philipp Mahler and Peter Ruckdeschel. Advanced calibration analysis and tools: Identifying influential observations in stochastic interest rate model calibration, 2026. arXiv preprint arXiv:2606.20420, DOI: 10.48550/arXiv.2606.20420.

[77] Philipp Mahler, Mark-Oliver Wolf, Marcus Willems, Jörg Wenzel, Andreas Wagner, and Ralf Korn. Konstruktion von Chancen-Risiko-Klassen zur Klassifizierung von Altersvorsorgeprodukten. *DAV Journal*, 2025. Also documented in the Fraunhofer-Publica record.

[78] Christoph Molnar. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Self-published, Munich, 2 edition, 2022.

[79] Charles R. Nelson and Andrew F. Siegel. Parsimonious modeling of yield curves. *The Journal of Business*, 60(4):473–489, 1987.

[80] Prudential Regulation Authority. Expectations for meeting the PRA's internal model requirements for insurers under Solvency II. Supervisory Statement SS1/24, 2024.

[81] Helmut Rieder. *Robust Asymptotic Statistics*. Springer Series in Statistics. Springer, New York, 1994.

[82] Valerio Russo and Giorgio Torri. Calibration of one-factor and two-factor Hull–White models using swaptions. *Computational Management Science*, 16(1):275–295, 2019.

[83] Jürgen F. Schmidt. *Die Deregulierung der Versicherungsaufsicht und die Versicherungsvermittlung in Deutschland*, volume 21 of *Beiträge zum Europäischen Wirtschaftsrecht*. Duncker & Humblot, Berlin, 2003.

[84] Daniel F. Schrager and Antoon Pelsser. Pricing swaptions and coupon bond options in affine term structure models. *Mathematical Finance*, 16(4):673–694, 2006.

[85] George A. F. Seber and Christopher J. Wild. *Nonlinear Regression*. Wiley Series in Probability and Statistics. Wiley, New York, 1989.

[86] Andrew Smith and Tim Wilson. Fitting yield curves with long term constraints. Technical report, Faculty and Institute of Actuaries, 2001. Research note commonly cited as the Smith–Wilson extrapolation reference.

[87] Lars E.O. Svensson. Estimating and interpreting forward interest rates: Sweden 1992-1994. IMF Working Papers 94/114, International Monetary Fund, 1994.

[88] Anders B. Trolle and Eduardo S. Schwartz. A general stochastic volatility model for the pricing of interest rate derivatives. *The Review of Financial Studies*, 22(5):2007–2057, 2009.

[89] Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.

[90] Mark-Oliver Wolf, Benedict Nikolaus Meyer, Philipp Mahler, and Maximilian Diehl. openIRM: Publicly accessible internal risk model of an artificial life insurer for analyzing and benchmarking actuarial methods in the Solvency II setting. *European Actuarial Journal*, 16:225–282, 2026.

A Explicit Entries of the Hessian Matrix

In this appendix, we provide the explicit analytical expressions required to assemble the cap Jacobian and the Hessian matrix $\mathbf{H}$ derived in Section 7.2.7. As established in (7.14), the (u, v)-th entry of the Hessian for the k-th cap is

$$H_{uv}^{k} = \sum_{m \in \mathcal{C}_k} \left[\mathcal{M}_m'(\Pi) \frac{\partial \Sigma_m^2}{\partial p_u} \frac{\partial \Sigma_m^2}{\partial p_v} + \mathcal{M}_m(\Pi) \frac{\partial^2 \Sigma_m^2}{\partial p_u \partial p_v} \right],$$

where $\Pi = (a_x, a_y, \sigma_x, \sigma_y, \rho_{xy})$. The first-order derivatives of the integrated variance function enter the Jacobian through (7.8); the same derivatives and the derivative of the weighting factor $\mathcal{M}'$ enter the Hessian through (7.14) and (7.15). This appendix lists the first- and second-order variance derivatives explicitly.

Since the Hessian is symmetric, we only need to derive the upper triangular entries. We categorise these derivatives based on the type of parameters involved.

A.1 Auxiliary Functions

For better readability, we recall the auxiliary function $C(a, t, T) = \frac{1 - e^{-a(T-t)}}{a}$ and define its derivatives with respect to a. As above, let $\delta_i = t_i - t_{i-1}$ denote the accrual period, then:

$$\begin{aligned} C_i(a) &= C(a, t_{i-1}, t_i) \\ C_i'(a) &= \frac{\partial C_i(a)}{\partial a} = \frac{e^{-a\delta_i}(a\delta_i + 1) - 1}{a^2} \\ C_i''(a) &= \frac{\partial^2 C_i(a)}{\partial a^2} = \frac{2 - e^{-a\delta_i}\left((a\delta_i)^2 + 2a\delta_i + 2\right)}{a^3}. \end{aligned}$$

Additionally, we define the auxiliary terms

$$\begin{aligned} \widetilde{C}_x &= C(2a_x, 0, t_{i-1}) \\ \widetilde{C}_y &= C(2a_y, 0, t_{i-1}) \\ \widetilde{C}_{xy} &= C(a_x + a_y, 0, t_{i-1}) \end{aligned}$$

Note that $\widetilde{C}_x, \widetilde{C}_y, \widetilde{C}_{xy}$ also depend on the mean reversion parameters and must be differentiated analogously to C_i' and C_i''. For example, $\frac{\partial \widetilde{C}_x}{\partial a_x} = 2C'(2a_x, 0, t_{i-1})$.

A.2 First-Order Variance Derivatives

For the i-th caplet, the integrated G2++ variance can be written as

$$\Sigma_i^2 = \sigma_x^2 C_i(a_x)^2 \widetilde{C}_x + \sigma_y^2 C_i(a_y)^2 \widetilde{C}_y + 2\rho_{xy}\sigma_x\sigma_y C_i(a_x) C_i(a_y) \widetilde{C}_{xy}.$$

The first-order variance derivatives used in (7.8) are therefore

$$\begin{aligned}
\frac{\partial \Sigma_i^2}{\partial \sigma_x} &= 2\sigma_x C_i(a_x)^2 \widetilde{C}_x + 2\rho_{xy}\sigma_y C_i(a_x) C_i(a_y) \widetilde{C}_{xy}, \\
\frac{\partial \Sigma_i^2}{\partial \sigma_y} &= 2\sigma_y C_i(a_y)^2 \widetilde{C}_y + 2\rho_{xy}\sigma_x C_i(a_x) C_i(a_y) \widetilde{C}_{xy}, \\
\frac{\partial \Sigma_i^2}{\partial \rho_{xy}} &= 2\sigma_x \sigma_y C_i(a_x) C_i(a_y) \widetilde{C}_{xy}.
\end{aligned}$$

For the mean-reversion speeds,

$$\begin{aligned}
\frac{\partial \Sigma_i^2}{\partial a_x} &= \sigma_x^2 \left[2C_i(a_x) C_i'(a_x) \widetilde{C}_x + C_i(a_x)^2 \frac{\partial \widetilde{C}_x}{\partial a_x} \right] \\
&\quad + 2\rho_{xy}\sigma_x\sigma_y \left[C_i'(a_x) C_i(a_y) \widetilde{C}_{xy} + C_i(a_x) C_i(a_y) \frac{\partial \widetilde{C}_{xy}}{\partial a_x} \right], \\
\frac{\partial \Sigma_i^2}{\partial a_y} &= \sigma_y^2 \left[2C_i(a_y) C_i'(a_y) \widetilde{C}_y + C_i(a_y)^2 \frac{\partial \widetilde{C}_y}{\partial a_y} \right] \\
&\quad + 2\rho_{xy}\sigma_x\sigma_y \left[C_i(a_x) C_i'(a_y) \widetilde{C}_{xy} + C_i(a_x) C_i(a_y) \frac{\partial \widetilde{C}_{xy}}{\partial a_y} \right].
\end{aligned}$$

A.3 Second Derivatives with Respect to Volatility and Correlation

These derivatives are the simplest, as the variance function Σ^2 is quadratic in volatilities and linear in correlation.

Second derivatives of volatilities:

$$\begin{aligned}
\frac{\partial^2 \Sigma_i^2}{\partial \sigma_x^2} &= 2C_i(a_x)^2 \widetilde{C}_x, \\
\frac{\partial^2 \Sigma_i^2}{\partial \sigma_y^2} &= 2C_i(a_y)^2 \widetilde{C}_y.
\end{aligned}$$

Mixed volatility and correlation derivatives:

$$\begin{aligned}
\frac{\partial^2 \Sigma_i^2}{\partial \sigma_x \partial \sigma_y} &= 2\rho_{xy} C_i(a_x) C_i(a_y) \widetilde{C}_{xy}, \\
\frac{\partial^2 \Sigma_i^2}{\partial \sigma_x \partial \rho_{xy}} &= 2\sigma_y C_i(a_x) C_i(a_y) \widetilde{C}_{xy}, \\
\frac{\partial^2 \Sigma_i^2}{\partial \sigma_y \partial \rho_{xy}} &= 2\sigma_x C_i(a_x) C_i(a_y) \widetilde{C}_{xy}.
\end{aligned}$$

Second derivative of correlation:

$$\frac{\partial^2 \Sigma_i^2}{\partial \rho_{xy}^2} = 0.$$

A.4 Second Derivatives Involving Mean-Reversion Speeds

Now, we look at the derivatives with respect to the mean reversion speeds a_x and a_y. We start by mixed derivatives and only list the ones with respect to a_x, since the derivatives for a_y follow by symmetry (swapping x and y).

$$\frac{\partial^2 \Sigma_i^2}{\partial \sigma_x \partial a_x} = 2\sigma_x \left[2C_i(a_x) C_i'(a_x) \widetilde{C}_x + C_i(a_x)^2 \frac{\partial \widetilde{C}_x}{\partial a_x} \right] + 2\rho_{xy}\sigma_y \left[C_i'(a_x) C_i(a_y) \widetilde{C}_{xy} + C_i(a_x) C_i(a_y) \frac{\partial \widetilde{C}_{xy}}{\partial a_x} \right],$$

$$\frac{\partial^2 \Sigma_i^2}{\partial \sigma_y \partial a_x} = 2\rho_{xy}\sigma_x \left[C_i'(a_x) C_i(a_y) \widetilde{C}_{xy} + C_i(a_x) C_i(a_y) \frac{\partial \widetilde{C}_{xy}}{\partial a_x} \right],$$

$$\frac{\partial^2 \Sigma_i^2}{\partial \rho_{xy} \partial a_x} = 2\sigma_x\sigma_y \left[C_i'(a_x) C_i(a_y) \widetilde{C}_{xy} + C_i(a_x) C_i(a_y) \frac{\partial \widetilde{C}_{xy}}{\partial a_x} \right].$$

The most involved computation is required when looking at the second derivative with respect to the mean-reversion speeds.

Second derivative w.r.t. a_x:

$$\frac{\partial^2 \Sigma_i^2}{\partial a_x^2} = \sigma_x^2 \Bigg[2\left((C_i'(a_x))^2 + C_i(a_x) C_i''(a_x) \right) \widetilde{C}_x + 4C_i(a_x) C_i'(a_x) \frac{\partial \widetilde{C}_x}{\partial a_x} + C_i(a_x)^2 \frac{\partial^2 \widetilde{C}_x}{\partial a_x^2} \Bigg] + 2\rho_{xy}\sigma_x\sigma_y \Bigg[C_i''(a_x) C_i(a_y) \widetilde{C}_{xy} + 2C_i'(a_x) C_i(a_y) \frac{\partial \widetilde{C}_{xy}}{\partial a_x} + C_i(a_x) C_i(a_y) \frac{\partial^2 \widetilde{C}_{xy}}{\partial a_x^2} \Bigg].$$

Mixed derivative a_x, a_y: Since the x-term of Σ^2 depends only on a_x and the y-term only on a_y, the only interaction occurs in the correlation term.

$$\frac{\partial^2 \Sigma_i^2}{\partial a_x \partial a_y} = 2\rho_{xy}\sigma_x\sigma_y \Bigg[C_i'(a_x) C_i'(a_y) \widetilde{C}_{xy} + C_i'(a_x) C_i(a_y) \frac{\partial \widetilde{C}_{xy}}{\partial a_y} + C_i(a_x) C_i'(a_y) \frac{\partial \widetilde{C}_{xy}}{\partial a_x} + C_i(a_x) C_i(a_y) \frac{\partial^2 \widetilde{C}_{xy}}{\partial a_x \partial a_y} \Bigg]$$

Note that for the term $\widetilde{C}_{xy} = C(a_x + a_y, 0, T_{i-1})$, the partial derivatives are straightforward:

$$\frac{\partial \widetilde{C}_{xy}}{\partial a_x} = \frac{\partial \widetilde{C}_{xy}}{\partial a_y} = C'(a_x + a_y, 0, T_{i-1}).$$

B Real-World Measure Derivations

This annex provides the mathematical details for the transition to the real-world measure $\mathbb{P}$ and for the moments and properties of the G2++ model under this measure. We first record the Girsanov drift-shift interpretation of the real-world factor dynamics and then derive the corresponding moments used in Section 4.4.

B.1 Girsanov Drift Shift and Real-World Factor Dynamics

Under the risk-neutral measure $\mathbb{Q}$, the two G2++ factors satisfy

$$\mathrm{d}x_t = -a_x x_t\,\mathrm{d}t + \sigma_x\,\mathrm{d}W_t^{x,\mathbb{Q}}, \qquad \mathrm{d}y_t = -a_y y_t\,\mathrm{d}t + \sigma_y\,\mathrm{d}W_t^{y,\mathbb{Q}},$$

with instantaneous correlation

$$\mathrm{d}\langle W^{x,\mathbb{Q}}, W^{y,\mathbb{Q}}\rangle_t = \rho_{xy}\,\mathrm{d}t.$$

The desired real-world dynamics are

$$\mathrm{d}x_t = a_x(\lambda_t^x - x_t)\,\mathrm{d}t + \sigma_x\,\mathrm{d}W_t^{x,\mathbb{P}}, \qquad \mathrm{d}y_t = a_y(\lambda_t^y - y_t)\,\mathrm{d}t + \sigma_y\,\mathrm{d}W_t^{y,\mathbb{P}}.$$

This form is obtained by a deterministic Brownian drift shift. Define

$$\theta_t^x = \frac{a_x\lambda_t^x}{\sigma_x}, \qquad \theta_t^y = \frac{a_y\lambda_t^y}{\sigma_y}.$$

Under the usual integrability assumptions, Girsanov's theorem gives an equivalent measure $\mathbb{P}$ under which

$$\mathrm{d}W_t^{x,\mathbb{P}} = \mathrm{d}W_t^{x,\mathbb{Q}} - \theta_t^x\,\mathrm{d}t, \qquad \mathrm{d}W_t^{y,\mathbb{P}} = \mathrm{d}W_t^{y,\mathbb{Q}} - \theta_t^y\,\mathrm{d}t$$

are Brownian increments under $\mathbb{P}$. Equivalently,

$$\mathrm{d}W_t^{x,\mathbb{Q}} = \mathrm{d}W_t^{x,\mathbb{P}} + \frac{a_x\lambda_t^x}{\sigma_x}\,\mathrm{d}t, \qquad \mathrm{d}W_t^{y,\mathbb{Q}} = \mathrm{d}W_t^{y,\mathbb{P}} + \frac{a_y\lambda_t^y}{\sigma_y}\,\mathrm{d}t.$$

Substituting these relations into the risk-neutral dynamics gives

$$\begin{aligned}\mathrm{d}x_t &= -a_x x_t\,\mathrm{d}t + \sigma_x\left(\mathrm{d}W_t^{x,\mathbb{P}} + \frac{a_x\lambda_t^x}{\sigma_x}\,\mathrm{d}t\right)\\ &= a_x(\lambda_t^x - x_t)\,\mathrm{d}t + \sigma_x\,\mathrm{d}W_t^{x,\mathbb{P}},\end{aligned}$$

and analogously

$$\mathrm{d}y_t = a_y(\lambda_t^y - y_t)\,\mathrm{d}t + \sigma_y\,\mathrm{d}W_t^{y,\mathbb{P}}.$$

The finite-variation drift shifts do not change quadratic variations or quadratic covariations. Therefore,

$$\mathrm{d}\langle W^{x,\mathbb{P}}, W^{y,\mathbb{P}}\rangle_t = \mathrm{d}\langle W^{x,\mathbb{Q}}, W^{y,\mathbb{Q}}\rangle_t = \rho_{xy}\,\mathrm{d}t.$$

Thus the measure change modifies the deterministic drift levels of the factors, while the diffusion coefficients and the instantaneous correlation remain unchanged. This is the reason why the risk-premium specification changes the expected values of x_t, y_t, and r_t, but not their variance formulas.

B.2 Properties of the Ornstein-Uhlenbeck Processes x_t and y_t

We start with the SDE for x_t under the real-world measure:

$$\mathrm{d}x_t = a_x(\lambda_t^x - x_t)\,\mathrm{d}t + \sigma_x\,\mathrm{d}W_t, \quad x_0 = 0.$$

Rearranging and multiplying by the integrating factor $e^{a_x t}$:

$$e^{a_x t}\,\mathrm{d}x_t + a_x e^{a_x t} x_t\,\mathrm{d}t = a_x \lambda_t^x e^{a_x t}\,\mathrm{d}t + \sigma_x e^{a_x t}\,\mathrm{d}W_t.$$

The left-hand side is the differential of $x_t e^{a_x t}$:

$$\mathrm{d}(x_t e^{a_x t}) = a_x \lambda_t^x e^{a_x t}\,\mathrm{d}t + \sigma_x e^{a_x t}\,\mathrm{d}W_t.$$

Integrating from 0 to t:

$$\int_0^t \mathrm{d}(x_u e^{a_x u}) = \int_0^t a_x \lambda_u^x e^{a_x u}\,\mathrm{d}u + \int_0^t \sigma_x e^{a_x u}\,\mathrm{d}W_u.$$

Since $x_0 = 0$, we get:

$$x_t e^{a_x t} = a_x \int_0^t \lambda_u^x e^{a_x u}\,\mathrm{d}u + \sigma_x \int_0^t e^{a_x u}\,\mathrm{d}W_u.$$

Therefore:

$$x_t = a_x \int_0^t \lambda_u^x e^{-a_x(t-u)}\,\mathrm{d}u + \sigma_x \int_0^t e^{-a_x(t-u)}\,\mathrm{d}W_u.$$

Since λ_u^x is deterministic and $\mathbb{E}[\int_0^t e^{-a_x(t-u)}\,\mathrm{d}W_u] = 0$:

$$\begin{aligned}\mathbb{E}[x_t] &= \mathbb{E}\left[a_x \int_0^t \lambda_u^x e^{-a_x(t-u)}\,\mathrm{d}u + \sigma_x \int_0^t e^{-a_x(t-u)}\,\mathrm{d}W_u\right] \\ &= a_x \int_0^t \lambda_u^x e^{-a_x(t-u)}\,\mathrm{d}u,\end{aligned}$$

$$\begin{aligned}\mathrm{Var}[x_t] &= \mathbb{E}[(x_t - \mathbb{E}[x_t])^2] \\ &= \mathbb{E}\left[\left(a_x \int_0^t \lambda_u^x e^{-a_x(t-u)}\,\mathrm{d}u + \sigma_x \int_0^t e^{-a_x(t-u)}\,\mathrm{d}W_u - \mathbb{E}[x_t]\right)^2\right] \\ &= \mathbb{E}\left[\left(a_x \int_0^t \lambda_u^x e^{-a_x(t-u)}\,\mathrm{d}u + \sigma_x \int_0^t e^{-a_x(t-u)}\,\mathrm{d}W_u - a_x \int_0^t \lambda_u^x e^{-a_x(t-u)}\,\mathrm{d}u\right)^2\right] \\ &= \mathbb{E}\left[\left(\sigma_x \int_0^t e^{-a_x(t-u)}\,\mathrm{d}W_u\right)^2\right] \\ &= {\sigma_x}^2 \mathbb{E}\left[\left(\int_0^t e^{-a_x(t-u)}\,\mathrm{d}W_u\right)^2\right] = {\sigma_x}^2 \int_0^t e^{-2a_x(t-u)}\,\mathrm{d}u = {\sigma_x}^2 \frac{1 - e^{-2a_x t}}{2a_x}.\end{aligned}$$

This confirms that the variance is invariant to the choice of the risk premium λ_t^x. □

B.3 Properties of the Short Rate r_t

The short rate is defined as $r_t = x_t + y_t + \psi_t$. The real-world risk premia enter through the mean-reversion levels of the factors x and y, not as an additional additive term in the short rate. Hence $\mathbb{E}[r_t] = \psi_t + \mathbb{E}[x_t] + \mathbb{E}[y_t]$.

Before calculating these expected values for specific risk premium functions (constant, two levels, n levels), we analyse the variance of the short rate.

We use the definition of variance $\mathsf{Var}[Z] = \mathbb{E}[(Z - \mathbb{E}[Z])^2]$. Since ψ_t is deterministic, it cancels out in the centred term $(r_t - \mathbb{E}[r_t])$.

$$\begin{aligned}\mathsf{Var}[r_t] &= \mathsf{Var}[x_t + y_t + \psi_t] = \mathsf{Var}[x_t + y_t] \\ &= \mathbb{E}\left[((x_t + y_t) - (\mathbb{E}[x_t] + \mathbb{E}[y_t]))^2\right] \\ &= \mathbb{E}\left[((x_t - \mathbb{E}[x_t]) + (y_t - \mathbb{E}[y_t]))^2\right].\end{aligned}$$

Let $\tilde{x}_t = x_t - \mathbb{E}[x_t]$ and $\tilde{y}_t = y_t - \mathbb{E}[y_t]$ be the zero-mean stochastic components.

$$\begin{aligned}\mathsf{Var}[r_t] &= \mathbb{E}\left[(\tilde{x}_t + \tilde{y}_t)^2\right] \\ &= \mathbb{E}\left[\tilde{x}_t^2 + 2\tilde{x}_t\tilde{y}_t + \tilde{y}_t^2\right] \\ &= \underbrace{\mathbb{E}[\tilde{x}_t^2]}_{\mathsf{Var}[x_t]} + \underbrace{\mathbb{E}[\tilde{y}_t^2]}_{\mathsf{Var}[y_t]} + 2\,\underbrace{\mathbb{E}[\tilde{x}_t\tilde{y}_t]}_{\mathsf{Cov}[x_t, y_t]}.\end{aligned}$$

The variances $\mathsf{Var}[x_t]$ and $\mathsf{Var}[y_t]$ have been derived above. We now derive the covariance term explicitly.

Covariance of x_t and y_t The covariance is defined as the expectation of the product of the centred processes, where we have:

$$\tilde{x}_t = \sigma_x \int_0^t e^{-a_x(t-u)}\,\mathrm{d}W_u^{x,\mathbb{Q}}, \quad \tilde{y}_t = \sigma_y \int_0^t e^{-a_y(t-v)}\,\mathrm{d}W_v^{y,\mathbb{Q}}.$$

The Brownian motions have a correlation $\mathrm{d}\langle W^x, W^y\rangle_t = \rho_{xy}\,\mathrm{d}t$. To compute the expectation of the product of these integrals, we apply the Itô isometry for correlated processes by integrating with respect to their quadratic covariation:

$$\mathbb{E}\left[\left(\int_0^t f(u)\,\mathrm{d}W_u^x\right)\left(\int_0^t g(u)\,\mathrm{d}W_u^y\right)\right] = \int_0^t f(u)g(u)\,\underbrace{\mathrm{d}\langle W^x, W^y\rangle_u}_{\rho_{xy}\,\mathrm{d}u}.$$

Applying this to our processes:

$$\begin{aligned}
\mathsf{Cov}[x_t, y_t] &= \mathbb{E}[\tilde{x}_t \tilde{y}_t] \\
&= \mathbb{E}\left[\left(\sigma_x \int_0^t e^{-a_x(t-u)}\, \mathrm{d}W_u^x\right)\left(\sigma_y \int_0^t e^{-a_y(t-u)}\, \mathrm{d}W_u^y\right)\right] \\
&= \int_0^t \left(\sigma_x e^{-a_x(t-u)}\right)\left(\sigma_y e^{-a_y(t-u)}\right)\rho_{xy}\, \mathrm{d}u \\
&= \sigma_x\sigma_y\rho_{xy}\int_0^t e^{-a_x(t-u)}e^{-a_y(t-u)}\, \mathrm{d}u \\
&= \sigma_x\sigma_y\rho_{xy}e^{-(a_x+a_y)t}\int_0^t e^{(a_x+a_y)u}\, \mathrm{d}u \\
&= \sigma_x\sigma_y\rho_{xy}e^{-(a_x+a_y)t}\left[\frac{1}{a_x+a_y}e^{(a_x+a_y)u}\right]_0^t \\
&= \frac{\sigma_x\sigma_y\rho_{xy}}{a_x+a_y}e^{-(a_x+a_y)t}\left(e^{(a_x+a_y)t}-1\right) \\
&= \frac{\sigma_x\sigma_y\rho_{xy}}{a_x+a_y}\left(1-e^{-(a_x+a_y)t}\right).
\end{aligned}$$

Combining the variances and the covariance, we obtain the total variance of the short rate:

$$\mathsf{Var}[r_t] = \frac{\sigma_x^2}{2a_x}(1-e^{-2a_x t}) + \frac{\sigma_y^2}{2a_y}(1-e^{-2a_y t}) + 2\rho_{xy}\frac{\sigma_x\sigma_y}{a_x+a_y}(1-e^{-(a_x+a_y)t}).$$

□

After having derived the variance, we calculate the expected values for the different types of risk premium.

B.3.1 Constant risk premiums

The expected value of x_t (and analogously of y_t), given constant mean reversion levels, is calculated as follows:

$$\begin{aligned}
\int_0^t e^{-a_x(t-u)}a_x\lambda_u^x\, \mathrm{d}u &= \int_0^t e^{-a_x t + a_x u}a_x d_x\, \mathrm{d}u \\
&= d_x a_x e^{-a_x t}\int_0^t e^{a_x u}\, \mathrm{d}u \\
&= d_x a_x e^{-a_x t}\frac{1}{a_x}\left(e^{a_x t}-1\right) \\
&= d_x\left(1-e^{-a_x t}\right)
\end{aligned}$$

□

B.3.2 Step Function: Two Levels

The expected value of x_t (and analogously of y_t) given step functions as mean reversion levels, is calculated as follows:

$$\begin{aligned}
\int_0^t e^{-a_x(t-u)} a_x \lambda_u^x \,\mathrm{d}u &= \int_0^t e^{-a_x t + a_x u} a_x \left(\mathbb{1}_{\{u \le \tau\}} d_x + \mathbb{1}_{\{u > \tau\}} l_x \right) \mathrm{d}u \\
&= e^{-a_x t} a_x \left(d_x \int_0^t e^{a_x u} \mathbb{1}_{\{u \le \tau\}} \,\mathrm{d}u + l_x \int_0^t e^{a_x u} \mathbb{1}_{\{u > \tau\}} \,\mathrm{d}u \right) \\
&= e^{-a_x t} a_x \left(d_x \int_0^{\min\{t,\tau\}} e^{a_x u} \,\mathrm{d}u + l_x \int_{\min\{t,\tau\}}^t e^{a_x u} \,\mathrm{d}u \right) \\
&= e^{-a_x t} a_x \left(d_x \frac{1}{a_x} \left(e^{a_x \min\{t,\tau\}} - 1 \right) + l_x \frac{1}{a_x} \left(e^{a_x t} - e^{a_x \min\{t,\tau\}} \right) \right) \\
&= d_x \left(e^{-a_x t + a_x \min\{t,\tau\}} - e^{-a_x t} \right) + l_x \left(1 - e^{-a_x t + a_x \min\{t,\tau\}} \right) \\
&= d_x \left(e^{-a_x (t - \min\{t,\tau\})} - e^{-a_x t} \right) + l_x \left(1 - e^{-a_x (t - \min\{t,\tau\})} \right)
\end{aligned}$$

□

B.3.3 Step Function: n Levels

Instead of just having one jump from d_x to l_x (and d_y to l_y resp.), we now want to look at the general case of having n levels $\lambda_{x,i}$ and $\lambda_{y,i}$, $i = \{1, \ldots, n\}$, and $(n-1)$ jump times τ_i, i.e.

$$\begin{aligned}
\lambda_t^x &:= \mathbb{1}_{\{t \le \tau_1\}} \lambda_{x,1} + \mathbb{1}_{\{\tau_1 < t \le \tau_2\}} \lambda_{x,2} + \cdots + \mathbb{1}_{\{t > \tau_{n-1}\}} \lambda_{x,n} \\
\lambda_t^y &:= \mathbb{1}_{\{t \le \tau_1\}} \lambda_{y,1} + \mathbb{1}_{\{\tau_1 < t \le \tau_2\}} \lambda_{y,2} + \cdots + \mathbb{1}_{\{t > \tau_{n-1}\}} \lambda_{y,n}.
\end{aligned}$$

As with the other cases, we start by determining the expected value of x_t (and analogously of y_t):

$$\begin{aligned}
&\int_0^t e^{-a_x(t-u)} a_x \lambda_u^x \,\mathrm{d}u \\
=&\int_0^t e^{-a_x t + a_x u} a_x \left(\mathbb{1}_{\{u \le \tau_1\}} \lambda_{x,1} + \mathbb{1}_{\{\tau_1 < u \le \tau_2\}} \lambda_{x,2} + \ldots + \mathbb{1}_{\{u > \tau_{n-1}\}} \lambda_{x,n} \right) \mathrm{d}u \\
=&e^{-a_x t} a_x \left(\lambda_{x,1} \int_0^{\min\{t,\tau_1\}} e^{a_x u} \,\mathrm{d}u + \lambda_{x,2} \int_{\min\{t,\tau_1\}}^{\min\{t,\tau_2\}} e^{a_x u} \,\mathrm{d}u + \ldots + \lambda_{x,n} \int_{\min\{t,\tau_{n-1}\}}^t e^{a_x u} \,\mathrm{d}u \right).
\end{aligned}$$

We look at the integral in the last line one by one:

$$\begin{aligned}
\lambda_{x,1} \int_0^{\min\{t,\tau_1\}} e^{a_x u} \,\mathrm{d}u &= \lambda_{x,1} \frac{1}{a_x} \left(e^{a_x \min\{t,\tau_1\}} - 1 \right), \\
\lambda_{x,2} \int_{\min\{t,\tau_1\}}^{\min\{t,\tau_2\}} e^{a_x u} \,\mathrm{d}u &= \lambda_{x,2} \frac{1}{a_x} \left(e^{a_x \min\{t,\tau_2\}} - e^{a_x \min\{t,\tau_1\}} \right), \\
\lambda_{x,n} \int_{\min\{t,\tau_{n-1}\}}^t e^{a_x u} \,\mathrm{d}u &= \lambda_{x,n} \frac{1}{a_x} \left(e^{a_x t} - e^{a_x \min\{t,\tau_{n-1}\}} \right).
\end{aligned}$$

The expressions for $\lambda_{x,3}, \ldots, \lambda_{x,n-1}$ with their corresponding τ_i and τ_{i-1} follow the same pattern as the one for $\lambda_{x,2}$ above.

In total, we get by regrouping the terms

$$\begin{aligned}\int_0^t e^{-a_x(t-u)} a_x \lambda_u^x \,\mathrm{d}u =& \lambda_{x,1}\left(e^{-a_x(t-\min\{t,\tau_1\})} - e^{-a_x t}\right)\\ &+\lambda_{x,2}\left(e^{-a_x(t-\min\{t,\tau_2\})} - e^{-a_x(t-\min\{t,\tau_1\})}\right)\\ &+\ldots\\ &+\lambda_{x,n-1}\left(e^{-a_x(t-\min\{t,\tau_{n-1}\})} - e^{-a_x(t-\min\{t,\tau_{n-2}\})}\right)\\ &+\lambda_{x,n}\left(1 - e^{-a_x(t-\min\{t,\tau_{n-1}\})}\right)\\ =& -\lambda_{x,1}e^{-a_x t} + \sum_{i=1}^{n-1}\left(\lambda_{x,i} - \lambda_{x,i+1}\right) e^{-a_x(t-\min\{t,\tau_i\})} + \lambda_{x,n}.\end{aligned}$$

Hence, for the expected value of the short rate in the case of n levels, we have

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}^{steps}[r_t] =& \psi_t - \lambda_{x,1}e^{-a_x t} + \sum_{i=1}^{n-1}\left(\lambda_{x,i} - \lambda_{x,i+1}\right) e^{-a_x(t-\min\{t,\tau_i\})} + \lambda_{x,n}\\ &- \lambda_{y,1}e^{-a_y t} + \sum_{i=1}^{n-1}\left(\lambda_{y,i} - \lambda_{y,i+1}\right) e^{-a_y(t-\min\{t,\tau_i\})} + \lambda_{y,n}.\end{aligned}$$

□

B.4 Expected Spot Rates and Calibration Equations

For the continuously compounded spot rate with maturity $S > 0$, the affine bond-price representation implies

$$i_t^S = \frac{1}{S}\int_t^{t+S} \psi_u \,\mathrm{d}u + \frac{1-e^{-a_x S}}{a_x S} x_t + \frac{1-e^{-a_y S}}{a_y S} y_t - \frac{V(S)}{2S}.$$

Taking expectations under $\mathbb{P}$ and inserting the general real-world expectations of x_t and y_t yields

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[i_t^S] = &\frac{1}{S}\int_t^{t+S} \psi_u \,\mathrm{d}u\\ &+ \frac{1-e^{-a_x S}}{a_x S}\int_0^t a_x e^{-a_x(t-u)}\lambda_u^x \,\mathrm{d}u\\ &+ \frac{1-e^{-a_y S}}{a_y S}\int_0^t a_y e^{-a_y(t-u)}\lambda_u^y \,\mathrm{d}u\\ &- \frac{V(S)}{2S}.\end{aligned}$$

Hence, once the $\mathbb{Q}$ parameters and the market curve are fixed, every forecast condition

$$\theta_S^F \stackrel{!}{=} \mathbb{E}_{\mathbb{P}}[i_t^S]$$

is linear in the risk premium parameters. This is the algebraic reason why the real-world calibration reduces to linear systems.

Constant risk premiums For constant levels $\lambda_t^x = d_x$ and $\lambda_t^y = d_y$, the previously derived formulas imply

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[i_t^S] = \frac{1}{S}\left(\log\left(\frac{P^M(0,t)}{P^M(0,t+S)}\right) + \frac{V(t+S)-V(t)-V(S)}{2}\right)\\ + \frac{(1-e^{-a_x S})(1-e^{-a_x t})}{a_x S}d_x + \frac{(1-e^{-a_y S})(1-e^{-a_y t})}{a_y S}d_y.\end{aligned}$$

Multiplying by S and writing $\theta_S^F \stackrel{!}{=} \mathbb{E}_{\mathbb{P}}[i_t^S]$ gives

$$\begin{aligned}S\theta_S^F - \log\left(\frac{P^M(0,t)}{P^M(0,t+S)}\right) - \frac{V(t+S)-V(t)-V(S)}{2}\\ = A_S(t)d_x + B_S(t)d_y,\end{aligned}$$

with

$$A_S(t) := \frac{(1-e^{-a_x S})(1-e^{-a_x t})}{a_x}, \qquad B_S(t) := \frac{(1-e^{-a_y S})(1-e^{-a_y t})}{a_y}.$$

Stacking two such equations yields exactly the 2×2 system used in the main text.

Two-level step functions For

$$\lambda_t^x = \mathbb{1}_{\{t\le\tau\}}d_x + \mathbb{1}_{\{t>\tau\}}l_x, \qquad \lambda_t^y = \mathbb{1}_{\{t\le\tau\}}d_y + \mathbb{1}_{\{t>\tau\}}l_y,$$

the expected spot rate becomes

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[i_t^S] = \frac{1}{S}\left(\log\left(\frac{P^M(0,t)}{P^M(0,t+S)}\right) + \frac{V(t+S)-V(t)-V(S)}{2}\right)\\ + \frac{A_S^1(t)}{S}d_x + \frac{A_S^2(t)}{S}l_x + \frac{B_S^1(t)}{S}d_y + \frac{B_S^2(t)}{S}l_y,\end{aligned}$$

where

$$\begin{aligned}A_S^1(t) &:= \frac{(1-e^{-a_x S})(e^{-a_x(t-\min\{t,\tau\})} - e^{-a_x t})}{a_x}, A_S^2(t) := \frac{(1-e^{-a_x S})(1-e^{-a_x(t-\min\{t,\tau\})})}{a_x},\\ B_S^1(t) &:= \frac{(1-e^{-a_y S})(e^{-a_y(t-\min\{t,\tau\})} - e^{-a_y t})}{a_y},\\ B_S^2(t) &:= \frac{(1-e^{-a_y S})(1-e^{-a_y(t-\min\{t,\tau\})})}{a_y}.\end{aligned}$$

Thus every forecast condition is affine in (d_x, l_x, d_y, l_y), and stacking four conditions yields the 4×4 system of the chapter.

Step functions with n levels For the general stepwise specification, define the coefficient kernels

$$\begin{aligned}\Delta_{x,1}(t) &:= e^{-a_x(t-\min\{t,\tau_1\})} - e^{-a_x t},\\ \Delta_{x,j}(t) &:= e^{-a_x(t-\min\{t,\tau_j\})} - e^{-a_x(t-\min\{t,\tau_{j-1}\})}, \qquad j = 2,\dots,n-1,\\ \Delta_{x,n}(t) &:= 1 - e^{-a_x(t-\min\{t,\tau_{n-1}\})},\end{aligned}$$

and analogously $\Delta_{y,j}(t)$ with a_x replaced by a_y. Then

$$S\theta_S^F - \left(\int_t^{t+S}\psi_u\,\mathrm{d}u - \frac{V(S)}{2}\right) = \frac{1-e^{-a_x S}}{a_x}\sum_{j=1}^n \Delta_{x,j}(t)\lambda_{x,j} + \frac{1-e^{-a_y S}}{a_y}\sum_{j=1}^n \Delta_{y,j}(t)\lambda_{y,j}.$$

Collecting $2n$ forecast conditions therefore yields one linear system in the $2n$ unknowns $(\lambda_{x,1},\dots,\lambda_{x,n},\lambda_{y,1},\dots,\lambda_{y,n})$.

Iterative construction The same algebra also explains the iterative calibration used in the excess-return work. Once the first $(j-1)$ blocks have been identified, their contribution is moved to the right-hand side, and the next pair $(\lambda_{x,j}, \lambda_{y,j})$ is determined from a residualized 2×2 system built from two forecast anchors on the j-th block. In other words, the high-dimensional problem can be decomposed into a sequence of low-dimensional identification steps. This is the technical basis of the iterative construction discussed in Chapter 8.

B.5 Representative risk-premium calibration slices

Table B.1 reports representative calibration slices across forecast families and premium specifications. The table documents the sensitivity of the calibrated mean-reversion levels to the selected target architecture.

Forecast family	Premium specification	Evidence window	Date	λ_x	λ_y	Reading
3M/10Y forwards after 1Y	Constant pair	1268 dates, 2018–2023	2023-12-29	−0.58%	−0.20%	Sparse short-horizon target; both factor premia are mildly negative on the latest date.
3M/10Y forwards after 10Y	Constant pair	1268 dates, 2018–2023	2023-12-29	−2.90%	2.69%	Longer-horizon target shifts the adjustment into an offsetting factor allocation.
Annualized forwards	Constant pair	1268 dates, 2018–2023	2023-12-29	0.48%	−0.87%	The annualized construction alters both the level and the sign pattern.
OIS 1Y	Constant pair	1687 dates, 2016–2023	2023-12-29	−0.54%	0.08%	Short OIS target produces a small local adjustment close to the openIRM-style one-year use case.
OIS 20Y	Constant pair	1687 dates, 2016–2023	2023-12-29	−0.43%	−0.01%	Longer OIS target remains moderate on the latest date, but the historical evidence differs materially from the 1Y target.
EIOPA 3M/10Y forwards after 20Y	Constant pair	8 dates, 2017–2023	2023-12-28	−3.13%	1.53%	Regulatory long-end targets can require much stronger opposing factor premia.
COVID OIS short/long targets	Two-level, short block	2020-03-18 case	2020-03-18	−3.48%	0.07%	The short block preserves the stressed near-term COVID rate regime.
COVID OIS short/long targets	Two-level, long block	2020-03-18 case	2020-03-18	2.27%	−1.19%	The long block moves the same case toward the prescribed long-run forecast geometry.

Table B.1: Representative forecast-family and premium-specification slices for the factor-specific real-world premia.

C Technical Status of Swaption Diagnostics in the G2++ Model

The main diagnostic benchmark uses caps, where the G2++ model gives an analytical pricing formula and hence an analytical Jacobian. Swaptions provide an instrument-class extension of the same local WLS geometry. In this case, however, the Jacobian used for the empirical diagnostics is numerical. Its construction and the limits of the resulting diagnostics therefore have to be stated explicitly. We do not repeat the pricing derivation, because the semi-analytical pricing formula is already stated in Section 4.3.1.

C.1 Swaption map and parameter dependence

For the diagnostic framework, the relevant object is the parameter-dependent pricing map

$$\Pi \longmapsto \mathsf{Swpt}_k(\Pi),$$

evaluated on the swaption instrument set. Here k indexes the swaption price after fixing the term structure, expiry, swap tenor, strike, notional, option type, and payment schedule. The local WLS diagnostics require the Jacobian of this pricing map at the calibrated parameter vector $\hat{\Pi}$.

The G2++ swaption price used in the main text is semi-analytical. It is written as a one-dimensional integral whose integrand contains deterministic cash-flow coefficients c_i, fixed by the payment schedule, and the implicit exercise-boundary component $\bar{y}(x;\Pi)$. Compared with the cap case, the parameter vector enters the swaption pricing map in more places. It enters not only through variance terms, but also through the Gaussian density, conditional normal terms, bond-price coefficients, and the numerical solution of the exercise-boundary equation.

This is a diagnostic difference, not a different WLS theory. Once a reliable Jacobian has been obtained, the same weighted tangent-space objects are available. What changes is the status of the Jacobian. Caps provide the analytical benchmark. Swaptions provide a numerical extension based on the implemented semi-analytical pricer.

The fixed-instrument reading is the same as in the cap case. In the local diagnostic calculation, the swaption expiry, underlying swap tenor, strike convention, term structure, notional, option type, and payment schedule are treated as part of the realised instrument design. The derivative is taken only with respect to the G2++ parameter vector. The swaption design has more moving parts at the quote-conversion and pricing stage, but the statistical interpretation remains a conditional fixed-design interpretation. This appendix therefore identifies the pricing map whose local derivative is needed once the market-convention layer has already been fixed.

C.2 Why an analytical Jacobian is not used

In principle, the swaption Jacobian can be obtained by differentiating the one-dimensional pricing integral. If $\mathcal{I}_k(x;\Pi)$ denotes the parameter-dependent integrand of the k-th swaption price, then formally

$$\frac{\partial}{\partial \Pi_p}\,\mathsf{Swpt}_k(\Pi) = \frac{\partial}{\partial \Pi_p}\int_{\mathbb{R}} \mathcal{I}_k(x;\Pi)\,\mathrm{d}x.$$

Under local smoothness and domination conditions, the derivative may be moved under the integral sign.

This analytical route is not used for the empirical diagnostics. The resulting expression would be long and would add little to the interpretation of the calibration. The parameter vector enters the integrand through the Gaussian density, variance and covariance terms, bond-price coefficients, payoff terms, and the implicit exercise boundary $\bar{y}(x;\Pi)$. Since this boundary is defined by a root equation, its parameter derivative would have to be obtained from the implicit function theorem. This introduces additional nested terms for every integration point.

The relevant conclusion is therefore limited. An analytical Jacobian is possible in principle, but it is not a useful object for the diagnostic comparison carried out here. The empirical swaption diagnostics therefore use a numerically checked Jacobian of the implemented semi-analytical pricing routine.

C.3 Numerical Jacobian used for the diagnostic comparison

The empirical swaption diagnostics treat the implemented G2++ swaption pricer as the map

$$g(\Pi) = \big(\mathsf{Swpt}_1(\Pi),\ldots,\mathsf{Swpt}_m(\Pi)\big)^{\top}.$$

For parameter component p, the k-th Jacobian entry at the calibrated parameter vector is approximated by the central finite difference

$$\frac{\partial\,\mathsf{Swpt}_k}{\partial \Pi_p}(\hat{\Pi}) \approx \frac{\mathsf{Swpt}_k(\hat{\Pi}+h_p e_p)-\mathsf{Swpt}_k(\hat{\Pi}-h_p e_p)}{2h_p}.$$

The step size h_p is parameter-specific. Each perturbed value is a full revaluation of the swaption pricing routine, including the one-dimensional integration and the numerical solution of the exercise-boundary equation. Away from active parameter bounds and for a sufficiently smooth pricing map, the central difference satisfies

$$\frac{\mathsf{Swpt}_k(\hat{\Pi}+h_p e_p)-\mathsf{Swpt}_k(\hat{\Pi}-h_p e_p)}{2h_p} = \frac{\partial\,\mathsf{Swpt}_k}{\partial \Pi_p}(\hat{\Pi}) + O(h_p^2),$$

provided the two prices are evaluated exactly. In the present application, this formal order is not the only relevant error source, because the two perturbed prices are themselves numerical outputs from integration and root finding.

The numerical experiment therefore has to control the finite-difference step, the pricing accuracy, and the admissibility of the perturbed parameters. Step sizes are checked for local stability. Integration tolerances and root-finding tolerances have to be tight enough that their numerical error is not amplified by the division by $2h_p$. If a perturbation would leave the admissible interval of the corresponding parameter, the perturbation must be clipped, replaced by a one-sided local difference, or carried out in a transformed coordinate. This issue is especially relevant for the correlation parameter near its boundary $\rho_{xy} = \pm 1$. The resulting matrix is also checked for

rank and conditioning, because leverage, influence, and effective degrees of freedom depend on products and pseudo-inverses of $\widehat{J}^\top W \widehat{J}$.

The diagnostic object is therefore the checked numerical Jacobian $\widehat{J}$, not the raw finite-difference output before numerical checks. A swaption diagnostic experiment should record the perturbation rule, the pricing tolerances, the treatment of active bounds, and the conditioning of $\widehat{J}^\top W \widehat{J}$. It should also separate pricing failures from diagnostic findings.

After the numerical Jacobian has been constructed, the diagnostic formulas are unchanged. With

$$X = W^{1/2}\widehat{J},$$

the weighted hat matrix is

$$\widehat{H}_w = W^{1/2}\widehat{J}\left(\widehat{J}^\top W \widehat{J}\right)^{+}\widehat{J}^\top W^{1/2}.$$

As before, leverage is read from the diagonal of this matrix, the effective degrees of freedom are its trace, and the influence vectors use the same WLS linearization as in the cap case. The swaption diagnostics are therefore interpreted as a numerically controlled instrument-class extension, while the main empirical claims remain based on the analytical cap benchmark.

D Boundary-Dominated Leverage in the G2++ Cap Calibration

This appendix records the technical details behind the boundary-leverage studies in Section 7.2.14. The main text states the claim, gives the interpretation, and reports the numerical evidence in Table 7.2. Here, we give the supporting algebra and definitions: the reconstruction identity, the row-normalised cumulative G2++ benchmark rows, and the grid and boundary-gap quantities used in the verification.

D.1 Exact reconstruction interpretation of leverage

This section proves the reconstruction identity stated in Proposition 7.2 in Section 7.2.8. The definition of the minimum-norm reconstruction cost is a direct Moore–Penrose projection interpretation of the weighted local design matrix used there.

Let $X \in \mathbb{R}^{m\times p}$ be this weighted local design matrix, for example $X = W^{1/2}J(\hat{\Pi})$. Its rows are denoted by $x_1^\top, \dots, x_m^\top$. Assume first that X has full column rank. The symmetric hat matrix is

$$H_X = X(X^\top X)^{-1}X^\top,$$

and the leverage score of row n is

$$h_n = x_n^\top (X^\top X)^{-1} x_n.$$

Delete row n and write X_{-n} for the remaining design. Define the minimum-norm reconstruction cost

$$\mu_n := \inf\left\{ \|a\|_2^2 : a \in \mathbb{R}^{m-1}, \quad X_{-n}^\top a = x_n \right\},$$

with $\mu_n = +\infty$ if exact reconstruction is impossible.

Proposition: leverage as reconstruction difficulty. If X has full column rank, then

$$h_n = \frac{\mu_n}{1+\mu_n},$$

where the right-hand side is interpreted as 1 if $\mu_n = +\infty$.

Proof. First assume that X_{-n} has full column rank and set

$$A_n = X_{-n}^\top X_{-n}.$$

Then A_n is positive definite and

$$X^\top X = A_n + x_n x_n^\top.$$

By the Sherman–Morrison formula,

$$(A_n + x_n x_n^\top)^{-1} = A_n^{-1} - \frac{A_n^{-1} x_n x_n^\top A_n^{-1}}{1 + x_n^\top A_n^{-1} x_n}.$$

Hence

$$\begin{aligned} h_n &= x_n^\top (A_n + x_n x_n^\top)^{-1} x_n \\ &= x_n^\top A_n^{-1} x_n - \frac{x_n^\top A_n^{-1} x_n x_n^\top A_n^{-1} x_n}{1 + x_n^\top A_n^{-1} x_n}. \end{aligned}$$

With

$$q_n := x_n^\top A_n^{-1} x_n,$$

this becomes

$$h_n = q_n - \frac{q_n^2}{1 + q_n} = \frac{q_n}{1 + q_n}.$$

Therefore

$$h_n = \frac{x_n^\top A_n^{-1} x_n}{1 + x_n^\top A_n^{-1} x_n}.$$

The minimum-norm solution of $X_{-n}^\top a = x_n$ is

$$a^* = X_{-n} A_n^{-1} x_n,$$

and hence

$$\mu_n = \|a^*\|_2^2 = x_n^\top A_n^{-1} x_n.$$

Substitution gives the identity.

If X_{-n} is not full column rank while X is full column rank, deleting row n reduces the row span. Then x_n cannot be reconstructed from the remaining rows, so $\mu_n = +\infty$. In this case the leave-one-out rank drops and $h_n = 1$, which is consistent with the stated convention. □

Since the map $\mu \mapsto \mu/(1+\mu)$ is increasing on $[0, \infty]$, leverage ranks rows in the same order as their minimum-norm reconstruction costs.

D.2 Why no unconditional U-shape claim is made

The weighted rows entering the hat matrix are $x_n = \sqrt{w_n} j_n$. Hence row scaling affects leverage, which rules out an unconditional model-only theorem for boundary dominance.

Lemma: no unconditional boundary theorem under arbitrary row scaling. Let $X \in \mathbb{R}^{m \times p}$ have full column rank. Fix an interior row $n \in \{2, \dots, m-1\}$, and assume that X_{-n} has full column rank and $x_n \neq 0$. If row n is replaced by the scaled row αx_n, with $\alpha > 0$, while all other rows are kept fixed, then the leverage score of the rescaled row is

$$h_n^{(\alpha)} = \frac{\alpha^2 u_n}{1 + \alpha^2 u_n}, \qquad u_n = x_n^\top (X_{-n}^\top X_{-n})^{-1} x_n > 0.$$

Consequently $h_n^{(\alpha)} \to 1$ as $\alpha \to \infty$.

Proof. Set

$$A_n = X_{-n}^\top X_{-n}.$$

By assumption, A_n is positive definite. After replacing row n by the scaled row αx_n, the cross-product matrix becomes

$$(X^{(\alpha)})^\top X^{(\alpha)} = A_n + \alpha^2 x_n x_n^\top.$$

The leverage score of the rescaled row is therefore

$$h_n^{(\alpha)} = (\alpha x_n)^\top \left(A_n + \alpha^2 x_n x_n^\top\right)^{-1} (\alpha x_n).$$

Applying the Sherman–Morrison formula gives

$$\left(A_n + \alpha^2 x_n x_n^\top\right)^{-1} = A_n^{-1} - \frac{\alpha^2 A_n^{-1} x_n x_n^\top A_n^{-1}}{1 + \alpha^2 x_n^\top A_n^{-1} x_n}.$$

With

$$u_n = x_n^\top A_n^{-1} x_n,$$

we obtain

$$\begin{aligned} h_n^{(\alpha)} &= \alpha^2 x_n^\top A_n^{-1} x_n - \frac{\alpha^4 x_n^\top A_n^{-1} x_n x_n^\top A_n^{-1} x_n}{1 + \alpha^2 x_n^\top A_n^{-1} x_n} \\ &= \alpha^2 u_n - \frac{\alpha^4 u_n^2}{1 + \alpha^2 u_n} \\ &= \frac{\alpha^2 u_n}{1 + \alpha^2 u_n}. \end{aligned}$$

Since $u_n > 0$, it follows that $h_n^{(\alpha)} \to 1$ as $\alpha \to \infty$. □

Thus, an interior row can be made arbitrarily high-leverage by row scaling alone. The empirical boundary result is therefore formulated for the realised and controlled benchmark designs, not as a theorem for all admissible designs.

D.3 Cumulative G2++ benchmark rows

The finite-grid verification in Table 7.2 is not carried out on the full weighted Jacobian. It uses the row-normalised cumulative G2++ benchmark described in the main text. This section defines the rows of that benchmark.

The price sensitivity factorisation used in Section 7.2.6 gives

$$\frac{\partial \,\mathsf{Caplet}_i}{\partial p} = \mathcal{M}_i(\Pi) \frac{\partial \Sigma_i^2}{\partial p}, \qquad \mathcal{M}_i(\Pi) > 0.$$

The benchmark keeps only the variance-sensitivity part and removes WLS weights, strike-specific price-to-variance factors, and pure row norms.

Under the six-month tenor convention, let $\delta = 1/2$ denote the accrual period and introduce

$$\chi_x = 2a_x, \qquad \chi_y = 2a_y, \qquad \chi_{xy} = a_x + a_y = \frac{\chi_x + \chi_y}{2},$$

and

$$c_x = \frac{1 - e^{-a_x \delta}}{a_x}, \qquad c_y = \frac{1 - e^{-a_y \delta}}{a_y},$$

$$A_x = \frac{\sigma_x^2 c_x^2}{2a_x}, \quad A_y = \frac{\sigma_y^2 c_y^2}{2a_y}, \quad A_{xy} = \frac{2\rho_{xy}\sigma_x \sigma_y c_x c_y}{a_x + a_y}.$$

With the vector

$$\vartheta = (A_x, A_y, A_{xy}, \chi_x, \chi_y)^\top,$$

the caplet variance at T is given by

$$\Sigma^2(T;\vartheta) = A_x(1 - e^{-\chi_x T}) + A_y(1 - e^{-\chi_y T}) + A_{xy}(1 - e^{-\chi_{xy} T}).$$

The corresponding variance-sensitivity vector is

$$\Gamma_\vartheta(T) = \begin{pmatrix} 1 - e^{-\chi_x T} \\ 1 - e^{-\chi_y T} \\ 1 - e^{-\chi_{xy} T} \\ A_x T e^{-\chi_x T} + \frac{A_{xy}}{2} T e^{-\chi_{xy} T} \\ A_y T e^{-\chi_y T} + \frac{A_{xy}}{2} T e^{-\chi_{xy} T} \end{pmatrix}.$$

This parametrisation is used only for the local benchmark construction and not as a global reparametrisation of the closed G2++ parameter domain.

For cap maturity τ_n, define the cumulative benchmark row and its row-normalised version by

$$z_n(\vartheta) = \sum_{i \in \mathcal{I}_n} \Gamma_\vartheta(t_{i-1}), \qquad \bar{x}_n(\vartheta) = \frac{z_n(\vartheta)}{\|z_n(\vartheta)\|_2}.$$

Putting these rows together gives $\bar{X}(\vartheta)$. Its projector is

$$\bar{H}(\vartheta) = \bar{X}(\vartheta)\big(\bar{X}(\vartheta)^\top \bar{X}(\vartheta)\big)^{+} \bar{X}(\vartheta)^\top,$$

and the benchmark boundary gap used in the verification is

$$\bar{\eta}(\vartheta) = \max\{\bar{h}_1(\vartheta), \bar{h}_m(\vartheta)\} - \max_{2 \le n \le m-1} \bar{h}_n(\vartheta).$$

If $\bar{\eta}(\vartheta) > 0$, the row-normalised cumulative benchmark is boundary dominated at ϑ.

D.4 Verification setup and boundary-gap definitions

The benchmark verification summarised in Table 7.2 uses the maturity grid

$$(3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30)$$

with six-month tenor. The mean-reversion grid is the equidistant grid

$$a_x, a_y \in \mathcal{A}, \qquad \mathcal{A} = \left\{ 0.02 + \ell \frac{1.50 - 0.02}{7} : \ell = 0, \dots, 7 \right\},$$

that is approximately

$$\mathcal{A} \approx \{0.02, 0.23, 0.44, 0.65, 0.87, 1.08, 1.29, 1.50\}.$$

The volatility and correlation grids are

$$\sigma_x, \sigma_y \in \{0.002, 0.005, 0.010, 0.020, 0.050\},$$

and

$$\rho_{xy} \in \{-0.95, -0.75, -0.50, 0, 0.50, 0.75, 0.95\}.$$

The finite benchmark grid therefore contains $8^2 \cdot 5^2 \cdot 7 = 11200$ candidate points. Candidate points that violate the nondegenerate benchmark construction are excluded, in particular the near-coincidence case $|a_x - a_y| < 10^{-3}$. This removes the $a_x = a_y$ grid diagonal and leaves $11200 - 8 \cdot 5^2 \cdot 7 = 9800$ valid benchmark points.

For the row-normalised benchmark, the reported boundary gap is $\bar{\eta}(\vartheta)$. A positive value means that the larger of the two endpoint leverages is above every interior maturity leverage. Thus the positive minimum over the valid grid checks boundary dominance after row normalisation. The local-search value checks the same quantity in neighbourhoods of the worst finite-grid points.

For realised calibration designs, define the ordinary boundary gap

$$\Delta_{\mathrm{int}}(t) = \max\{h_1(t), h_m(t)\} - \max_{2 \le n \le m-1} h_n(t).$$

A positive value means that, on day t, the largest leverage score is attained at one of the two boundary maturities rather than at an interior maturity.

The empirical verification separates days with interior correlation estimates from days where $\hat{\rho}_{xy} = -1$ is active. The ordinary full-Jacobian Moore–Penrose diagnostic keeps all five Jacobian columns and uses the pseudo-inverse for singular directions. On the lower-boundary days, a stricter active-set diagnostic is also computed by removing the active correlation direction from the local tangent space. Its boundary gap is denoted by $\Delta_{\mathrm{int}}^{\mathcal{A}}$.

For the lower-boundary days, the additional active-set summary quantities are

$$\min \Delta_{\mathrm{int}}^{\mathcal{A}} = -0.107366, \qquad \text{median } \Delta_{\mathrm{int}}^{\mathcal{A}} = 0.484407,$$

and an endpoint maturity remains the leverage maximum on a share of 0.960894. The negative minimum shows that the stricter active-set diagnostic is not boundary dominated on every lower-boundary day. The positive median shows that the typical active-set day remains strongly boundary dominated. The share 0.960894 records the proportion of lower-boundary days on which one of the two endpoint maturities still has the largest leverage after the active correlation direction has been removed. These quantities support the ordinary full-Jacobian diagnostic.

E Statistical Background of the Local Diagnostic Approximation

The key-concept sections in Chapter 7 introduce the general nonlinear-regression, asymptotic-statistical, and functional-delta-method terminology used in the diagnostic analysis; see Sections 7.2.1, 7.2.2, and 7.2.3. The WLS formulation, the tangent-space geometry, the Gauss–Newton adequacy check, and the main diagnostic tools are developed in Sections 7.2.4, 7.2.5, 7.2.7, and 7.2.8.

Consequently, this appendix fixes the statistical terminology used in Section 7.2 for one realised daily WLS calibration problem. The estimator and functional-delta-method terminology follows the standard asymptotic-statistical usage of [89]; the nonlinear least-squares terminology follows [85], and the influence terminology follows [61, 57]. The terminology is used for a local diagnostic interpretation of the weighted calibration problem and does not specify a full stochastic error model for market quotes. In this setting, the weighted nonlinear least-squares solution can be read locally as an M-estimator and, through its first-order condition, as a Z-estimator. Terms such as score, influence, covariance, Hadamard differentiability, and the Functional Delta Method are used in this local diagnostic sense. The Fisher-information-type matrix describes local curvature and sensitivity of the weighted pricing map. The covariance approximation gives the corresponding first-order uncertainty measure. Both objects are used diagnostically, without specifying a likelihood model for market-price errors.

E.1 Conditional daily design

For a fixed calibration date, the maturity grid, market conventions, discount curve, strikes, notionals, quote conversion, and weights define the realised daily design in the fixed-design sense of regression diagnostics. Let

$$y = (y_1^{\mathsf{Mkt}}, \dots, y_m^{\mathsf{Mkt}})^\top, \qquad g(\Pi) = (g_1(\Pi), \dots, g_m(\Pi))^\top,$$

and let $W = \mathsf{diag}(w_1, \dots, w_m)$. The calibrated parameter vector is

$$\hat{\Pi} = \arg\min_{\Pi \in \mathcal{D}} Q_m(\Pi), \qquad Q_m(\Pi) = (y - g(\Pi))^\top W (y - g(\Pi)).$$

The diagnostic quantities in Chapter 7 are computed conditional on the realised daily instrument set, the market conventions, the curve construction, and the weights. They describe the local geometry of the realised weighted pricing map g on this fixed design. They do not require the maturities on a single day to be treated as i.i.d. observations from a population of possible instruments.

This fixed-design reading is also why the WLS weights are kept inside the diagnostic object. The matrix W defines the price-space metric induced by the RMSRE objective. A different

weight matrix would define a different local calibration target and therefore a different diagnostic geometry.

E.2 M- and Z-estimator reading of the WLS optimum

In the terminology used here, an M-estimator is an estimator defined as a minimiser of a criterion function. For a parameter $\theta \in \Theta$ and a criterion $M_m : \Theta \to \mathbb{R}$, this means

$$\hat{\theta} \in \operatorname*{arg\,min}_{\theta \in \Theta} M_m(\theta).$$

A Z-estimator is an estimator defined as a zero of an estimating equation. For an estimating function $\Psi_m : \Theta \to \mathbb{R}^p$, this means

$$\Psi_m(\hat{\theta}) = 0.$$

For differentiable criterion functions, the two descriptions are linked by the first-order condition: an interior locally regular minimiser of M_m solves the corresponding estimating equation given by the gradient of the criterion.

With $M_m = Q_m$, the WLS optimum can therefore be read as an M-estimator because it minimises the criterion Q_m. At an interior locally regular solution, the same estimator can also be read through its first-order condition. Define the residual vector

$$r(\Pi) = y - g(\Pi),$$

and let $J(\Pi)$ denote the Jacobian of the pricing map. Differentiating the WLS criterion gives

$$\nabla Q_m(\Pi) = -2J(\Pi)^\top W r(\Pi).$$

Thus, at an interior optimum, the fitted parameter satisfies the weighted normal equation

$$J(\hat{\Pi})^\top W r(\hat{\Pi}) = 0.$$

Equivalently, the WLS calibration can be read as a Z-estimator with estimating function

$$\Psi_m(\Pi) = J(\Pi)^\top W r(\Pi).$$

The calibrated parameter is then a zero of this first-order condition. The M- and Z-estimator descriptions give the objective-function and first-order-condition views of the same WLS optimum.

The derivative of this first-order condition is the local sensitivity matrix. Differentiating exactly would also produce terms involving residuals and second derivatives of g. Section 7.2.7 shows empirically that these residual-weighted second-order terms are negligible for the cap calibration examined here. At the diagnostic level used in Chapter 7, the sensitivity matrix is therefore read as

$$-J(\hat{\Pi})^\top W J(\hat{\Pi}).$$

This is the formal reason why the same matrix $J^\top W J$ appears in the local curvature, rank, leverage, influence, and covariance diagnostics. In rank-deficient cases, the corresponding Moore–Penrose inverse is used consistently with the weighted hat matrix in Section 7.2.5.

E.3 Influence, covariance, and transformed uncertainty

The influence and covariance expressions used in Chapter 7 are read here as local sensitivity formulae for the realised WLS problem.

For a functional T, the Gâteaux derivative at F in a direction h is the directional derivative

$$\dot{T}_F(h) = \lim_{\varepsilon \downarrow 0} \frac{T(F + \varepsilon h) - T(F)}{\varepsilon},$$

whenever this limit exists. Hadamard differentiability strengthens this requirement by requiring stability of the derivative along converging perturbation directions. More precisely, if $t_n \downarrow 0$, $h_n \to h$, and $F + t_n h_n$ belongs to the domain of T, then

$$\frac{T(F + t_n h_n) - T(F)}{t_n} \to \dot{T}_F(h).$$

In this terminology, the influence function is the directional derivative along the contamination path $(1 - \varepsilon)F + \varepsilon\delta_z$, equivalently in the signed direction $h_z = \delta_z - F$.

Under the local linearization

$$g(\Pi) \approx g(\hat{\Pi}) + J(\hat{\Pi})(\Pi - \hat{\Pi}),$$

a small perturbation of the price vector moves the fitted parameter vector through the inverse (or pseudo-inverse) of the local sensitivity matrix. Using the residual notation from above,

$$r(\Pi) = y - g(\Pi),$$

the local influence vector for the k-th quote is

$$\mathrm{IF}_k(\hat{\Pi}) = r_k(\hat{\Pi})\big(J(\hat{\Pi})^\top W J(\hat{\Pi})\big)^+ J_k(\hat{\Pi})^\top w_k.$$

Here $J_k(\hat{\Pi})$ denotes the k-th row of the Jacobian. The vector measures the local parameter displacement associated with the realised residual of quote k. The influence score plotted in Chapter 7 is the Euclidean norm of this vector.

The covariance approximation has the same local status. Conditional on the realised daily design and under the relative-error scale used for the WLS criterion, the local diagnostic covariance estimate is

$$\widehat{\mathrm{Cov}}(\hat{\Pi}) \approx \hat{\sigma}^2 \big(J(\hat{\Pi})^\top W J(\hat{\Pi})\big)^+.$$

The scale $\hat{\sigma}$ is estimated in the main text by the robust residual scale chosen for the cap panel. This expression is used as a local diagnostic approximation of how the WLS solution reacts to small price perturbations in the cap calibration examined here. It does not specify a likelihood-based covariance model for quote errors.

The functional delta method enters only at the last step. Once a differentiable transformation or scalar diagnostic $\eta(\Pi)$ is applied to the calibrated parameter, the local covariance is propagated by the derivative at $\hat{\Pi}$:

$$\widehat{\mathrm{Cov}}\{\eta(\hat{\Pi})\} \approx D\eta(\hat{\Pi})\, \widehat{\mathrm{Cov}}(\hat{\Pi})\, D\eta(\hat{\Pi})^\top.$$

For the boundary-respecting parameter intervals in Chapter 7, η is the componentwise transformation that maps positive parameters to log coordinates and the correlation coordinate to Fisher-z coordinates. Intervals are constructed in the transformed coordinates and mapped back. This does not change the calibration estimate; it only expresses local uncertainty on a scale compatible with the constrained parameter domain.