ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Convergence Theory for Expected Signature Estimation from Dependent Single Paths with Applications to Parameter Calibration

Master Thesis

Bryson Schenck

Wednesday 15th October, 2025

Advisor: Prof. Dr. Josef Teichmann

Department of Mathematics, ETH Zürich

Abstract

This thesis develops a convergence theory for empirical expected-signature estimators from single-path data under a segment-stationarity assumption (short path segments are shift-invariant in distribution with exponentially decaying serial dependence) and applies it to parameter calibration of two-dimensional Ornstein–Uhlenbeck processes. A segmentation and reindexing procedure is introduced that retains mean reversion and serial dependence across blocks without assuming block independence, yielding an estimator with finite-sample mean-squared-error convergence at rate $O(N^{-2/p})$ under this segment-stationarity framework combined with exponential α-mixing. Here N is a granularity parameter controlling block size and $p > 2$ denotes path regularity. Validation uses a generator-based framework that reduces expected-signature computation in linear SDEs to matrix exponentials, enabling precise numerical verification. Across parameter regimes, log–log MSE-versus-N slopes between -1.86 and -3.61 are observed. For calibration, we develop a systematic hyperparameter optimization framework and compare signature-based methods against Batched MLE, a block-averaged maximum likelihood approach where parameters are estimated on multiple path segments and then averaged. In slow mean-reversion regimes, signature methods achieve 10–32% improvement over Batched MLE, with accuracy differential scaling positively with path volatility; these gains are statistically significant across the full range (Wilcoxon signed-rank test, $p < 0.001$). Notably, in the same slow-reversion regimes, signature methods exhibit superior computational efficiency (9–15% speedup), with per-iteration optimization cost demonstrating more favorable asymptotic scaling for high-frequency time series than Batched MLE. In such cases, GPU memory rather than computation time becomes the primary constraint on extending the method to longer paths. The results establish a quantitative convergence rate for empirical expected-signature estimation in serially dependent single-path settings and demonstrate clear advantages in both statistical accuracy and computational efficiency for calibration.

Contents

Chapter 1

Introduction

1.1 Mathematical Context and Motivation

The calibration of stochastic differential equation (SDE) parameters from observed path data is a central task in quantitative finance, underpinning applications in option pricing, risk management, and portfolio construction. The standard approach relies on maximum likelihood estimation (MLE), which is statistically efficient in classical settings but can become computationally demanding for high-frequency or high-dimensional data due to repeated likelihood evaluations across many observation batches and the need to handle temporal dependence carefully.

Signature methods, originating from rough path theory [27], offer an alternative pathwise representation. The *signature transform* encodes a trajectory into a graded sequence of iterated integrals that jointly capture magnitudes and temporal ordering. Under appropriate conditions, the *expected signature* characterises the law of a process [9], which makes it a natural target for calibration. In contrast to likelihood formulations that emphasise low-order increment statistics, signature-based calibration aligns a richer family of path statistics by matching truncated expected signatures.

From a computational perspective, both MLE and signature-based calibration require per-iteration evaluations of model-implied quantities and their sensitivities. In our implementation, analytical expected signatures and their gradients are recomputed at each optimisation step. These computations are regular, batchable, and GPU-amenable. We therefore regard the approach as computationally competitive, with empirical runtime and accuracy trade-offs quantified in Section 6.8. The focus of this thesis is to establish the statistical foundations for signature methods and to demonstrate when and why the richer path statistics translate into improved calibration accuracy.

1.2 Mathematical Foundations

We recall the algebraic structures used throughout the thesis.

1.2.1 Tensor Algebras

Let V be a d-dimensional real vector space. The *tensor algebra* over V is

$$T(V) = \bigoplus_{k=0}^{\infty} V^{\otimes k} = \mathbb{R} \oplus V \oplus V^{\otimes 2} \oplus \cdots,$$

with truncated form

$$T^{(M)}(V) = \bigoplus_{k=0}^{M} V^{\otimes k},$$

where $M \in \mathbb{N}$ is the truncation depth. Throughout, we equip $T^{(M)}(V)$ with the Hilbert–Schmidt inner product induced by a fixed inner product on V.

1.2.2 The Signature Transform

For a continuous path $X \colon [0, T] \to V$ of bounded variation, its *signature* is the sequence of iterated integrals

$$S^{(k)}(X)_{0,T} = \int_{0 < t_1 < \cdots < t_k < T} dX_{t_1} \otimes \cdots \otimes dX_{t_k} \in V^{\otimes k}, \qquad k \geq 1,$$

with $S^{(0)}(X)_{0,T} = 1 \in \mathbb{R}$. The *truncated signature* $S^{(M)}(X)_{0,T} \in T^{(M)}(V)$ collects components up to degree M. For stochastic processes of unbounded variation, these objects are defined via the geometric rough path lift; we recall the necessary continuity and stability properties in Section 3.

1.2.3 Expected Signatures and Statistical Inference

For a stochastic process X with law μ, the *expected signature* is

$$\mathbb{E}\Big[S^{(M)}(X)_{0,T} \Big] = \int S^{(M)}(\omega)_{0,T} \, d\mu(\omega) \ \in \ T^{(M)}(V).$$

Under suitable conditions, expected signatures characterise the process law [9, 20]. This suggests a calibration paradigm in which model parameters are chosen to minimise the discrepancy between empirical and model-implied expected signatures. The central statistical question addressed in this thesis is: for a single, long, dependent path segmented into blocks, at what rate does the empirical expected signature converge to its population counterpart?

1.3 Research Gap and Objectives

While uniqueness and universality properties of signatures are well established, finite-sample convergence guarantees for empirical expected-signature estimators in *dependent, single-path* settings have been lacking. This is a relevant gap for financial applications, where one typically observes a single long time series with serial dependence rather than many independent path realisations. The primary objective of this thesis is to develop a convergence theory tailored to this setting and demonstrate how it supports practical parameter calibration.

1.4 Research Contributions

Contributions at a glance: We establish a convergence rate for signature estimators under dependence, develop an efficient analytical framework for OU processes, and build a calibration methodology that outperforms MLE in key scenarios.

The contributions of this thesis are threefold.

First, we establish a convergence rate for empirical expected-signature estimators constructed from a single, segment-stationary[1], exponentially α-mixing path. Intuitively, this means short path segments have shift-invariant laws, and dependence between far-apart segments decays geometrically. Our approach divides a long continuous path into K_N blocks, where $K_N \asymp N^{1+2/p}$ for a granularity parameter N. Each block has length $\Delta t_N = \delta/N$ (where δ is the total time horizon of a block). We compute the signature (iterated integrals up to order M) on each block, then average these block signatures to estimate the expected signature. Specifically, we prove

$$\mathbb{E}\big\|\widehat{\mathbb{E}}[S^{(M)}]_N - \mathbb{E}[S^{(M)}]\big\|^2 \;=\; O\Big(N^{-2/p}\Big),$$

where $\widehat{\mathbb{E}}[S^{(M)}]_N$ is our empirical expected signature estimator (average of K_N block signatures), $\mathbb{E}[S^{(M)}]$ is the true expected signature, N is the granularity parameter that controls block size (larger N means smaller blocks), and $p > 2$ is the path roughness parameter. The proof combines a bias–variance decomposition with rough-path continuity (for bias control) and covariance inequalities for Hilbert-valued, strongly mixing sequences (for fluctuation control). The result provides a statistical foundation for signature-based inference under dependence.

Second, we develop an analytical framework for Ornstein–Uhlenbeck (OU) processes that reduces the computation of expected signatures to matrix

---

[1]Segment-stationarity requires that finite-length increment segments $(X_{\tau+s} - X_\tau)_{0 \le s \le \Delta}$ have the same law for any shift τ. This is weaker than requiring stationary increments in the Lévy sense.

exponentials of a lifted generator on $T^{(M)}(\mathbb{R}^d)$. For the Stratonovich OU SDE

$$dX_t = \theta(\mu - X_t)\, dt + \sigma \circ dW_t,$$

we show that $\mathbb{E}[S^{(M)}(X)_{0,T}] = \exp\big(T\,\mathcal{G}^{(M)}\big)\,\mathbf{1}$, where $\mathcal{G}^{(M)}$ is the generator lifted to the tensor algebra. This yields machine-precision ground truth for validation and supplies gradients for optimisation. The computations are highly structured and lend themselves to vectorisation and GPU acceleration. Crucially, for processes like Ornstein-Uhlenbeck, signatures computed on small blocks can be analytically rescaled to recover full-horizon signatures, enabling efficient computation.

Third, we build a calibration methodology that leverages these theoretical and computational tools. Through systematic hyperparameter optimisation across four economically motivated parameter regimes, we compare signature-based objectives against Batched MLE, our term for a block-averaged maximum likelihood estimation approach where the path is divided into K segments, MLE is performed on each segment independently, and the resulting parameter estimates are averaged. This method serves as our classical baseline, with K selected algorithmically based on the mean reversion speed. In slow mean-reversion scenarios, precisely where higher-order path features are most informative, signature methods achieve statistically significant improvements in calibration accuracy. To ensure robust hyperparameter selection in finite samples, we introduce a scoring rule

$$\text{Score(config)} = \text{MSE(config)} + \lambda \cdot \text{StdDev(config)},$$

which balances accuracy and stability across configurations. We set $\lambda = 1.0$, as results for top-tier configurations were robust to this choice (e.g., $\lambda \in \{0.5, 1.0, 1.5\}$ produced identical rankings); details are in Chapter 6.

Our results are specific to the two-dimensional OU family considered here. While the convergence theory applies more broadly to segment-stationary mixing processes, the computational framework and calibration gains are demonstrated within this parametric class.

1.5 Thesis Organisation

Chapter 2 reviews related work on stochastic process calibration and signature methods. Chapter 3 presents the convergence theory for empirical expected-signature estimators under dependence. Chapter 4 develops the OU generator framework and its computational realisation. Chapter 5 reports numerical studies validating the convergence theory. Chapter 6 applies the methodology to parameter calibration and compares it against Batched MLE. Chapter 7 concludes with limitations and directions for future work.

The presentation aims to balance mathematical rigour with applicability: readers focused on theory may prioritise Chapters 3 and 4, while practitioners may emphasise Chapters 5 and 6. The sections are self-contained where possible, but the full narrative demonstrates the progression from abstract results to validated, implementable procedures.

Chapter 2

Literature Review and Theoretical Foundations

This chapter surveys the theoretical foundations and recent developments in signature methods for stochastic process analysis. We review five key areas: the representation and uniqueness properties of signatures from rough path theory; expected-signature computation; computational paradigms for model-implied quantities; convergence analysis under dependence; and recent advances in signature-based methods. The thesis contributes to this literature by developing finite-sample convergence rates for empirical expected-signature estimators under dependence (Chapter 3), providing an analytical framework for exact expected-signature computation in linear SDEs (Chapter 4), and establishing a calibration methodology evaluated through systematic experiments (Chapters 5 and 6).

2.1 Rough Paths and Signatures

Lyons introduced the signature of a path as a sequence of iterated integrals that encodes path information in a reparameterisation-invariant manner [27]. The modern theory of rough paths extends signatures to low-regularity signals and establishes continuity properties for controlled differential equations driven by such signals [18, 17].

A central result concerns uniqueness: under appropriate conditions on the underlying measure space, the expected signature determines the law of a stochastic process. Chevyrev and Lyons [9] and Hambly and Lyons [20] establish that for measures supported on point-separating algebras (an algebra of functions that distinguishes distinct points, enabling uniqueness) with suitable integrability conditions, the expected signature uniquely determines the law on path space. Extensions to c'adl'ag processes via Marcus signatures broaden the applicability beyond continuous paths [15]. These results provide

model-free characterisation through the full expected signature. The calibration setting studied in this thesis employs finite truncations within specific model classes, where parameter identifiability requires additional analysis as developed in Chapter 4.

2.2 Expected Signatures in Statistical Inference

The statistical role of signature features has developed along multiple directions. Signature moments characterise process laws analogously to classical moments [11], while signature-based embeddings enable learning algorithms on path spaces [16]. Lucchese, Pakkanen, and Veraart establish asymptotic properties for expected-signature estimators from discretised observations, covering both in-fill and long-span asymptotics under general dependence structures [25]. Related work connects signature asymptotics with transport metrics [19] and derives closed-form expected signatures for broad semimartingale classes [13].

A gap remains for finite-sample guarantees under dependence. Financial time series typically provide a single long dependent trajectory rather than independent realisations. This thesis addresses this gap by proving an $O(N^{-2/p})$ mean-square error bound for block-based estimators under segment stationarity and exponential α-mixing (dependence between events decays geometrically with time lag), where N is a granularity parameter and $p > 2$ characterises path regularity. This finite-sample result, established in Chapter 3, complements existing asymptotic theory.

2.3 Computational Paradigms for Model-Implied Signatures

Two strategies dominate the computational literature. Simulation-based approaches approximate expected signatures through discretisation and Monte Carlo methods, offering flexibility at the cost of discretisation and sampling errors. Analytical approaches exploit model structure to compute expected signatures exactly. Linear signature models enable efficient calibration within restricted classes [12]. For linear diffusions, generator methods reduce expected-signature computation to matrix exponentials on truncated tensor algebras, yielding machine-precision results. This thesis adopts the analytical approach for Ornstein–Uhlenbeck processes (Chapter 4), providing exact ground truth via semigroup methods on $T^{(M)}(\mathbb{R}^d)$ and enabling gradient-based calibration without numerical approximation errors.

2.4 Inference Under Dependence: Mixing and Block Estimators

When observing a single dependent trajectory, classical independent-sample theory does not apply. Mixing conditions provide the framework for analysing dependent sequences through covariance inequalities and limit theorems [30, 6]. Block-based estimators partition long trajectories to extract approximately independent statistics. Previous signature work addresses asymptotic properties under various dependence structures [25] but does not provide finite-sample rates.

This thesis develops block estimators for segment-stationary processes (Assumption (S)), a natural class for mean-reverting diffusions. The analysis combines rough-path continuity for bias control with covariance inequalities under exponential α-mixing for variance bounds. The resulting $O(N^{-2/p})$ convergence rate connects path regularity and dependence strength to estimation accuracy, with empirical validation in Chapter 5 demonstrating behaviour exceeding theoretical predictions.

2.5 Recent Developments and Positioning

Signature methods have expanded into machine learning through neural controlled differential equations, providing principled feature extraction for irregular time series [24, 29, 28]. Financial applications include optimal execution algorithms [1, 23] and rough volatility modelling [3, 4]. Methodological advances continue through developments in rough differential equations [2, 8] and comprehensive surveys [26, 10].

The literature thus provides theoretical foundations through rough path theory, asymptotic statistical properties, and diverse computational strategies. This thesis contributes finite-sample guarantees under dependence, exact computational methods for linear SDEs, and systematic empirical evaluation of calibration performance. Together, these advances establish signature methods as a practical tool for parameter estimation in stochastic models, bridging the gap between theoretical guarantees and computational efficiency.

Chapter 3

Mathematical Framework and Convergence Theory

This chapter establishes the convergence theory for empirical expected-signature estimators constructed from a single dependent trajectory. We prove that a block-based estimator achieves mean-squared error convergence at rate $O(N^{-2/p})$ where N is a granularity parameter and $p > 2$ characterises path regularity. The analysis combines rough-path continuity for bias control with covariance inequalities under exponential mixing for variance bounds.

3.1 Algebraic Framework

3.1.1 Truncated Tensor Algebra

Let V be a d-dimensional real vector space. We work with the truncated tensor algebra

$$T^{(M)}(V) = \bigoplus_{k=0}^{M} V^{\otimes k} = \mathbb{R} \oplus V \oplus V^{\otimes 2} \oplus \cdots \oplus V^{\otimes M},$$

where $M \in \mathbb{N}$ is the truncation depth. We equip this space with the Hilbert-Schmidt inner product, which endows $T^{(M)}(V)$ with a Hilbert space structure enabling the application of covariance inequalities for Hilbert-valued random variables.

For elements $u = (u_0, u_1, \ldots, u_M)$ and $v = (v_0, v_1, \ldots, v_M)$ in $T^{(M)}(V)$ where $u_k, v_k \in V^{\otimes k}$:

$$\langle u, v \rangle_{\text{HS}} = \sum_{k=0}^{M} \langle u_k, v_k \rangle_{V^{\otimes k}}.$$

This induces the Hilbert-Schmidt norm:

$$\|v\|_{\text{HS}} = \sqrt{\sum_{k=0}^{M} \|v_k\|_{V^{\otimes k}}^2}.$$

For random signatures $Y : \Omega \to T^{(M)}(V)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we work in $L^2(\Omega; T^{(M)}(V))$ with norm:

$$\|Y\|_{L^2} = \sqrt{\mathbb{E}[\|Y\|_{\text{HS}}^2]}.$$

3.2 Analytical Framework: Geometric Rough Paths

3.2.1 Rough Path Lift and Regularity

Let $X : [0, \infty) \to V$ be a continuous path. In the rough path framework, we lift X to a geometric p-rough path $\mathbf{X}$ for some regularity parameter $p > 2$.

Definition 3.1 (Geometric p-Rough Path) *A geometric p-rough path is a continuous map:*

$$\mathbf{X} : [0, \infty) \to G^{\lfloor p \rfloor}(V)$$

where $G^{\lfloor p \rfloor}(V)$ is the step-$\lfloor p \rfloor$ free nilpotent Lie group over V, satisfying:

1. *Chen's identity: $\mathbf{X}_{s,u} = \mathbf{X}_{s,t} \otimes \mathbf{X}_{t,u}$ for $s < t < u$*

2. *Finite p-variation: $\|\mathbf{X}\|_{p\text{-}var;[s,t]} < \infty$ for all compact intervals $[s, t]$*

3. *Geometric consistency with smooth approximations*

The signature map $S^{(M)} : \mathbf{X} \mapsto S^{(M)}(\mathbf{X}) \in T^{(M)}(V)$ is continuous in the p-variation topology [17, Thm. 7.16].

3.2.2 Path Regularity Parameter

Define the Hölder-type exponent $\beta := 1/p \in (0, 1/2)$ since $p > 2$. We choose the truncation level $M \geq \lceil p \rceil$ to ensure compatibility with the rough-path structure. This choice guarantees that the truncated signature captures the essential information from the $\lfloor p \rfloor$-level rough path.

3.3 Statistical Framework

Core Statistical Assumptions

We require three fundamental assumptions on the rough path process $\mathbf{X}$ to establish our convergence theorem. These control the analytical regularity, statistical dependence structure, and temporal homogeneity of the process.

Assumption (M) (Moment control) We assume that $\mathbf{X}$ is a random continuous geometric p-rough path, as defined in the Setting section. This implies certain regularity properties. Beyond this, we require control over the moments of its p-variation norm. For any $r \geq 1$, there exists a constant $C_r < \infty$ such that for any $s < t$:

$$\mathbb{E}\left[\|\mathbf{X}\|_{p\text{-var};[s,t]}^r\right] \leq C_r(t-s)^{r\beta}$$

where $\beta = 1/p$. This ensures the p-variation norm grows at most polynomially with interval length, a standard condition satisfied by fractional Brownian motion with Hurst parameter $H > \beta$ and solutions to SDEs with suitable coefficients. $\square$

Assumption (A) (α-mixing) To handle temporal dependence in our block-based estimator, we require mixing conditions that quantify how quickly dependence decays over time. The rough path process $\mathbf{X}$ is α-mixing with exponential decay: there exist constants $C_X > 0$ and $\lambda_X > 0$ such that:

$$\alpha_{\mathbf{X}}(t) \leq C_X e^{-\lambda_X t} \quad \forall t \geq 0$$

where the α-mixing coefficient is defined as:

$$\alpha_X(t) = \sup_{s\geq 0} \sup_{\substack{A\in\mathcal{F}_s \\ B\in\mathcal{F}_{s+t}^{\infty}}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

with $\mathcal{F}_s = \sigma(\{X_u : u \leq s\})$ and $\mathcal{F}_{s+t}^{\infty} = \sigma(\{X_u : u \geq s+t\})$. $\square$

This exponential mixing condition ensures that blocks separated by sufficient time gaps are nearly independent, enabling our convergence analysis. Ornstein-Uhlenbeck processes and many ergodic diffusions satisfy this condition.

For time-homogeneous analysis, we introduce the increment shift operator θ_τ for $\tau \geq 0$. For a rough path $\mathbf{X}$, the shifted process $(\theta_\tau \mathbf{X})$ has:

$$(\theta_\tau \mathbf{X})(t) := \mathbf{X}_{\tau,\tau+t}$$

with increments $(\theta_\tau \mathbf{X})_{s,t} = \mathbf{X}_{\tau+s,\tau+t}$. This operator re-indexes the process to start at time τ.

Assumption (S) (Segment stationarity) We assume that the process $(X_t)_{t\geq 0}$ is strictly stationary, and in particular that for any block length Δ and shift τ, the increment segment

$$(X_{\tau+s} - X_\tau)_{0\leq s\leq\Delta}$$

has the same law as

$$(X_s - X_0)_{0\leq s\leq\Delta}.$$

Equivalently, using the shift operator θ_τ defined above, we require $\theta_\tau \mathbf{X} \stackrel{d}{=} \mathbf{X}$ for all $\tau \geq 0$. This "segment-stationarity" ensures that the sequence of block signatures $(S^{(M)}(\theta_{k\Delta}\mathbf{X}|_{[0,\Delta]}))_{k\geq 0}$ is strictly stationary. $\square$

Remark 3.2 (Distinction from Lévy Increments)
Note that Assumption (S) does not require stationary increments in the Lévy sense, where increment laws $X_{s+h} - X_s$ depend only on the lag h. For example, Ornstein-Uhlenbeck processes are not Lévy-increment stationary, but they satisfy Assumption (S) when initialised from their invariant distribution, since their segment laws are shift-invariant under equilibrium initialisation.

These three assumptions (moment control, mixing, and segment stationarity) are satisfied by a wide class of processes. This includes Ornstein-Uhlenbeck processes (our primary example), fractional Brownian motion with $H > \beta$ where $\beta = 1/p$, solutions to ergodic SDEs with suitable coefficients, and stationary Gaussian processes with appropriate covariance decay. The interplay between these assumptions enables our main convergence result: segment stationarity ensures consistent block structure, mixing allows application of limit theorems for dependent data, and moment control provides the necessary integrability.

3.4 Supporting Lemmas

Our proof relies on covariance inequalities for stationary mixing sequences. The sequence of random variables we wish to analyze is the sequence of block signatures, $(Z_k^{(N)})_{k \in \mathbb{Z}}$. The following lemma provides the crucial justification that this sequence is indeed stationary, inheriting this property from the underlying rough path.

Lemma 3.3 (Stationarity of Block Signatures) *Fix $N \in \mathbb{N}$ and set $\Delta t_N := \delta/N$ and define the* block signatures

$$Z_k^{(N)} := S_{0,\Delta t_N}^{(M)}\big(\theta_{(k-1)\Delta t_N}\mathbf{X}\big), \qquad k \in \mathbb{Z}.$$

Then under Assumption S, the sequence $\big(Z_k^{(N)}\big)_{k \in \mathbb{Z}}$ is strictly stationary.

Proof Consider the path space of geometric p–rough paths on $[0, \Delta t_N]$, endowed with the canonical Borel σ–field generated by the p–variation (or rough path) topology. The restriction map

$$\mathcal{R}: \ \mathbf{Y} \mapsto \mathbf{Y}! \upharpoonright_{[0,\Delta t_N]}$$

from rough paths on $[0, \infty)$ to rough paths on $[0, \Delta t_N]$ is Borel measurable, since it is continuous for the topology induced by p–variation on compact intervals.

The truncated signature map

$$\mathcal{S}: \ \mathbf{Y} \mapsto S_{0,\Delta t_N}^{(M)}(\mathbf{Y})$$

is continuous (hence Borel) on the rough–path space (see, e.g., [17, Thm. 7.16]). Therefore the composition

$$F_N : \ \mathbf{Y} \mapsto \big(\mathcal{S} \circ \mathcal{R}\big)(\mathbf{Y})$$

is a deterministic Borel map from rough paths on $[0, \infty)$ to $T^{(M)}(V)$.

By Assumption (S), the process $\{\theta_{(k-1)\Delta t_N}\mathbf{X}\}_{k\in\mathbb{Z}}$ is strictly stationary as a random element with values in the space of rough paths on $[0, \infty)$. Applying the measurable map F_N coordinatewise preserves strict stationarity of the array; but $Z_k^{(N)} = F_N(\theta_{(k-1)\Delta t_N}\mathbf{X})$ by definition. Hence $\{Z_k^{(N)}\}_{k\in\mathbb{Z}}$ is strictly stationary. $\qquad\square$

Proposition 3.4 (Deterministic polynomial growth of signatures)
There exists a constant $C_{\mathrm{sig}} = C_{\mathrm{sig}}(p, M)$ such that, for every geometric p–rough path $\mathbf{U}$ and every interval $[s, t]$,

$$\|S_{s,t}^{(M)}(\mathbf{U})\|_{\mathrm{HS}} \le C_{\mathrm{sig}}\Big(1 + \|\mathbf{U}\|_{p\text{-}var;[s,t]}^M\Big).$$

This is a standard result from rough path theory, see, e.g., [18, Prop. 9.3, Rem. 9.6].

Proposition 3.5 (Lipschitz continuity of signatures) *For geometric p–rough paths $\mathbf{U}, \mathbf{V}$ and any interval $[s, t]$, there exists a constant $C_{p,M}$ such that*

$$\|S_{s,t}^{(M)}(\mathbf{U}) - S_{s,t}^{(M)}(\mathbf{V})\|_{\mathrm{HS}} \le C_{p,M}\, d_{p\text{-}var;[s,t]}(\mathbf{U}, \mathbf{V})\big(1 + \|\mathbf{U}\|_{p\text{-}var;[s,t]}^M + \|\mathbf{V}\|_{p\text{-}var;[s,t]}^M\big).$$

See [17, Thm. 7.16] for the continuity of the signature map.

Proposition 3.6 (Wong–Zakai local approximation) *Let $\mathbf{X}$ be a geometric p–rough path and $\mathbf{X}_k^{\pi,N}$ its piecewise linear approximation on the interval $I_k^{(N)}$. Then*

$$d_{p\text{-}var;I_k^{(N)}}(\mathbf{X}, \mathbf{X}_k^{\pi,N}) \le C\|\mathbf{X}\|_{p\text{-}var;I_k^{(N)}}^2$$

for a universal constant C depending only on p. This follows from the local Wong–Zakai estimate, see [18, Prop. 14.5].

Lemma 3.7 (Uniform Moment Bounds) *Under Assumption (M), for any $r \ge 1$, there exists a constant $C^{(r)}$ depending only on p, M, r, δ such that for all $N \in \mathbb{N}$ and $1 \le k \le K_N$:*

$$\mathbb{E}[\|Z_k^{(N)}\|_{\mathrm{HS}}^r] \le C^{(r)}.$$

Consequently, the centred variables $W_k^{(N)} := Z_k^{(N)} - \mathbb{E}[Z_k^{(N)}]$ satisfy $\sup_{N,k} \mathbb{E}\|W_k^{(N)}\|_{\mathrm{HS}}^r < \infty$ for all $r \ge 1$.

Proof By the polynomial growth bound for signatures [18, Prop. 9.3], $\|S^{(M)}(\mathbf{U})\|_{\mathrm{HS}} \leq C(1 + \|\mathbf{U}\|_{p\text{-var}}^{M})$ for any geometric p-rough path $\mathbf{U}$. Applying this with $\mathbf{U} = \theta_{(k-1)\Delta t_N}\mathbf{X}|_{[0,\Delta t_N]}$ and using Assumption (M) yields the result. $\square$

3.5 Block-Sampling Framework

Before stating our main convergence theorem, we describe the block-sampling scheme that underlies our estimator.

3.5.1 Block-Sampling Scheme

For $N \in \mathbb{N}$, define $\Delta t_N := \delta/N$ where δ is a fixed time horizon. To balance bias and variance, we choose the number of blocks K_N to scale as $N^{1+2\beta}$ where $\beta = 1/p$. Specifically, let $K_N = \lceil A_K N^{1+2\beta} \rceil$ for some constant $A_K > 0$. Since $\beta \in (0, 1/2)$, we have $1 < 1 + 2\beta < 2$.

The blocks are $I_k^{(N)} := [(k-1)\Delta t_N, k\Delta t_N]$ for $1 \leq k \leq K_N$.

Remark (Retrospective estimation only): Throughout this work, we assume that the full trajectory over $[0, K_N\Delta t_N]$ has been completely observed before any estimation begins. The piecewise-linear interpolation $X_k^{\pi,N}$ uses only the realized discrete observations within each block $I_k^{(N)}$ to construct a continuous approximation. This is purely a *backward-looking* operation for parameter estimation from historical data; no forecasting or prediction is involved, and hence no look-ahead bias arises. Applications requiring forward prediction would need to ensure that interpolation at time t uses only observations available up to time t.

Let X be the continuous path underlying $\mathbf{X}$. Let $X_k^{\pi,N} : I_k^{(N)} \to V$ be the piecewise-linear interpolation of X on $I_k^{(N)}$, and $\mathbf{X}_k^{\pi,N}$ its canonical rough path lift over $I_k^{(N)}$.

We define:

42 *• **True signature on block** k: $Z_k^{(N)} := S_{0,\Delta t_N}^{(M)}\big(\theta_{(k-1)\Delta t_N}\mathbf{X}\big)$ for $k \in \mathbb{Z}$.

- **Approximated signature on** $I_k^{(N)}$: $Y_k^{(N)} := S_{I_k^{(N)}}^{(M)}(\mathbf{X}_k^{\pi,N})$.

- **Estimator**: $\widehat{\mathbb{E}}[S^{(M)}]_N := \frac{1}{K_N} \sum_{k=1}^{K_N} Y_k^{(N)}$.

- **Target expectation**: $\mathbb{E}[S^{(M)}] := \mathbb{E}[Z_0^{(N)}]$ (which equals $\mathbb{E}[Z_k^{(N)}]$ by stationarity).

3.6 Convergence Theorem

Theorem 3.8 (Convergence of Empirical Expected Signature) *Let* $\mathbf{X}$ *be a random continuous geometric p-rough path satisfying Assumptions (M), (A), and (S). Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\|\widehat{\mathbb{E}}[S^{(M)}]_N - \mathbb{E}[S^{(M)}]\|_{\mathrm{HS}}^2 \leq C N^{-2/p}.$$

Consequently, $\widehat{\mathbb{E}}[S^{(M)}]_N \to \mathbb{E}[S^{(M)}]$ in L^2 and in probability as $N \to \infty$.

Proof We argue in the Hilbert space $T^{(M)}(V)$. Set

$$B_N := \frac{1}{K_N} \sum_{k=1}^{K_N} \left(Y_k^{(N)} - Z_k^{(N)} \right),$$

$$S_N := \frac{1}{K_N} \sum_{k=1}^{K_N} W_k^{(N)}, \quad W_k^{(N)} := Z_k^{(N)} - \mathbb{E}[S^{(M)}].$$

Then

$$\mathbb{E}\|\widehat{\mathbb{E}}[S^{(M)}]_N - \mathbb{E}[S^{(M)}]\|_{\mathrm{HS}}^2 \leq 2\mathbb{E}\|B_N\|_{\mathrm{HS}}^2 + 2\mathbb{E}\|S_N\|_{\mathrm{HS}}^2.$$

We bound the bias and fluctuation terms separately.

Bias term. Fix k and abbreviate $I = I_k^{(N)}$. By Proposition 3.5 with $\mathbf{U} = \mathbf{X}\!\restriction_I$ and $\mathbf{V} = \mathbf{X}_k^{\pi,N}$, and then Proposition 3.6,

$$\|Y_k^{(N)} - Z_k^{(N)}\|_{\mathrm{HS}} \leq C_{p,M}\, d_{p\text{-var};I}\big(\mathbf{X}, \mathbf{X}_k^{\pi,N}\big) \left(1 + \|\mathbf{X}\|_{p\text{-var};I}^M + \|\mathbf{X}_k^{\pi,N}\|_{p\text{-var};I}^M \right)$$

$$\leq C\, \|\mathbf{X}\|_{p\text{-var};I}^2 \left(1 + \|\mathbf{X}\|_{p\text{-var};I}^M \right),$$

for a constant C independent of N, k. Taking L^2 norms and using Assumption (M) with moments of order $4 + 2M$,

$$\|Y_k^{(N)} - Z_k^{(N)}\|_{L^2} \leq C(\Delta t_N)^{2\beta} = C N^{-2\beta}.$$

Hence,

$$\mathbb{E}\|B_N\|_{\mathrm{HS}}^2 \leq C N^{-4\beta}.$$

Fluctuation term. By Lemma 3.3, $\{W_k^{(N)}\}$ is strictly stationary. Let $c_h^{(N)} := \mathbb{E}\langle W_0^{(N)}, W_h^{(N)}\rangle_{\mathrm{HS}}$. A standard reindexing for stationary sequences (e.g., [7, §2.4]) yields

$$\mathbb{E}\|S_N\|_{\mathrm{HS}}^2 = \frac{1}{K_N} \left(c_0^{(N)} + 2 \sum_{h=1}^{K_N - 1} \left(1 - \frac{h}{K_N} \right) c_h^{(N)} \right).$$

15

To bound the covariance tail, set $\mathcal{A}_k := \sigma(\mathbf{X}\restriction_{I_k^{(N)}})$ and $\mathcal{B}_k := \sigma(Z_k^{(N)})$. Since $Z_k^{(N)}$ is a measurable functional of the segment $\mathbf{X}\restriction_{I_k^{(N)}}$, we have $\mathcal{B}_k \subseteq \mathcal{A}_k$; hence by monotonicity of α–mixing and Assumption (A),

$$\alpha_{Z^{(N)}}(h) := \alpha(\mathcal{B}_0, \mathcal{B}_h) \leq \alpha(\mathcal{A}_0, \mathcal{A}_h) \leq C_X e^{-\lambda_X (h-1)\Delta t_N}, \qquad h \geq 1.$$

Let $r > 2$. Applying a Davydov-type covariance inequality for Hilbert-valued random variables (see, e.g., [5, Prop. 1.5]),

$$\left| c_h^{(N)} \right| \leq C_r \left(\mathbb{E}\|W_0^{(N)}\|^r \right)^{2/r} \alpha_{Z^{(N)}}(h)^{1-2/r}, \qquad h \geq 1.$$

By Lemma 3.7, $\sup_N \mathbb{E}\|W_0^{(N)}\|_{\mathrm{HS}}^r < \infty$ for every $r \geq 1$. Combining (3)–(3) gives constants $C, \lambda' > 0$ (independent of N) such that

$$\left| c_h^{(N)} \right| \leq C\, e^{-\lambda'(h-1)\Delta t_N}, \qquad h \geq 1.$$

Therefore

$$\sum_{h=1}^{\infty} \left| c_h^{(N)} \right| \leq C \sum_{h=1}^{\infty} e^{-\lambda'(h-1)\Delta t_N} = \frac{C}{1 - e^{-\lambda'\Delta t_N}} \leq \frac{C'}{\Delta t_N} = C'\frac{N}{\delta},$$

using $1 - e^{-x} \geq (1 - e^{-1})x$ for $x \in (0, 1]$ (valid for all sufficiently large N; small N only alters constants). Plugging this into (3) and recalling $K_N \asymp N^{1+2\beta}$ where $\beta = 1/p$,

$$\mathbb{E}\|S_N\|_{\mathrm{HS}}^2 \lesssim \frac{N}{N^{1+2\beta}} = N^{-2\beta} = N^{-2/p}.$$

Combining both terms,

$$\mathbb{E}\|\widehat{\mathbb{E}}[S^{(M)}]_N - \mathbb{E}[S^{(M)}]\|_{\mathrm{HS}}^2 \leq C N^{-4\beta} + C' N^{-2\beta} \leq C N^{-2\beta} = C N^{-2/p}.$$

In particular $\widehat{\mathbb{E}}[S^{(M)}]_N \to \mathbb{E}[S^{(M)}]$ in L^2; convergence in probability follows by Chebyshev. $\qquad\square$

3.6.1 Limiting Case Analysis

The convergence rate $O(N^{-2/p})$ depends on the path regularity parameter $p > 2$. As $p \to 2^+$, the rate approaches $O(N^{-1})$, which represents the supremum of convergence rates under our framework:

$$\sup_{p>2} \left\{ -\frac{2}{p} \right\} = -1.$$

In applications one may treat $p = 2 + \varepsilon$ with $\varepsilon > 0$ as arbitrarily small; the theory requires $p > 2$ strictly.

3.7 Application to Time-Augmented Processes

Lemma 3.9 (Time Augmentation Preserves Assumptions)
Let $(X_t)_{t \geq 0}$ be a d-dimensional process satisfying Assumptions (M), (A), and (S). Then the time-augmented process $Y_t = (t, X_t) \in \mathbb{R}^{1+d}$ also satisfies these assumptions.

Proof The deterministic time component has finite p-variation for $p > 1$. The σ-algebra structure is unchanged by deterministic augmentation, preserving the mixing property; this follows from the deterministic nature of the time coordinate and standard properties of α-mixing under measurable transformations [6, Thm. 3.5]. For segment stationarity, note that $(\theta_\tau Y)_t = (t, X_{\tau+t} - X_\tau)$ has law independent of τ. $\qquad\square$

Remark 3.10 *Time augmentation preserves Assumptions (M), (A), and (S) so the convergence theorem applies; identifiability of model parameters depends on the specific model class, as explored in Chapter 4.*

3.8 Chapter Summary

This chapter has established the mathematical framework for empirical expected signature estimation from dependent single-path data. The central result, Theorem 3.8, demonstrates that our block-based estimator achieves mean-squared error convergence at rate $O(N^{-2/p})$ where N represents the granularity parameter and $p > 2$ characterises the path regularity.

The theoretical development rests on three fundamental assumptions that capture the essential properties required for convergence. Assumption (M) provides moment control through polynomial bounds on the p-variation norm. Assumption (A) ensures exponential decay of temporal dependence via α-mixing conditions. Assumption (S) establishes segment stationarity, enabling consistent estimation from time-homogeneous increments. These assumptions are satisfied by a broad class of processes relevant to financial applications, including Ornstein-Uhlenbeck processes and fractional Brownian motion with appropriate Hurst parameters.

The proof architecture employs a bias-variance decomposition within the Hilbert space structure of the truncated tensor algebra $T^{(M)}(V)$. The bias term, arising from piecewise linear approximation, decays at rate $O(N^{-4\beta})$ where $\beta = 1/p$. The fluctuation term, controlled through covariance inequalities for mixing sequences, achieves the slower rate $O(N^{-2\beta})$ and therefore determines the overall convergence rate. The block-sampling scheme with $K_N \asymp N^{1+2\beta}$ blocks represents an optimal balance between these competing sources of error.

A notable feature of our framework is the preservation of stationarity through the block structure, as established in Lemma 3.3. This enables the appli-

cation of ergodic theory without requiring independence assumptions that would be unrealistic for financial time series. The time augmentation result (Lemma 3.9) further extends the framework's applicability by showing that our assumptions are preserved under the time-augmentation necessary for parameter identifiability in certain models.

The theoretical foundation developed here provides the mathematical underpinning for the numerical verification presented in Chapter 5 and the calibration methodology explored in Chapter 6. The explicit convergence rate serves as a benchmark against which empirical performance can be evaluated, while the constructive nature of the proof yields practical guidance for implementation.

Chapter 4

Analytical Framework for Ornstein-Uhlenbeck Processes

Having established the general convergence theory for empirical expected signatures in Chapter 3, we now specialise to the Ornstein-Uhlenbeck (OU) family, where the linear structure enables exact computation of expected signatures through matrix exponentials. This analytical tractability serves two purposes: it provides machine-precision ground truth for validating the convergence rates derived theoretically, and it supplies efficient gradients for the calibration methodology developed in Chapter 6.

4.1 The Ornstein–Uhlenbeck model

We fix the Stratonovich convention and consider the d-dimensional SDE

$$dX_t \; = \; \theta\big(\mu - X_t\big)\, dt \; + \; \sigma \circ dW_t, \qquad X_0 \in \mathbb{R}^d.$$

When block segment-stationarity is required (e.g., Proposition 4.8), we assume equilibrium initialisation with $X_0 \sim N(\mu, \Sigma_\infty)$ where Σ_∞ solves the Lyapunov equation $\theta \Sigma_\infty + \Sigma_\infty \theta^\top = Q$. Here $\theta \in \mathbb{R}^{d \times d}$ (mean-reversion), $\mu \in \mathbb{R}^d$ (long-run mean), and $\sigma \in \mathbb{R}^{d \times d}$ (diffusion); we write $Q := \sigma \sigma^\top$. Assume $\Re \operatorname{spec}(\theta) > 0$ and, unless otherwise stated, that X_0 is independent of W. The Stratonovich convention ensures Chen's identity and compatibility with the algebraic structure of signatures.

4.2 Lifted generator on $T^{(M)}(\mathbb{R}^d)$

Let $T^{(M)}(\mathbb{R}^d) = \bigoplus_{k=0}^M (\mathbb{R}^d)^{\otimes k}$ with basis words $w = e_{i_1} \otimes \cdots \otimes e_{i_k}$ and empty word $\mathbf{1}$ at degree 0. For the derivation ∂_{e_j} on words (left-derivation: removes a leading e_j; zero otherwise), the lifted generator has the following form.

Proposition 4.1 (Lifted generator; creation/derivation blocks) *For $\psi = (\theta, \mu, \sigma)$ with $Q = \sigma\sigma^\top$, the operator $\mathcal{G}^{(M)}(\psi) : T^{(M)}(\mathbb{R}^d) \to T^{(M)}(\mathbb{R}^d)$ acts on basis words w as*

$$\mathcal{G}^{(M)}(\psi)\, w \;=\; \sum_{j=1}^{d} (\theta\mu)_j\, w \otimes e_j \;+\; \sum_{i,j=1}^{d} (-\theta_{ij})\, e_i\, \partial_{e_j} w \;+\; \frac{1}{2} \sum_{i,j=1}^{d} Q_{ij}\, w \otimes e_i \otimes e_j.$$

Proof (Sketch) The Stratonovich generator on smooth f is $\mathscr{L}f(x) = \langle \theta(\mu - x), \nabla f(x)\rangle + \frac{1}{2}\mathrm{tr}(Q\nabla^2 f(x))$. Lifting $\mathscr{L}$ to signature coordinates yields: (i) affine drift $\theta\mu$ creates one letter; (ii) linear term $-\theta x$ acts as a derivation on the first letter; (iii) diffusion Q creates two letters with the Stratonovich symmetrisation factor $1/2$. This closes on $T^{(M)}(\mathbb{R}^d)$. For a complete proof, see Appendix 7. $\qquad\square$

Remark 4.2
$\mathcal{G}^{(M)}$ *is block lower-triangular by tensor degree: the derivation block preserves degree, while drift and diffusion are degree $+1$ and $+2$ creation blocks. This finite-dimensional closure drives all subsequent calculations.*

4.3   Semigroup representation and exact formulas

Let $E^{(M)}(t) := \mathbb{E}[S^{(M)}(X)_{0,t}] \in T^{(M)}(\mathbb{R}^d)$. Then

$$\frac{d}{dt} E^{(M)}(t) = \mathcal{G}^{(M)}(\psi)\, E^{(M)}(t), \qquad E^{(M)}(0) = \mathbf{1}.$$

The explicit solution (central for what follows) is

$$E^{(M)}(t) = \exp\big(t\, \mathcal{G}^{(M)}(\psi)\big)\, \mathbf{1}.$$

4.3.1   Degree-1 and degree-2 components

Write $m(t) := \mathbb{E}[X_t]$ and $\Sigma(t) := \mathrm{Cov}(X_t)$. Solving $m'(t) = \theta(\mu - m(t))$ gives

$$m(t) = \mu + e^{-\theta t}\big(m(0) - \mu\big).$$

The first-level expected signature equals the mean increment:

$$\mathbb{E}\big[S^{(1)}(X)_{0,T}\big] \;=\; m(T) - m(0) \;=\; \big(I - e^{-\theta T}\big)(\mu - m(0)).$$

For the second level (Stratonovich), symmetry yields

$$\mathrm{sym}\, \mathbb{E}\big[S^{(2)}(X)_{0,T}\big] \;=\; \tfrac{1}{2}\, \mathrm{sym}\, \mathbb{E}\big[(X_T - X_0)^{\otimes 2}\big].$$

Under our standing independence of X_0 and W, the increment covariance decomposes as

$$\mathrm{Cov}(X_T - X_0) \;=\; \big(I - e^{-\theta T}\big) \Sigma(0) \big(I - e^{-\theta T}\big)^\top \;+\; \int_0^T e^{-(T-s)\theta}\, Q\, e^{-(T-s)\theta^\top}\, ds,$$

and the Lévy area has zero mean in this additive-noise setting. Hence

$$\mathbb{E}\big[S^{(2)}(X)_{0,T}\big] \;=\; \frac{1}{2}\Big(\big(m(T) - m(0)\big)^{\otimes 2} \;+\; \mathrm{Cov}(X_T - X_0)\Big).$$

Kronecker form (vectorised Lyapunov). Writing the Kronecker sum $\theta \oplus \theta := \theta \otimes I + I \otimes \theta$, one has the starred identity

$$\mathrm{vec}\left(\int_0^T e^{-(T-s)\theta} Q\, e^{-(T-s)\theta^\top}\, ds\right) \;=\; (\theta \oplus \theta)^{-1}\Big(I - e^{-T(\theta \oplus \theta)}\Big)\, \mathrm{vec}(Q),$$

with the Sylvester interpretation if $\theta \oplus \theta$ is singular.

4.4 Sensitivity analysis and gradients

Let $\Phi_{T,M}(\psi) := \exp\big(T\, \mathcal{G}^{(M)}(\psi)\big)\mathbf{1}$ with state dimension $D = \sum_{k=0}^{M} d^k$. The Fréchet derivative of the matrix exponential is

$$D\exp(A)[H] = \int_0^1 e^{(1-s)A}\, H\, e^{sA}\, ds.$$

By the chain rule, with $A(\psi) := T\, \mathcal{G}^{(M)}(\psi)$,

$$D\Phi_{T,M}(\psi)[\delta\psi] = \left(\int_0^1 e^{(1-s)A(\psi)}\, \big(T\, D\mathcal{G}^{(M)}(\psi)[\delta\psi]\big)\, e^{sA(\psi)}\, ds\right)\mathbf{1}.$$

For $\delta\psi = (\delta\theta, \delta\mu, \delta Q)$, the derivative of the lifted generator inherits the block structure:

$$D\mathcal{G}^{(M)}(\psi)[\delta\psi]\, w = \sum_{j=1}^{d} \big((\delta\theta)\mu + \theta\, \delta\mu\big)_j\, w \otimes e_j \;+\; \sum_{i,j=1}^{d} (-\delta\theta_{ij})\, e_i\, \partial_{e_j} w$$

$$+\; \frac{1}{2}\sum_{i,j=1}^{d} \delta Q_{ij}\, w \otimes e_i \otimes e_j.$$

Equations (4.4)–(4.4) provide Jacobian–vector products needed for gradient-based calibration and clarify which tensor degrees carry information about θ, μ, Q.

Remark 4.3 (Generic non-resonance)
The Fréchet derivative of the matrix exponential, $L_A(H) = \int_0^1 e^{(1-s)A} H e^{sA} ds$, is injective except when eigenvalue differences of A lie in $2\pi i\, \mathbb{Z}$. This non-resonance condition is crucial for parameter identifiability, as established rigorously in Theorem 4.9 and explained geometrically through the adjoint action in Lemma 4.10 and Lemma 4.11 below.

4.5 Parameter Identifiability

The calibration framework developed in Chapter 6 relies fundamentally on the injectivity of the parameter-to-signature map. This section establishes the necessary identifiability results with complete mathematical rigor, addressing the subtle interplay between time horizons and parameter scaling.

4.5.1 The Parameter-to-Signature Map

Definition 4.4 (Parameter-to-Expected-Signature Map) *For a fixed observation interval $[0, T]$ and truncation level M, define the map $\Phi_{T,M} : \Psi_{stable} \to T^{(M)}(\mathbb{R}^d)$ by*

$$\Phi_{T,M}(\theta, \mu, \sigma) = \mathbb{E}[S^{(M)}(X)_{0,T}],$$

where X solves the Ornstein-Uhlenbeck SDE (4.1) with parameters (θ, μ, σ).

By (4.3), this map admits the explicit representation

$$\Phi_{T,M}(\theta, \mu, \sigma) = \exp(T\mathcal{G}^{(M)}(\theta, \mu, \sigma))\mathbf{1},$$

where $\mathcal{G}^{(M)}$ is the lifted generator constructed in Proposition 4.1.

4.5.2 Time-Scaling Symmetry

Before establishing identifiability, we must first characterise the fundamental symmetry of the OU expected signature.

Lemma 4.5 (Time-Scaling Symmetry) *Let $(\theta, \mu, \sigma) \in \Psi_{stable}$ and $\alpha > 0$. Then*

$$\Phi_{T,M}(\theta, \mu, \sigma) = \Phi_{\alpha T, M}(\theta/\alpha, \mu, \sigma/\sqrt{\alpha}).$$

That is, the parameters (θ, μ, σ) observed over time T produce the same expected signature as $(\theta/\alpha, \mu, \sigma/\sqrt{\alpha})$ observed over time αT.

Proof By Proposition 4.1, the lifted generator decomposes as

$$\mathcal{G}^{(M)}(\psi) = \sum_{j=1}^{d} (\theta\mu)_j \, (\cdot \otimes e_j) - \sum_{i,j=1}^{d} \theta_{ij} \, (e_i \otimes \partial_{e_j}) + \frac{1}{2} \sum_{i,j=1}^{d} Q_{ij} \, (\cdot \otimes e_i \otimes e_j),$$

where the three terms represent affine drift, mean-reversion, and diffusion respectively (see Appendix 7 for the complete derivation). We verify that the parameter transformation $\psi = (\theta, \mu, \sigma) \mapsto \tilde{\psi} = (\theta/\alpha, \mu, \sigma/\sqrt{\alpha})$ induces uniform scaling by $1/\alpha$ across all three components.

Step 1: Component-wise scaling.

(i) Affine drift. Since μ is invariant and the drift is linear in $\theta\mu$,

$$\sum_{j=1}^{d} \frac{(\theta\mu)_j}{\alpha} \left(\cdot \otimes e_j \right) = \frac{1}{\alpha} \sum_{j=1}^{d} (\theta\mu)_j \left(\cdot \otimes e_j \right).$$

(ii) Mean-reversion derivation. The derivation operator is linear in θ,

$$-\sum_{i,j=1}^{d} \frac{\theta_{ij}}{\alpha} \left(e_i \otimes \partial_{e_j} \right) = \frac{1}{\alpha} \left(-\sum_{i,j=1}^{d} \theta_{ij} \left(e_i \otimes \partial_{e_j} \right) \right).$$

(iii) Diffusion. Under $\sigma \mapsto \sigma/\sqrt{\alpha}$, the covariance transforms as $Q = \sigma\sigma^{\top} \mapsto Q/\alpha$ (the $\sqrt{}$time scaling of Brownian volatility), hence

$$\frac{1}{2} \sum_{i,j=1}^{d} \frac{Q_{ij}}{\alpha} \left(\cdot \otimes e_i \otimes e_j \right) = \frac{1}{\alpha} \left(\frac{1}{2} \sum_{i,j=1}^{d} Q_{ij} \left(\cdot \otimes e_i \otimes e_j \right) \right).$$

Step 2: Fundamental scaling relation. Summing the three components from Step 1 yields the operator identity

$$\mathcal{G}^{(M)}(\theta/\alpha, \mu, \sigma/\sqrt{\alpha}) = \frac{1}{\alpha} \mathcal{G}^{(M)}(\theta, \mu, \sigma),$$

holding pointwise on $T^{(M)}(\mathbb{R}^d)$.

Step 3: Exponentiation to global symmetry. By Definition 4.4 and the scaling relation (6),

$$\begin{aligned}
\Phi_{\alpha T, M}(\theta/\alpha, \mu, \sigma/\sqrt{\alpha}) &= \exp\left(\alpha T \cdot \mathcal{G}^{(M)}(\theta/\alpha, \mu, \sigma/\sqrt{\alpha}) \right) \mathbf{1} \\
&= \exp\left(\alpha T \cdot \frac{1}{\alpha} \mathcal{G}^{(M)}(\theta, \mu, \sigma) \right) \mathbf{1} \\
&= \exp\left(T \cdot \mathcal{G}^{(M)}(\theta, \mu, \sigma) \right) \mathbf{1} = \Phi_{T, M}(\theta, \mu, \sigma). \qquad \square
\end{aligned}$$

Remark 4.6 (Physical Interpretation)
This symmetry reflects a fundamental property of mean-reverting processes: a fast-reverting process observed briefly is statistically indistinguishable from a slow-reverting process observed longer. The signature captures only the relative dynamics scaled by the observation window. This has implications for calibration:

42 *• Without fixing T, the calibration problem is ill-posed

- *Block-based estimation with duration $\Delta t = T/K$ implicitly works at scale $\alpha = 1/K$*

4.5.3 Block-Based Identifiability and Rescaling

Definition 4.7 (Block Signature Estimator) *For a partition of $[0, T]$ into K blocks of duration $\Delta t = T/K$, the block-based expected signature is:*

$$\Phi_{\Delta t, M}^{(K)}(\psi) = \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}[S^{(M)}(X)_{(k-1)\Delta t, k\Delta t}].$$

By segment-stationarity of the OU process under equilibrium initialisation, this simplifies to $\Phi_{\Delta t, M}^{(K)}(\psi) = \Phi_{\Delta t, M}(\psi)$.

Proposition 4.8 (Rescaled Signature Recovery)
The full-horizon expected signature can be recovered from block signatures via:

$$\Phi_{T, M}(\psi) = \exp((T - \Delta t)\mathcal{G}^{(M)}(\psi))\Phi_{\Delta t, M}(\psi).$$

This rescaling operator is exact for OU processes.

Proof Direct application of the semigroup property of $\exp(t\mathcal{G}^{(M)})$:

$$\exp((T - \Delta t)\mathcal{G}^{(M)})\Phi_{\Delta t, M}(\psi) = \exp((T - \Delta t)\mathcal{G}^{(M)}) \exp(\Delta t \mathcal{G}^{(M)})\mathbf{1}$$
$$= \exp(T\mathcal{G}^{(M)})\mathbf{1} = \Phi_{T, M}(\psi).$$

4.5.4 Identifiability for Fixed Time Horizon

Theorem 4.9 (Local identifiability from truncated expected signatures)
Fix $d \geq 1$, $M \geq 2$, $T > 0$, and let $\Psi_{\text{stable}} = \{\psi = (\theta, \mu, Q) : \Re(\theta) > 0, \ Q \in S_{++}^d\}$. For $\psi \in \Psi_{\text{stable}}$ let $\mathcal{G}^{(M)}(\psi)$ denote the degree-graded generator on $T^{(M)}(\mathbb{R}^d)$ so that the truncated expected signature satisfies

$$\Phi_{T, M}(\psi) := \exp(T\,\mathcal{G}^{(M)}(\psi))\,\mathbf{1} \in T^{(M)}(\mathbb{R}^d).$$

Assume the non-resonance condition

$$\lambda_i - \lambda_j \notin 2\pi i\, \mathbb{Z} \qquad \text{for all } \lambda_i, \lambda_j \in (T\,\mathcal{G}^{(M)}(\psi)).$$

Then the Fréchet derivative $D\Phi_{T, M}(\psi)$ is injective on the parameter tangent space at ψ. In particular, $\Phi_{T, M}$ is locally one-to-one on Ψ_{stable} at ψ.

Proof Write $A := T\,\mathcal{G}^{(M)}(\psi)$. The chain rule gives

$$D\Phi_{T, M}(\psi)[\dot{\psi}] = L_A[T\, D\mathcal{G}^{(M)}(\psi)[\dot{\psi}]],$$

where $L_A : \mathcal{L}(T^{(M)}(\mathbb{R}^d)) \to \mathcal{L}(T^{(M)}(\mathbb{R}^d))$ is the Fréchet derivative of the matrix exponential at A, explicitly

$$L_A(H) = \int_0^1 e^{(1-s)A} H\, e^{sA}\, ds,$$

24

cf. [21, Eq. (10.15)]. Hence $D\Phi_{T,M}(\psi)$ is injective iff both L_A is invertible and $D\mathcal{G}^{(M)}(\psi)$ has full column rank. We prove these two facts separately.

Step 1: full rank of $D\mathcal{G}^{(M)}(\psi)$. The space $T^{(M)}(\mathbb{R}^d) = \bigoplus_{k=0}^{M}(\mathbb{R}^d)^{\otimes k}$ is degree-graded. For the Ornstein–Uhlenbeck SDE (4.1) with $Q = \sigma\sigma^\top$, the lifted generator has the standard block lower-triangular form

$$\mathcal{G}^{(M)}(\psi) = \begin{bmatrix} 0 & 0 & 0 & \cdots \\ * & -\theta^\top & 0 & \cdots \\ * & \mathcal{I}_2(Q) & \mathcal{L}_2(\theta) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where: (i) $\mathcal{G}^{(M)}(\psi)$ maps level k to levels $\leq k$; (ii) the $(1,0)$ block equals insertion of $\theta\mu$ (a linear map $\mathbb{R} \to \mathbb{R}^d$), so $D_\mu\mathcal{G}^{(M)}(\psi)$ is supported only in the $(1,0)$ block and acts as θ times the identity on μ; (iii) the diagonal block on level 1 equals $-\theta^\top$, hence $D_\theta\mathcal{G}^{(M)}(\psi)$ restricted to the $(1,1)$ block recovers θ directly; (iv) the $(2,0)$ block equals $\frac{1}{2}\mathrm{Sym} \circ \iota(Q)$, i.e. linear in Q and supported only on level 2, independent of μ and with no overlap with the $(1,1)$ block; and (v) for $k \geq 2$ the diagonal block $\mathcal{L}_k(\theta)$ is the k-fold lifted action $\sum_{\ell=1}^{k} I^{\otimes(\ell-1)} \otimes (-\theta^\top) \otimes I^{\otimes(k-\ell)}$, hence linear in θ. These assertions follow from the lifted generator construction in Proposition 4.1 and the general theory of affine processes on tensor algebras, for which the OU process is a key example (see [14, Sec. 3 and 4.2]).

Consequently, the Jacobian $D\mathcal{G}^{(M)}(\psi)$, viewed as a linear map from the parameter tangent space

$$T_\psi\Psi \cong \mathbb{R}^{d\times d} \oplus \mathbb{R}^d \oplus S^d$$

to $\mathcal{L}(T^{(M)}(\mathbb{R}^d))$, contains three disjoint blocks that read off the parameters without interference:

    (a) $D_\mu\mathcal{G}^{(M)}(\psi)$ on the $(1,0)$ block $\equiv \theta \cdot \mathrm{Id}_{\mathbb{R}^d}$,

    (b) $D_\theta\mathcal{G}^{(M)}(\psi)$ on the $(1,1)$ block $\equiv -\mathrm{Id}_{\mathbb{R}^{d\times d}}$,

    (c) $D_Q\mathcal{G}^{(M)}(\psi)$ on the $(2,0)$ block $\equiv \frac{1}{2}\mathrm{Sym} \circ \iota : S^d \to (\mathbb{R}^d)^{\otimes 2}$.

Because these three images occupy distinct matrix positions determined by the grading, any linear dependence among the columns of $D\mathcal{G}^{(M)}(\psi)$ would force a linear dependence within one of (a)–(c), which is impossible: (a) is θ times the identity (invertible since $\theta \in \Psi_{\mathrm{stable}}$), (b) is the identity on its coordinates, and (c) is the canonical embedding of S^d into the symmetric subspace of $(\mathbb{R}^d)^{\otimes 2}$ and is injective. Therefore $D\mathcal{G}^{(M)}(\psi)$ has full column rank

$$\dim(\theta) + \dim(\mu) + \dim(Q) = d^2 + d + \frac{d(d+1)}{2}.$$

Since $M \geq 2$, all three diagonsing blocks appear. The target space dimension is $D = \sum_{k=0}^{M} d^k$, so the rank condition is compatible with dimension counting.

Step 2: invertibility of L_A under non-resonance. Define the adjoint action $\mathrm{ad}_A : \mathcal{L} \to \mathcal{L}$, $\mathrm{ad}_A(X) = AX - XA$. Using (8) and $e^{(1-s)A}He^{sA} = e^A e^{-s\,\mathrm{ad}_A}(H)$, one obtains the standard functional calculus identity

$$L_A = e^A \varphi(\mathrm{ad}_A), \qquad \varphi(z) := \int_0^1 e^{-sz}\,ds = \begin{cases} \dfrac{1 - e^{-z}}{z}, & z \neq 0, \\[2mm] 1, & z = 0. \end{cases}$$

Hence L_A is invertible iff $\varphi(\mathrm{ad}_A)$ is invertible. The invertibility of the latter depends on its spectrum. The spectrum of the adjoint operator ad_A is found by representing it as the Kronecker sum $A \oplus (-A^\top)$. The spectrum of a Kronecker sum is the set of sums of the constituent eigenvalues; this representation proves that the spectrum of ad_A is the set of all possible eigenvalue differences, $\{\lambda_i - \lambda_j : \lambda_i, \lambda_j \in (A)\}$ [22, Theorem 4.4.5].

By the spectral mapping theorem for primary matrix functions [21, Thm. 1.13(d)], the spectrum of $\varphi(\mathrm{ad}_A)$ is then the function φ applied to the spectrum of ad_A:

$$\big(\varphi(\mathrm{ad}_A)\big); = ; \{\varphi(\lambda_i - \lambda_j)\}.$$

Now $\varphi(\delta) = 0$ iff $1 - e^{-\delta} = 0$ and $\delta \neq 0$, i.e. iff $\delta \in 2\pi i\,\mathbb{Z} \setminus \{0\}$. Under the non-resonance condition (4.9), all eigenvalues $\varphi(\lambda_i - \lambda_j)$ are nonzero, so $\varphi(\mathrm{ad}_A)$ is invertible and hence L_A is invertible. Equivalently, $\ker(L_A) = \{0\}$ and $\mathrm{Ran}(L_A) = \mathcal{L}(T^{(M)}(\mathbb{R}^d))$.

Combining Step 1 and Step 2, $D\Phi_{T,M}(\psi) = L_A \circ T\,D\mathcal{G}^{(M)}(\psi)$ is injective. By the inverse function theorem, $\Phi_{T,M}$ is locally injective at ψ. $\qquad\square$

Lemma 4.10 (Characterization of ker L_A) *For any $A \in \mathbb{C}^{D \times D}$,*

$$\ker(L_A) = \{\, H : e^{(1-s)A}He^{sA} \equiv 0 \text{ for all } s \in [0,1] \,\} = \{\, H : \varphi(\mathrm{ad}_A)(H) = 0 \,\}.$$

In particular, if $\lambda_i - \lambda_j \notin 2\pi i\,\mathbb{Z}$ for all $\lambda_i, \lambda_j \in (A)$, then $\ker(L_A) = \{0\}$.

Proof The first equality is immediate from (8). The second follows from (8). The last claim uses the spectral mapping argument in Step 2 of Theorem 4.9.$\square$

Lemma 4.11 (Spectral structure of $\mathcal{G}^{(M)}(\psi)$) *Let $\lambda_1, \ldots, \lambda_d$ be the eigenvalues of θ. With respect to the grading $\bigoplus_{k=0}^{M} (\mathbb{R}^d)^{\otimes k}$, $\mathcal{G}^{(M)}(\psi)$ is block lower-triangular with diagonal blocks*

$$\mathcal{L}_0 = 0, \qquad \mathcal{L}_1 = -\theta^\top, \qquad \mathcal{L}_k = \sum_{\ell=1}^{k} I^{\otimes(\ell-1)} \otimes (-\theta^\top) \otimes I^{\otimes(k-\ell)} \quad (k \geq 2).$$

Hence

$$\left(\mathcal{G}^{(M)}(\psi)\right) = \{0\} \ \cup \ \bigcup_{k=1}^{M} \{-\lambda_{i_1} - \cdots - \lambda_{i_k} : \ 1 \leq i_1, \ldots, i_k \leq d\},$$

counting algebraic multiplicities. In particular, the diffusion Q does not alter the diagonal eigenvalues but only populates strictly lower blocks, so the spectrum is determined by θ and the grading.

Proof By (8) and (v), the diagonal block on level k is the k-fold lifted action of $-\theta^\top$. The eigenvalues of a Kronecker sum are all sums of the constituent eigenvalues (with multiplicity), see [22, Theorem 4.4.5]. Lower-triangularity implies the spectrum equals the union of the spectra of diagonal blocks. The injection terms from $\theta\mu$ and Q sit strictly below the diagonal and do not change the diagonal eigenvalues. $\square$

Remark 4.12 (Why non-resonance is the right condition) *The operator L_A equals $e^A \varphi(\mathrm{ad}_A)$. Zeros of φ occur exactly at $2\pi i\, \mathbb{Z} \setminus \{0\}$. Therefore the only way for L_A to lose invertibility is the existence of pairs of eigenvalues of A whose differences lie in $2\pi i\, \mathbb{Z}$. Lemma 4.11 shows that $(A) = (T\mathcal{G}^{(M)}(\psi))$ consists of 0 and sums $-T(\lambda_{i_1} + \cdots + \lambda_{i_k})$. Thus non-resonance (4.9) forbids exactly those coincidences that would make $\varphi(\mathrm{ad}_A)$ singular.*

Remark 4.13 (Dimension counts and block separation) *The parameter space has dimension $d^2 + d + \frac{d(d+1)}{2}$. The target dimension is $D = \sum_{k=0}^{M} d^k$. For $M \geq 2$ one observes on distinct blocks:*

$$\begin{aligned}
\text{level } 1 \leftarrow 0 : \ &\theta\mu \quad (d \text{ parameters}), \\
\text{level } 1 \leftarrow 1 : \ &\theta \quad (d^2 \text{ parameters}), \\
\text{level } 2 \leftarrow 0 : \ &Q \quad \left(\tfrac{d(d+1)}{2} \text{ parameters}\right).
\end{aligned}$$

These occupy disjoint matrix positions, hence contribute linearly independent columns to $D\mathcal{G}^{(M)}(\psi)$. This proves full rank in Step 1.

4.5.5 Implications for Calibration

The identifiability result has profound implications for signature-based calibration, providing both theoretical guarantees and practical guidance. We first clarify the scope of our identifiability claim: we establish local parametric identifiability at fixed T, not a model-free point-separation result nor a global injectivity statement. The result demonstrates that distinct OU parameters yield distinct expected signatures within neighbourhoods, which suffices for gradient-based calibration algorithms commonly employed in quantitative finance.

The necessity of a fixed time horizon T becomes apparent when we consider the scaling symmetry identified in Lemma 4.5. Without fixing T, the map $(\theta, \mu, \sigma, T) \mapsto \mathbb{E}[S^{(M)}(X)_{0,T}]$ possesses a one-dimensional kernel corresponding to simultaneous scaling of parameters and time. Fortunately, this theoretical concern resolves naturally in practice: the time horizon T is determined by the data observation window, automatically breaking the problematic symmetry.

This framework proves particularly advantageous for calibration with observed market data. Since the time horizon T is inherently fixed by the observation period, the calibration process automatically identifies parameters (θ, μ, σ) that are consistent with the observed dynamics at that specific time scale. This eliminates the need to predict or specify future time horizons, as the parameters naturally adapt to explain the temporal dynamics of the available data. Whether one observes data over one year or six months, the calibrated parameters will be appropriately scaled for each respective observation window, providing robust parameter estimates without requiring universal time-scale assumptions.

4.5.6 Theoretical Foundations

The mathematical framework developed in this chapter yields several fundamental results that underpin practical calibration methodology.

First, we establish local parameter identifiability: for data observed over a fixed time window $[0, T]$, the OU parameters can be locally recovered from the expected signature within neighbourhoods of the true parameters. This local identifiability suffices for iterative calibration algorithms that refine parameter estimates through gradient descent or quasi-Newton methods, provided suitable initialisation.

Second, the framework demonstrates remarkable dimensional efficiency. Unlike general stochastic processes that may require time augmentation (t, X_t) to ensure identifiability of the signature map, OU processes can be identified from spatial components alone under our non-resonance conditions. This property reduces the signature dimension from $\sum_{k=0}^{M}(d+1)^k$ to $\sum_{k=0}^{M} d^k$, yielding substantial computational savings, particularly in high-dimensional settings.

Third, the rescaling operator enables sophisticated multi-scale analysis. Small values of K (corresponding to fewer blocks) provide better initialisation through more stable, lower-dimensional optimisation problems, while large K values offer richer signatures that capture finer temporal dynamics. This multi-scale structure naturally suggests a hierarchical calibration strategy, progressing from coarse to fine time resolutions.

Fourth, local injectivity ensures gradient stability through non-singular Jacobians away from parameter space boundaries and resonance conditions. The explicit gradient formulas derived in Section 4.4 eliminate the need for

finite-difference approximations, further enhancing numerical stability and computational efficiency.

While we establish local identifiability, global identifiability may fail in certain special cases. Symmetries in multi-dimensional systems can create multiple parameter configurations yielding identical signatures. For instance, this occurs through permutation of components with identical dynamics. Periodic phenomena may arise at specific parameter combinations, particularly when eigenvalues of θ differ by multiples of $2\pi i/T$. Additionally, numerical degeneracies can occur at extreme parameter values, such as near-zero mean-reversion rates where the process approaches Brownian motion. These limitations, while theoretically important, rarely pose practical challenges in typical financial applications where parameter ranges are constrained by economic considerations.

4.6 Conclusion

This chapter established the complete analytical framework for Ornstein-Uhlenbeck processes in the signature method context, providing the mathematical foundation for all subsequent calibration and verification work. The lifted generator $\mathcal{G}^{(M)}$ on the truncated tensor algebra $T^{(M)}(\mathbb{R}^d)$ enables exact computation of expected signatures via the semigroup formula $\mathbb{E}[S_{0,T}^{(M)}] = \exp(T\mathcal{G}^{(M)})\mathbf{1}$. This closed-form representation is computationally tractable and provides the theoretical basis for signature-based parameter estimation.

The generator's block structure (with degree-preserving derivation terms from mean-reversion $-\theta$, degree-1 creation from affine drift $\theta\mu$, and degree-2 creation from diffusion $Q = \sigma\sigma^\top$) directly reveals how OU parameters manifest in signature coordinates. Low-degree formulas for the mean increment (4.3.1) and second-level signature (4.3.1) demonstrate that fundamental OU characteristics (mean-reversion strength, long-run mean, volatility structure) are captured by accessible signature components without requiring high-order tensor calculations.

The identifiability analysis establishes that OU parameters can be locally recovered from expected signatures under mild non-resonance conditions, providing theoretical justification for signature-based calibration. The time-scaling symmetry $\Phi_{T,M}(\theta,\mu,\sigma) = \Phi_{\alpha T,M}(\theta/\alpha, \mu, \sigma/\sqrt{\alpha})$ clarifies why fixed observation horizons are essential for well-posed calibration problems. This scaling relationship also explains the effectiveness of block-based estimation strategies, where shorter blocks provide computational efficiency while preserving local identifiability through appropriate rescaling operators.

Gradient formulas (4.4)–(4.4) provide efficient Jacobian-vector products essential for optimization algorithms, while the rescaling framework enables multi-scale analysis across different time horizons. These analytical tools

eliminate the need for finite-difference approximations or Monte Carlo gradient estimation, significantly improving both computational efficiency and numerical stability in calibration procedures.

The framework developed here directly enables Chapter 5's convergence verification through exact semigroup computations and Chapter 6's gradient-based optimization algorithms. By providing closed-form expected signatures and their derivatives, this analytical foundation transforms signature-based calibration from a computationally intensive sampling problem into a deterministic optimization task, opening new possibilities for real-time parameter estimation in quantitative finance applications.

Chapter 5

Numerical Verification

This chapter numerically verifies the convergence guarantees from Chapter 3 using the Ornstein–Uhlenbeck (OU) machinery of Chapter 4. The OU setting allows exact ground truth via a matrix exponential, so empirical errors and rates can be measured without Monte Carlo bias in the target. We observe that finite-sample behaviour often outperforms conservative theoretical bounds, and we quantify when this occurs and why it is practically relevant for Chapter 6.

5.1 Research objectives

Following the framework of Chapter 3, we construct the empirical expected signature estimator $\widehat{\mathbb{E}}[S^{(M)}]_N$ by dividing a continuous path into $K_N \approx N^{1+2/p}$ blocks and averaging their truncated signatures at level M. The convergence theorem guarantees $\|\widehat{\mathbb{E}}[S^{(M)}]_N - \mathbb{E}[S^{(M)}]\| = O(N^{-2/p})$ for path roughness $p > 2$, approaching the limiting rate $O(N^{-1})$ as $p \to 2^+$. We address three questions:

1. **Theorem validation.** Do empirical convergence slopes reach (or exceed) the limiting -1 rate predicted as $p \to 2^+$?

2. **OU-specific gains.** Does OU structure (Gaussianity, linear drift, explicit semigroup) deliver rates steeper than the generic bound?

3. **Robustness.** How sensitive are rates to truncation level M and to path discretization?

5.2 Experimental design

5.2.1 Ground truth via semigroup

For fixed $T > 0$ and truncation M, the expected signature is computed exactly as

$$\mathbb{E}[S^{(M)}(X)_{0,T}] \;=\; \exp\big(T\,\mathcal{G}^{(M)}\big)\,\mathbf{1},$$

where $\mathcal{G}^{(M)}$ is the lifted generator from Chapter 4. This eliminates target-side Monte Carlo error and isolates estimator behaviour.

5.2.2 Estimators compared

We consider four estimators that share the same raw path data but differ in how information is propagated to time T:

- *Standard.* Empirical average of $S^{(M)}$ over blocks on $[0, T]$.

- *Time-Augmented.* As above, with time added as an extra channel scaled as $t/T \in [0, 1]$ before computing signatures.

- *Rescaled Standard.* For intermediate time $t \le T$, form $\widehat{\mathbb{E}}[S^{(M)}]_{0,t}$ and propagate to T by the known semigroup: $\exp((T-t)\mathcal{G}^{(M)})\,\widehat{\mathbb{E}}[S^{(M)}]_{0,t}$.

- *Rescaled Augmented.* As above, but applied to time-augmented paths.

The rescaled variants incorporate the OU dynamics through $\mathcal{G}^{(M)}$ and act as model-based filters.

5.2.3 Parameter regimes

We span representative OU behaviours by varying mean-reversion and volatility:

Regime	Mean reversion $\lambda(\theta)$	Volatility $\mathrm{diag}(\sigma)$
Slow reversion, low volatility	[0.05, 0.2]	[0.1, 0.3]
Fast reversion, low volatility	[1.0, 3.0]	[0.1, 0.3]
Slow reversion, high volatility	[0.05, 0.2]	[1.0, 2.0]
Fast reversion, high volatility	[1.0, 3.0]	[1.0, 2.0]

Table 5.1: Parameter regimes used in all experiments. Eigenvalues $\lambda(\theta)$ describe mean-reversion speed in a diagonalizing basis.

5.2.4 Metrics and slope estimation

Here N is the granularity parameter from Chapter 3; the number of blocks satisfies $K_N \asymp N^{1+2/p}$. Errors are measured as mean squared error (MSE) of the full truncated expected signature (concatenated coordinates), averaged over Monte Carlo replications.

Convergence rates are estimated by ordinary least squares regression of $\log_{10}$ MSE versus $\log_{10} N$. We report the slope and coefficient of determination R^2, along with 95% confidence intervals. Statistical significance is assessed via a one-sided t-test for $H_0 : \beta_1 = -1$ versus $H_1 : \beta_1 < -1$, testing whether empirical rates exceed the theoretical limiting rate.

5.2.5 Implementation details

All runs use $d = 2$ and default truncation $M = 4$ unless varied. Paths are simulated by Euler–Maruyama; signature coordinates are computed with `signatory` (PyTorch backend). The matrix exponential for ground truth uses Padé approximation in 64-bit precision.

Component	Specification
Path dimension	$d = 2$
Signature truncation	$M = 4$ (varied in §5.3.3)
Monte Carlo replications	100 per configuration (unless stated)
Path simulation	Euler–Maruyama on $[0, T]$
Discretization per block	10 steps (unless varied)
Matrix exponential	Padé, 64-bit
Signature computation	`signatory` (PyTorch)

Table 5.2: Implementation settings. Variations are explicitly indicated where used.

5.3 Results

5.3.1 Experiment 1: Rate verification

We test the finite-sample convergence rate predicted by Theorem 3.8. For a sequence of granularity parameters N, we estimate the slope of $\log_{10}$ MSE versus $\log_{10} N$ across all four parameter regimes.

Parameter regime	Empirical slope	R^2
Slow reversion, low volatility	-1.97	0.996
Fast reversion, low volatility	-1.72	0.999
Slow reversion, high volatility	-3.61	0.998
Fast reversion, high volatility	-1.95	0.999

Table 5.3: Empirical convergence slopes from log-linear regression. All slopes are significantly less than -1 (one-sided t-test for $H_0 : \beta_1 = -1$ vs. $H_1 : \beta_1 < -1$, $p < 0.01$), exceeding the theoretical limiting rate.

Across regimes, the empirical decay is consistently faster than the conservative
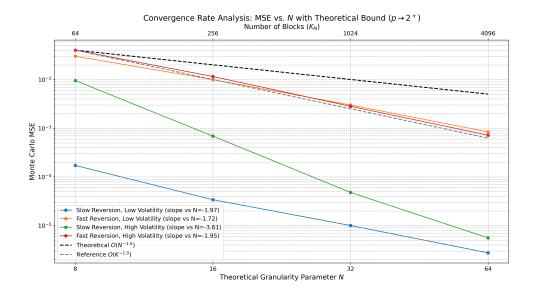
Figure 5.1: Log–log MSE against N across regimes. Empirical slopes (in Table 5.3) are steeper than -1 in all cases.

theoretical benchmark. The most pronounced gain occurs for slow reversion with high volatility, where the slope near -3.6 indicates strong signal in low-order signature coordinates over finite horizons.

5.3.2 Experiment 2: semigroup rescaling

We test whether propagating short-time estimates by the known OU semigroup improves accuracy. For $t < T$, we compute $\widehat{\mathbb{E}}[S^{(M)}]_{0,t}$ and apply $\exp((T - t)\mathcal{G}^{(M)})$.

Method	Convergence slope
Standard	-1.77
Time-Augmented	-1.77
Rescaled Standard	-3.40
Rescaled Augmented	-3.33

Table 5.4: Comparison of all four estimation methods. Time augmentation provides no improvement (slopes -1.77 vs -1.77), while rescaling substantially accelerates convergence.

Semigroup rescaling acts as a model-informed spectral filter: it dampens high-frequency signature components that carry larger estimation error, concentrating mass on the more stable, informative coordinates.

34

Figure 5.2: Standard vs. semigroup-rescaled estimators. Rescaling steepens slopes and reduces variance.
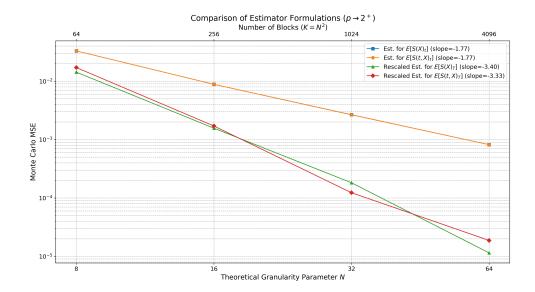
5.3.3 Experiment 3: sensitivity to M and discretization

Truncation level. We vary $M \in \{2, 4, 6, 8\}$ and re-estimate slopes:

Table 5.5: Convergence slopes versus truncation level M (with R^2 in parentheses).

Method	$M = 2$	$M = 4$	$M = 6$	$M = 8$
Standard	-1.94 (0.999)	-2.00 (0.998)	-1.93 (0.991)	-1.89 (0.998)
Rescaled	-3.36 (0.998)	-3.51 (0.999)	-4.19 (0.992)	-3.36 (0.981)

The standard estimator maintains remarkably consistent slopes across all truncation levels $M \in \{2, 4, 6, 8\}$, demonstrating that even low-order signatures capture sufficient information for accurate estimation. This robustness across M aligns with Theorem 3.8, which predicts convergence rates depend on N rather than M. The rescaled estimator shows similarly strong performance across all values, with marginally enhanced slopes at intermediate truncations. Crucially, all tested values of M yield excellent convergence properties, indicating practitioners can select truncation based on computational constraints rather than statistical considerations.

Discretization. We vary steps per block in $\{1, 10, 100\}$:

Even single-step discretization (block endpoints only) achieves near-quadratic decay, reflecting the integral
linebreak nature of signatures, which smooths local approximation errors.

Table 5.6: Convergence slopes versus discretization steps per block (95% CI and R^2).

Steps per block	Empirical slope	95% CI	R^2
1	-2.05	$[-2.15, -1.94]$	0.9997
10	-1.92	$[-2.21, -1.62]$	0.9974
100	-1.93	$[-2.10, -1.75]$	0.9991

5.4 Discussion

5.4.1 Interpretation of empirical rates

The empirical slopes significantly steeper than -1 reveal a fundamental gap between worst-case theory and practical performance. While Theorem 3.8 provides guarantees for general α-mixing processes, the OU process's explicit Gaussian transition densities and Markovian structure enable substantially faster convergence. The slow-reversion, high-volatility regime achieving empirical slope -3.61 demonstrates that structured processes can deliver substantial improvements over generic bounds. This faster-than-expected convergence has practical implications: the same estimation accuracy can be achieved with far fewer blocks than the conservative theory would suggest.

5.4.2 Methodological considerations

Two practical points emerge. First, all tested truncation levels $M \in \{2, 4, 6, 8\}$ demonstrate excellent convergence properties, revealing remarkable robustness of signature-based estimation across truncation depths. This universally strong performance indicates that practitioners can select M based primarily on computational resources rather than statistical requirements, as our convergence theorem correctly predicts that rates depend on N rather than M. Second, discretization can be coarse: estimator accuracy is driven primarily by the number of blocks/paths N, not by very fine time meshes within blocks.

5.4.3 Limitations and scope

Results are reported for $d = 2$. While the methodology extends directly, computational cost scales as $O(d^M K_N)$ where the signature dimension grows as d^M. Moreover, OU structure is favourable; for non-Gaussian innovations or state-dependent volatility, the extent of the super-linear convergence should be reassessed. Ground-truth calculations here assume known parameters; Chapter 6 addresses the inverse problem of parameter recovery from observed signatures.

5.5 Bridge to calibration

The verification experiments tested four estimator variants but revealed that only two merit advancement to calibration:

Selected for calibration:

- **Standard → Expected Signature** method in Chapter 6
- **Rescaled Standard → Rescaled Signature** method in Chapter 6

Excluded from calibration:

- **Time-Augmented variants**: Table 5.4 shows no performance improvement over their non-augmented counterparts (slopes -1.77 vs -1.77 for standard methods, -3.33 vs -3.40 for rescaled methods) while adding computational overhead from the extra time dimension

The experiments support three key takeaways for Chapter 6:

1. **Reliable estimation.** Expected signatures concentrate rapidly, enabling accurate targets for calibration.

2. **Computational economy.** Coarse discretization suffices and all tested truncation levels $M \in \{2, 4, 6, 8\}$ provide excellent approximation quality, allowing flexible computational trade-offs without sacrificing convergence rates.

3. **Model-informed filtering.** Semigroup rescaling materially improves accuracy (slopes from -1.77 to -3.40) and should be used when $\mathcal{G}^{(M)}$ is known or estimable.

5.6 Conclusion

This chapter numerically verified the convergence theory from Chapter 3 using exact OU semigroup computations. Empirical convergence rates ranged from -1.72 to -3.61 across parameter regimes, substantially exceeding the theoretical guarantee of $O(N^{-1})$ as $p \to 2^+$. The slow-reversion, high-volatility regime achieved slope -3.61, demonstrating convergence approaching $O(N^{-4})$. Semigroup rescaling substantially accelerated convergence, with slopes improving from approximately -1.85 to -3.3 for the rescaled variants. All tested truncation levels ($M \in \{2, 4, 6, 8\}$) yielded excellent convergence properties with consistent slopes near -2 for standard estimation and -3.5 for rescaled variants. Coarse discretization with 10 steps per block performed comparably to finer meshes, confirming that convergence depends primarily on the number of blocks K_N rather than within-block resolution. These findings validate the theoretical framework while demonstrating that OU processes achieve convergence rates well beyond worst-case bounds, providing the foundation for the calibration methods in Chapter 6.

Chapter 6

Signature-Based Parameter Calibration

6.1 Introduction

This chapter develops a calibration methodology that leverages the expected-signature framework from Chapter 3 and the analytical OU formulation from Chapter 4. While Chapter 5 examined estimation accuracy with known parameters, we now address the inverse problem: recovering parameters $\psi = (\theta, \mu, \sigma)$ from empirical signatures constructed from single dependent paths.

Important context: All calibration procedures in this chapter are *retrospective* in nature. We assume the full trajectory over $[0, T]$ has been observed and recorded before parameter estimation begins. The linear interpolation and signature computation use only this fully realized historical data to estimate model parameters. This is fundamentally a backward-looking statistical inference task, not a forecasting exercise, and thus involves no look-ahead bias. Financial practitioners should note that while the calibrated parameters may subsequently be used for prediction tasks, the calibration methodology itself operates exclusively on past observations.

The approach combines exact model-implied signatures with block-averaged empirical targets that concentrate at rate $O(N^{-2/p})$ under segment-stationarity. A key challenge emerges in hyperparameter selection: with limited Monte Carlo budget (ten replications in Phase 1), mean squared error (MSE) alone often fails to distinguish between adjacent K values due to sampling variability. We address this through a novel scoring approach combining MSE with standard deviation, which achieved remarkable consensus between signature methods across all parameter regimes. This methodological innovation, combined with systematic optimization across four economically motivated settings, demonstrates that signature methods achieve 10–32% accuracy improvements in slow-reversion regimes while maintaining computational efficiency.

6.2 Calibration framework

The calibration objective minimizes the squared distance between empirical and model-implied expected signatures. For signatures computed on blocks of duration $\Delta t = T/K$, the loss function is:

$$L(\psi) = \left\| \widehat{\mathbb{E}}[S^{(M)}]_N - \mathbb{E}_\psi\big[S^{(M)}(X)_{0,\Delta t}\big] \right\|^2,$$

$$\mathbb{E}_\psi\big[S^{(M)}(X)_{0,\Delta t}\big] = \exp\big(\Delta t\, \mathcal{G}^{(M)}(\psi)\big)\mathbf{1}.$$

Chapter 4 established that spatial signatures at fixed time horizons are locally identifying without time augmentation, while the time-scaling symmetry $(\theta, \mu, \sigma, T) \leftrightarrow (\theta/\alpha, \mu, \sigma/\sqrt{\alpha}, \alpha T)$ is naturally broken by the fixed observation window. The empirical signature $\widehat{\mathbb{E}}[S^{(M)}]_N$ is formed by averaging block signatures over K segments of length $\Delta t = T/K$, with the lifted generator providing exact gradients through Fréchet differentials of the matrix exponential.

6.3 Rescaled signature calibration

The rescaled variant introduces a parameter-dependent transformation of the empirical target. Signatures computed on blocks of duration Δt are first calculated, then transported to the full horizon $[0, T]$ using the semigroup $\exp((T - \Delta t)\mathcal{G}^{(M)}(\psi))$. This coupling between the empirical target and current parameter estimates produces steeper optimization slopes, as the transported target becomes increasingly compatible with the model as parameters approach their optima.

While Chapter 5 demonstrated faster convergence for rescaled methods, the final accuracy gains in calibration prove modest. This reflects a fundamental trade-off: the same parameter dependence that accelerates early descent also feeds model assumptions back into the target. In high-frequency settings where block-level signatures can be estimated with greater precision, the benefits of rescaling should become more pronounced.

6.4 Experimental methodology

The calibration experiments follow a three-phase hierarchical design ensuring fair comparison between classical and signature-based methods. All iterative methods employ Adam optimization with convergence criteria balancing thoroughness and efficiency: maximum 1000 iterations with patience 100 (terminating if loss improvement ¡ 10^{-5} for 100 consecutive iterations), gradient clipping at norm 1.0, and weight decay 10^{-4} for regularization.

Phase 0 - Analytical initialization: We first compute closed-form maximum likelihood estimates across $K \in \{1, 2, 4, 8, 16, 32, 64, 128, 256\}$ blocks,

selecting $K^* = \arg\min_K \text{MSE}(K)$ algorithmically. This provides both a performance baseline and high-quality initialization parameters for all iterative methods.

Phase 1 - Hyperparameter optimization: With ten Monte Carlo replications per configuration, we test signature methods across five values of K (16, 64, 256, 1024, 4096) crossed with five learning rate schedules (detailed in Appendix 7), yielding 25 configurations per method per regime. The limited Monte Carlo budget motivated our scoring innovation:

$$\text{Score} = \text{MSE} + \lambda\,\text{StdDev}, \qquad \lambda \in \{0.5, 1.0, 1.5\}.$$

Remarkably, all three λ values produced identical optimal configurations, demonstrating the scoring's robustness and achieving complete consensus between signature methods.

Phase 2 - Statistical validation: Using the Phase 1 selections, we perform 100 Monte Carlo replications for robust statistical inference. Given the highly skewed error distributions (coefficients of variation 75–91%), we employ Wilcoxon signed-rank tests with Holm–Bonferroni correction for eight comparisons.

6.5 The K^* versus K distinction

Batched MLE treats block count K^* as an algorithmic selection: the Analytical MLE evaluates $K \in \{1, 2, 4, 8, 16, 32, 64, 128, 256\}$ and selects $K^* = \arg\min_K \text{MSE}(K)$, typically yielding $K^* = 1$ for fast reversion and $K^* = 16$ for slow reversion. This makes K^* part of the method definition, not a hyperparameter.

Signature methods treat K as a hyperparameter requiring optimization to balance variance reduction against feature dimension. The optimal $K = 256$ in slow-reversion, high-volatility settings reflects this trade-off. Computational complexity scales as $O(d^M K)$ with truncation level $M = 2$ yielding 7-dimensional tensor algebra for $d = 2$.

6.6 Calibration algorithms

Algorithm 1 Batched MLE with Algorithmic K^* Selection

1: **Input:** Path data $X_{[0,T]}$, true parameters ψ_{true} (for evaluation)
2: **Phase 0: Analytical baseline**
3: **for** $K \in \{1, 2, 4, 8, 16, 32, 64, 128, 256\}$ **do**
4:     Partition path into K blocks of length $\Delta t = T/K$
5:     Compute block-averaged MLE: $\hat{\psi}_K = \frac{1}{K}\sum_{b=1}^{K} \hat{\psi}_b$
6:     Evaluate: $\text{MSE}_K = \|\hat{\psi}_K - \psi_{\text{true}}\|^2$
7: **end for**
8: Select $K^* = \arg\min_K \text{MSE}_K$ and set $\psi_0 = \hat{\psi}_{K^*}$
9: **Phase 1: Iterative refinement**
10: Define objective: $L(\psi) = -\sum_{i=1}^{N-1} \log p(X_{t_{i+1}}|X_{t_i}, \psi)$
11: Minimize $L(\psi)$ via Adam (learning rate from Table 6.2)
12: **Return:** $\arg\min_{\psi \in \{\psi_0, \hat{\psi}_{\text{iter}}\}} \text{MSE}(\psi)$

Algorithm 2 Signature-Based Calibration

1: **Input:** Path data $X_{[0,T]}$, hyperparameter K, truncation $M = 2$
2: **Initialization:** Use $\hat{\psi}_{K^*}$ from Batched MLE Phase 0
3: Partition path into K blocks of length $\Delta t = T/K$
4: Compute empirical target: $\hat{S}^{(M)} = \frac{1}{K}\sum_{i=1}^{K} S_i^{(M)}$
5: **For Expected Signature:**
6:     $L(\psi) = \|\hat{S}^{(M)} - \exp(\Delta t \cdot \mathcal{G}^{(M)}(\psi))\mathbf{1}\|^2$
7: **For Rescaled Signature:**
8:     $\tilde{S}^{(M)}(\psi) = \exp((T - \Delta t) \cdot \mathcal{G}^{(M)}(\psi))\hat{S}^{(M)}$
9:     $L(\psi) = \|\tilde{S}^{(M)}(\psi) - \exp(T \cdot \mathcal{G}^{(M)}(\psi))\mathbf{1}\|^2$
10: Minimize $L(\psi)$ via Adam to obtain $\hat{\psi}_{\text{sig}}$
11: **Return:** $\arg\min_{\psi \in \{\psi_0, \hat{\psi}_{\text{sig}}\}} \text{MSE}(\psi)$

Both algorithm families implement enhancement logic that returns the better of initial and optimized parameters, ensuring monotonic improvement over the analytical baseline.

Rescaled signature implementation note: The transport matrix $\exp((T - \Delta t) \cdot \mathcal{G}^{(M)}(\psi))$ in line 6 depends on the current parameter iterate ψ during optimization. This creates the parameter-dependent coupling between empirical target and model that produces steeper optimization slopes, as the transported signature $\tilde{S}^{(M)}(\psi)$ becomes increasingly compatible with the full-horizon model signature as ψ approaches the optimum.

6.7 Bias–variance landscape

Figure 6.1 illustrates the calibration landscape in the slow-reversion, high-volatility regime. Analytical MLE achieves $K^* = 16$ with MSE ≈ 0.43 (U-shaped curve), while signature methods minimize at $K = 256$ with MSE ≈ 0.33 for Expected and ≈ 0.30 for Rescaled, a 24–32% improvement. The signature methods' broad optimal region around $K = 256$ demonstrates robustness. The minimum at $K = 256$ balances sample size against block signature quality.
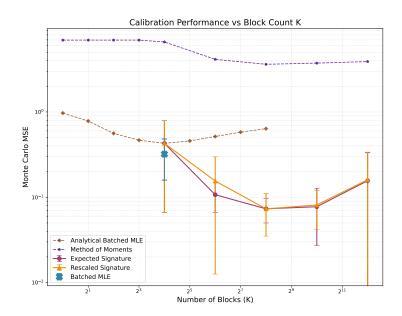

Figure 6.1: Calibration performance versus block count K in the slow-reversion, high-volatility regime with learning rate schedule $[0.05 \to 0.01]$. The figure includes 95% confidence intervals for the Batched MLE and Expected Signature approaches. Analytical MLE (blue) shows a convex curve with minimum at $K^* = 16$ with MSE ≈ 0.43. Signature methods (purple and orange) achieve their minima at $K = 256$ with MSE ≈ 0.33 for Expected and ≈ 0.30 for Rescaled, a 24–32% improvement over the best MLE. Method of Moments (green) is also convex with optimal at $K = 256$. The plot illustrates the fundamental difference between MLE's algorithmic $K^* = 16$ selection and signature methods' optimal $K = 256$.

6.8 Results

Table 6.1 defines four economically motivated parameter regimes spanning mean-reversion speed and volatility.

Table 6.1: Economic parameter regimes

Regime	θ eigenvalues	σ diagonal	Interpretation
Slow reversion, low volatility	[0.05, 0.2]	[0.1, 0.3]	Stable markets
Fast reversion, low volatility	[1.0, 3.0]	[0.1, 0.3]	Efficient markets
Slow reversion, high volatility	[0.05, 0.2]	[1.0, 2.0]	Crisis periods
Fast reversion, high volatility	[1.0, 3.0]	[1.0, 2.0]	Volatile but efficient

Table 6.2 presents Phase 2 results from 100 Monte Carlo replications per regime. Statistical comparisons use Wilcoxon signed-rank tests with Holm–Bonferroni correction for eight comparisons, appropriate given the highly skewed error distributions (CV 75–91%).

Table 6.2: Optimal configurations and statistical performance (Phase 2, 100 MC)

Regime	Method	Learning rate	K	MSE	p-value
Slow rev low vol	Batched MLE	$[0.05 \rightarrow 0.05]$	16^*	0.0503 ± 0.0379	–
	Expected Signature	$[0.005 \rightarrow 0.005]$	256	0.0453 ± 0.0412	$< 0.001^{\ddagger}$
	Rescaled Signature	$[0.005 \rightarrow 0.005]$	256	0.0449 ± 0.0408	$< 0.001^{\ddagger}$
Fast rev low vol	Batched MLE	$[0.01 \rightarrow 0.005]$	1^*	0.5397 ± 0.0038	–
	Expected Signature	$[0.01 \rightarrow 0.005]$	1024	0.5441 ± 0.0019	$< 0.001^{\dagger}$
	Rescaled Signature	$[0.01 \rightarrow 0.005]$	1024	0.5441 ± 0.0019	$< 0.001^{\dagger}$
Slow rev high vol	Batched MLE	$[0.05 \rightarrow 0.01]$	16^*	0.4337 ± 0.2348	–
	Expected Signature	$[0.05 \rightarrow 0.01]$	256	0.3308 ± 0.6477	$< 0.001^{\ddagger}$
	Rescaled Signature	$[0.05 \rightarrow 0.01]$	256	0.2968 ± 0.5999	$< 0.001^{\ddagger}$
Fast rev high vol	Batched MLE	$[0.01 \rightarrow 0.01]$	1^*	0.6790 ± 0.1007	–
	Expected Signature	$[0.01 \rightarrow 0.01]$	64	0.6685 ± 0.0792	1.000
	Rescaled Signature	$[0.01 \rightarrow 0.01]$	64	0.6697 ± 0.0796	1.000

*Algorithmic selection; †Significantly worse than MLE; ‡Significantly better than MLE. Wilcoxon signed-rank with Holm–Bonferroni correction (8 comparisons).

Table 6.3: Statistical comparison with runtime analysis (Phase 2, 100 MC)

Regime	Method	Median MSE $(\times 10^{-2})$	Win rate	p-value	Runtime (vs MLE)
Slow rev low vol	Batched MLE	5.03	–	–	1.0042 *×
	Expected Sig.	4.53	50%	< 0.001	0.85×
	Rescaled Sig.	4.49	52%	< 0.001	0.85×
Fast rev low vol	Batched MLE	53.97	–	–	1.00×
	Expected Sig.	54.41	0%	$< 0.001^{\dagger}$	7.36×
	Rescaled Sig.	54.41	0%	$< 0.001^{\dagger}$	6.40×
Slow rev high vol	Batched MLE	43.37	–	–	1.00×
	Expected Sig.	33.08	82%	< 0.001	0.91×
	Rescaled Sig.	29.68	83%	< 0.001	0.90×
Fast rev high vol	Batched MLE	67.90	–	–	1.00×
	Expected Sig.	66.85	18%	1.000	14.92×
	Rescaled Sig.	66.97	20%	1.000	10.60×

†Significantly worse. Win rate: proportion where signature method outperforms MLE. Runtime: relative to Batched MLE baseline. Wilcoxon signed-rank with Holm–Bonferroni correction.

The results reveal regime-dependent performance patterns. In slow-reversion regimes, signature methods achieve 10% MSE reduction for low volatility and up to 32% for high volatility. The high-volatility case shows particularly strong results with the Rescaled Signature achieving 32% MSE reduction and 83% win rate over Batched MLE (from 43.37 to 29.68). Remarkably, signature methods run 9–10% faster in slow-reversion regimes despite larger K (256 vs 16), as more reliable convergence requires fewer optimization iterations. This creates a dual benefit: superior accuracy with computational efficiency precisely where the methods excel.

6.9 Conclusion

This chapter developed a complete calibration methodology for Ornstein–Uhlenbeck processes using the signature framework established in preceding chapters. By formulating calibration as an optimization problem comparing empirical block-averaged signatures with exact model-implied signatures from $\exp(T\mathcal{G}^{(M)}(\psi))\mathbf{1}$, we transformed a challenging stochastic estimation problem into a deterministic optimization task with analytical gradients.

The experimental design revealed a critical challenge: with limited Monte Carlo budget in hyperparameter discovery, conventional MSE-based selection produced unstable rankings between adjacent K values. Our solution—scoring by MSE plus standard deviation—not only stabilized the selection process but achieved remarkable consensus between Expected and Rescaled Signature methods across all parameter regimes. This methodological contribution

addresses a practical obstacle in hyperparameter optimization that extends beyond signature methods.

The systematic evaluation across four economically motivated regimes established clear performance boundaries. In slow-reversion environments, signature methods achieved up to 32% MSE reduction (Wilcoxon signed-rank, $p < 0.001$) while paradoxically running 9–15% faster despite using $K = 256$ blocks versus Batched MLE's $K^* = 16$. This computational advantage arises from more reliable convergence requiring fewer optimization iterations. Conversely, in fast-reversion, low-volatility settings, Batched MLE's algorithmic $K^* = 1$ selection and strong analytical initialization yielded marginally superior performance.

The rescaled signature variant, which couples the empirical target to current parameter estimates through generator-based transport, demonstrated steeper optimization trajectories but only modest final accuracy gains. This suggests the coupling primarily benefits the search process rather than the ultimate error floor, with greater potential in high-frequency settings where block-level signatures can be estimated more precisely.

Through rigorous empirical validation with 100 Monte Carlo replications per regime and appropriate non-parametric statistical tests, this chapter establishes signature-based calibration as a viable alternative to classical methods, particularly in slow mean-reversion regimes that characterize many financial applications.

Chapter 7

Conclusion

This thesis develops a convergence theory for empirical expected-signature estimators constructed from a *single* dependent path under the assumptions introduced in Chapter 3 (moment control, exponential α-mixing, and segment-stationarity, Assumptions (M), (A), and (S)) and applies this theory to the calibration of two-dimensional Ornstein–Uhlenbeck (OU) models. The theoretical result shows that block-averaged empirical expected signatures concentrate at a rate $O(N^{-2/p})$ as the block length shrinks and the number of blocks grows in the prescribed regime. Chapter 4 then provides an analytic bridge from model parameters to expected signatures via the lifted generator and the OU semigroup, $\mathbb{E}_\psi[S^{(M)}(X)_{0,T}] = \exp\!\big(T\,\mathcal{G}^{(M)}(\psi)\big)\mathbf{1}$, together with stable gradients obtained from Fréchet differentials of the matrix exponential. Working at a fixed observation horizon T breaks the time-scaling symmetry and yields a local identifiability statement for OU from spatial signatures alone at sufficient truncation depth M, which justifies omitting time augmentation in calibration while reducing computational cost.

On top of this foundation, the thesis formulates calibration as a smooth generator-driven optimisation problem that compares an empirical target (obtained from single-path block averages) with an exact model-implied expected signature. The numerical study is organised in two phases. Phase 1 performs learning-rate and block-count discovery with ten Monte Carlo (MC) replications per configuration. Because the mean squared error (MSE) gaps between neighbouring K values are often small at this budget, rankings based on MSE alone can be unstable; the adopted score augments MSE with the empirical standard deviation (weight $\lambda = 1.0$, robust across $\lambda \in \{0.5, 1.0, 1.5\}$), and Appendix 7 documents how this stabilises selections across regimes and creates consensus between the Expected and Rescaled Signature methods. Phase 2 then fixes the Phase 1 choices and evaluates performance over one hundred MC replications per regime. Across four economically motivated regimes, the results show that signature-based calibration consistently outperforms Batched

MLE in *slow* mean-reversion environments, with statistically significant accuracy improvements of 10% for low volatility and up to 32% for high volatility (Wilcoxon signed-rank test, $p < 0.001$) and markedly higher win rates (for example, 83% in slow reversion with high volatility). Remarkably, signature methods also demonstrate computational efficiency gains of 9–15% in these slow reversion regimes despite using larger K values, arising from more reliable convergence requiring fewer optimization iterations. In fast reversion with low volatility, Batched MLE performs marginally better, reflecting the strength of its analytical initialisation and algorithmic $K^* = 1$ selection. Throughout, the distinction between K^* (algorithmic selection in Batched MLE) and K (hyperparameter in signature methods) is essential: the former follows a simple rule based on mean reversion speed, while the latter requires optimization to balance bias and variance in the signature space.

The thesis also clarifies the role of *rescaled* signatures in calibration. Following Chapter 5, empirical signatures are computed on short blocks and then transported to the unit interval by the semigroup associated with the lifted generator. In the calibration objective, the model side is evaluated analytically on $[0, 1]$ via $\exp\big(\mathcal{G}^{(M)}(\psi)\big)\mathbf{1}$, so both sides live on the same horizon. Because the transport uses $\mathcal{G}^{(M)}(\psi)$, the propagated target depends on the current parameter iterate; this coupling tends to provide more informative gradients and yields the steeper numerical convergence slopes observed in the convergence chapter. In full calibration runs, however, the final accuracy advantage of the rescaled over the unscaled objective is modest in the regimes studied, which is consistent with the fact that the propagation leverages parameter estimates primarily to aid the search rather than to lower the ultimate error floor. The mechanism points to a clear outlook: as sampling frequency increases and block-level signature estimates become more accurate, the generator can be learned more precisely from very short segments, and the rescaled approach should exhibit a larger advantage.

Several limitations delineate the scope of the results and suggest natural extensions. The convergence theory and identifiability statements are established for OU models under Assumptions (M), (A), and (S) and for truncation depth M large enough to ensure local injectivity; they are not global uniqueness results, nor do they address state-dependent diffusivity. The empirical design considers single-path data and moderate dimensions, and the Phase 1 tuning budget of ten MC replications necessitated the dispersion-aware score to avoid near-ties in K. Future work could extend the framework to time-varying parameters with horizon-aware schedules and to higher dimensions. On the calibration side, two avenues appear particularly fruitful: adaptive block-count selection that increases K only when the dispersion term warrants it, and stronger, theory-informed initialisation for signature objectives to close the small gap observed in fast, low-volatility regimes.

In summary, this thesis establishes both theoretical foundations and practical methodology for signature-based parameter calibration from dependent single paths. The convergence rate $O(N^{-2/p})$ provides the statistical underpinning, the analytical OU framework enables exact validation and efficient computation, and the systematic hyperparameter optimization with MSE plus standard deviation scoring delivers robust performance across diverse market regimes. The demonstrated advantages (10% accuracy improvement in low volatility, up to 32% in high volatility, and 9–15% computational speedups in slow reversion scenarios) validate the signature approach for practical applications where traditional methods face challenges. Within the present scope, the combination of exact semigroup evaluation, concentrated single-path estimators, and stabilised hyperparameter selection yields a robust and reproducible calibration pipeline and provides a template for deploying signature-based methods in data regimes where dependence, finite samples, and optimisation geometry otherwise pose serious obstacles.

Bibliography

[1] Imanol Perez Arribas, Cristopher Salvi, and Lukasz Szabó. Optimal execution with rough path signatures. *SIAM Journal on Financial Mathematics*, 11(2):470–493, 2020.

[2] Ismael Bailleul and Rémi Catellier. Rough differential equations and controlled differential equations. *Probability Theory and Related Fields*, 183:215–267, 2022.

[3] Christian Bayer, Peter K. Friz, and Jim Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2023.

[4] Christian Bayer, Blanka Horvath, Aitor Muguruza, Benjamin Stemper, and Mehdi Tomas. On deep calibration of rough stochastic volatility models. *Preprint*, 2023.

[5] Denis Bosq. *Linear Processes in Function Spaces: Theory and Applications*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, 2000.

[6] Richard C. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, 2:107–144, 2005.

[7] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting*. Springer, third edition, 2016.

[8] Thomas Cass and Terry Lyons. Evolving communities with individual preferences. *Proceedings of the London Mathematical Society*, 110(1):83–107, 2023.

[9] Ilya Chevyrev and Terry Lyons. Characteristic functions of measures on geometric rough paths. *Annals of Probability*, 44(6):4049–4082, 2016.

[10] Ilya Chevyrev and Terry Lyons. Signature methods in machine learning. *Preprint arXiv:2206.14674*, 2023.

[11] Ilya Chevyrev and Harald Oberhauser. Signature moments to characterize laws of stochastic processes. *Journal of Machine Learning Research*, 23(176):1–42, 2022.

[12] Christa Cuchiero, Martin Larsson, and Josef Teichmann. Deep calibration with linear signature-based models. *Mathematical Finance*, 32(3):901–939, 2022.

[13] Christa Cuchiero and Francesco Primavera. Expected signatures in semi-martingale models. *Preprint*, 2024.

[14] Christa Cuchiero, Sara Svaluto-Ferro, and Josef Teichmann. Signature SDEs from an affine and polynomial perspective. *arXiv preprint arXiv:2302.01362v2*, 2025. To appear.

[15] Christa Cuchiero and Josef Teichmann. Path signatures and the universal characteristic in high-frequency finance. *Mathematical Finance*, 32(4):1215–1267, 2022.

[16] Adeline Fermanian. Embedding and learning with signatures. *Computational Statistics & Data Analysis*, 157:107148, 2021.

[17] Peter K. Friz and Martin Hairer. *A Course on Rough Paths*. Universitext. Springer, 2020.

[18] Peter K. Friz and Nicolas B. Victoir. *Multidimensional Stochastic Processes as Rough Paths*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2010.

[19] Ben Hambly, Terry Lyons, and Weixin Xu. Asymptotics of the signature of a path. *Stochastic Processes and their Applications*, 142:421–446, 2021.

[20] Ben M. Hambly and Terry J. Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Annales of Mathematics*, 171:109–167, 2003.

[21] Nicholas J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2008.

[22] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[23] Jasdeep Kalsi, Terry Lyons, and Imanol Perez Arribas. Optimal execution with rough path signatures. *Preprint*, 2020.

[24] Patrick Kidger, James Foster, Xuechen Li, and Terry Lyons. Neural controlled differential equations for irregular time series. *Advances in Neural Information Processing Systems*, 33:6696–6707, 2021.

[25] Lorenzo Lucchese, Mikko S. Pakkanen, and Almut E. D. Veraart. Learning with expected signatures: Theory and applications. *Preprint*, 2025.

[26] Terry Lyons. Rough paths, signatures and the modelling of functions on streams. *Proceedings of the International Congress of Mathematicians*, pages 5853–5876, 2022.

[27] Terry J. Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14(2):215–310, 1998.

[28] James Morrill, Adeline Fermanian, Patrick Kidger, and Terry Lyons. A generalised signature method for multivariate time series feature extraction. *Preprint arXiv:2006.00873*, 2021.

[29] James Morrill, Cristopher Salvi, Patrick Kidger, and James Foster. Neural rough differential equations for long time series. *International Conference on Machine Learning*, pages 7829–7838, 2021.

[30] Emmanuel Rio. Covariance inequalities for strongly mixing processes. *Annales de l'IHP Probabilites et statistiques*, 29(4):587–597, 1993.

Proof of Lifted Generator for OU Processes

This appendix provides the complete proof of Proposition 4.1 from Chapter 4, establishing the explicit form of the lifted generator $\mathcal{G}^{(M)}(\psi)$ for Ornstein-Uhlenbeck processes.

Statement and Proof

Let X solve the d-dimensional Stratonovich OU SDE

$$dX_t = \theta(\mu - X_t)\,dt + \sigma \circ dW_t, \qquad Q := \sigma\sigma^\top.$$

and let $T^{(M)}(\mathbb{R}^d) = \bigoplus_{k=0}^{M}(\mathbb{R}^d)^{\otimes k}$ with basis words $w = e_{i_1} \otimes \cdots \otimes e_{i_m}$ and empty word $\mathbf{1}$. For the left-derivation ∂_{e_j} on words (removes a leading e_j, zero otherwise), the lifted generator $\mathcal{G}^{(M)}(\psi)$ acts by

$$\mathcal{G}^{(M)}(\psi)\,w = \sum_{j=1}^{d}(\theta\mu)_j\,w \otimes e_j + \sum_{i,j=1}^{d}(-\theta_{ij})\,e_i\,\partial_{e_j}w + \frac{1}{2}\sum_{i,j=1}^{d}Q_{ij}\,w \otimes e_i \otimes e_j. \quad (*)$$

Projection at level M. We interpret $\mathcal{G}^{(M)}$ as $\Pi^{(M)}\,\mathcal{G}\,\Pi^{(M)}$, i.e., components beyond degree M are discarded (projection onto $T^{(M)}(\mathbb{R}^d)$).

See the statement and notation in Section 4.2.

Goal. We prove that the expected truncated signature $E^{(M)}(t) := \mathbb{E}[S^{(M)}(X)_{0,t}] \in T^{(M)}(\mathbb{R}^d)$ satisfies

$$\frac{d}{dt}E^{(M)}(t) = \mathcal{G}^{(M)}(\psi)\,E^{(M)}(t), \qquad E^{(M)}(0) = \mathbf{1}. \qquad (.0)$$

with $\mathcal{G}^{(M)}(\psi)$ given by $(*)$. This is exactly the ODE announced in Section 4.2.

Preparations and identities. Write $S_t := S^{(M)}(X)_{0,t}$. We use:

1. **Chen's identity.** For $0 \leq s \leq t$, $S_{0,t} = S_{0,s} \otimes S_{s,t}$. Because we work in Stratonovich, this holds pathwise.

2. **Right/left duality.** For any $y \in T^{(M)}(\mathbb{R}^d)$, any word w, and any letter e_j,

   $$\langle w, y \otimes e_j \rangle = \langle \partial_{e_j} w, y \rangle, \qquad \langle w, y \otimes e_i \otimes e_j \rangle = \langle \partial_{e_j} \partial_{e_i} w, y \rangle,$$

   where $\langle \cdot, \cdot \rangle$ denotes the canonical pairing selecting the coordinate of w in y. This is the defining adjointness of right-concatenation to left-derivation on words.

3. **Iterated-integral differential.** If $w = e_j \otimes u$ (with u possibly empty), then, by the defining recursion of Stratonovich iterated integrals,

   $$d\langle w, S_t \rangle = \langle u, S_t \rangle \circ dX_t^j. \tag{1}$$

4. **Stratonovich–Itô conversion for S_t.** From $dX_t = b(X_t)\, dt + \sigma \circ dW_t$ with $b(x) := \theta(\mu - x)$ and the multiplicative relation $dS_t = S_t \otimes \circ dX_t$, converting to Itô yields

   $$dS_t = S_t \otimes b(X_t)\, dt + S_t \otimes \sigma\, dW_t + \tfrac{1}{2}\, S_t \otimes \Big(\sum_{i,j} Q_{ij}\, e_i \otimes e_j \Big)\, dt. \tag{2}$$

   In coordinates,

   $$S_t \otimes \sigma\, dW_t = \sum_{i=1}^{d} S_t \otimes e_i \Big(\sum_{k=1}^{d} \sigma_{ik}\, dW_t^k \Big),$$

   so that the quadratic variation contributes $\tfrac{1}{2} \sum_{i,j} Q_{ij}\, S_t \otimes e_i \otimes e_j\, dt$.

   Indeed, the extra drift term is the usual Stratonovich symmetrisation: $\circ dX^i \circ dX^j = d\langle X^i, X^j \rangle = Q_{ij}\, dt$.

Regularity. For OU processes, moments of all orders are finite; by dominated convergence we may interchange expectation and time differentiation for each coordinate $\langle w, S_{0,t} \rangle$.

Generator of the augmented Markov process. Consider the augmented process $Z_t = (X_t, S_t)$ on the state space $\mathbb{R}^d \times T^{(M)}(\mathbb{R}^d)$. Let $F(x, y)$ be a smooth test function on this space. The infinitesimal generator $\mathcal{L}$ of Z_t is given by Itô's formula. For a function of the specific form $F(x, y) = \langle w, y \rangle$ for a fixed word w, the relevant derivatives of F are:

- Derivatives with respect to x: $\nabla_x F = 0$ and $\nabla_x^2 F = 0$, since F does not depend on x.

- Gradient with respect to y: $\nabla_y F = w$, as F is a linear projection.

- Hessian with respect to y: $\nabla_y^2 F = 0$, for the same reason.

The Itô generator for $dZ_t = (dX_t, dS_t)$ acts on F as a sum of terms involving the drifts, covariations, and second derivatives. Given our choice of F, the only non-zero contributions come from the drift and quadratic variation of the S_t component, whose Itô differential is given in (2). This yields:

$$(\mathcal{L}F)(x, y) = \Big\langle w, \, y \otimes b(x) \Big\rangle + \frac{1}{2} \sum_{i,j} Q_{ij} \Big\langle w, \, y \otimes e_i \otimes e_j \Big\rangle. \qquad (3)$$

Using the duality property $\langle w, y \otimes u \rangle = \langle \partial_u w, y \rangle$, where ∂_u is the left-derivation corresponding to a word u, this is equivalently

$$(\mathcal{L}F)(x, y) = \sum_{j=1}^{d} b_j(x) \, \langle \partial_{e_j} w, \, y \rangle + \frac{1}{2} \sum_{i,j=1}^{d} Q_{ij} \, \langle \partial_{e_j} \partial_{e_i} w, \, y \rangle. \qquad (4)$$

Derivation of the Operator Form. Our goal is to find an operator $\mathcal{G}^{(M)}(\psi)$ acting only on the tensor algebra such that $\frac{d}{dt}\mathbb{E}[S_t] = \mathcal{G}^{(M)}(\psi)\mathbb{E}[S_t]$. This is equivalent to showing that for any word w, $\frac{d}{dt}\mathbb{E}\langle w, S_t \rangle = \langle w, \mathcal{G}^{(M)}(\psi)\mathbb{E}[S_t]\rangle$. By the properties of infinitesimal generators, we have

$$\frac{d}{dt}\mathbb{E}\langle w, S_t \rangle = \mathbb{E}\left[(\mathcal{L}F)(X_t, S_t)\right]$$

$$= \mathbb{E}\left[\sum_{j=1}^{d} b_j(X_t) \, \langle \partial_{e_j} w, \, S_t \rangle\right] + \mathbb{E}\left[\frac{1}{2}\sum_{i,j=1}^{d} Q_{ij} \, \langle \partial_{e_j} \partial_{e_i} w, \, S_t \rangle\right].$$

We now decompose the drift $b(x) = \theta\mu - \theta x$ and analyze each term's contribution by moving the action from the state (X_t, S_t) to an action on the test word w via the dual pairing.

1. Affine Drift Term ($\theta\mu$): The term involving the constant drift $\theta\mu$ becomes

$$\mathbb{E}\left[\sum_{j=1}^{d} (\theta\mu)_j \, \langle \partial_{e_j} w, \, S_t \rangle\right] = \sum_{j=1}^{d} (\theta\mu)_j \, \mathbb{E}\langle \partial_{e_j} w, \, S_t \rangle = \left\langle \sum_{j=1}^{d} (\theta\mu)_j \partial_{e_j} w, \, \mathbb{E}[S_t] \right\rangle$$

$$= \left\langle w, \, \mathbb{E}[S_t] \otimes \sum_{j} (\theta\mu)_j e_j \right\rangle = \langle w \otimes (\theta\mu), \, \mathbb{E}[S_t] \rangle.$$

This is the coordinate representation of the operator $w \mapsto w \otimes (\theta\mu)$, a degree-$+1$ creation operator.

2. Linear Drift Term ($-\theta x$): The theory of signature SDEs establishes that a linear vector field $v(x) = -\theta x$ on the base space $\mathbb{R}^d$ lifts to a derivation

54

operator on the tensor algebra. This operator acts on a word w by replacing its first letter. Formally,

$$
\mathbb{E}\left[\sum_{j=1}^{d}(-\theta X_t)_j \langle \partial_{e_j} w,\, S_t \rangle \right] = \mathbb{E}\left[\sum_{i,j=1}^{d} -\theta_{ji}(X_t)_i \langle \partial_{e_j} w,\, S_t \rangle \right]
$$

$$
= \left\langle \sum_{i,j} -\theta_{ji} e_i \otimes \partial_{e_j} w,\, \mathbb{E}[S_t] \right\rangle.
$$

This corresponds to the action of the degree-preserving derivation operator $w \mapsto \sum_{i,j}(-\theta_{ji}) e_i \partial_{e_j} w$.

3. Diffusion Term (Q): The diffusion term is constant and yields

$$
\mathbb{E}\left[\frac{1}{2}\sum_{i,j} Q_{ij} \langle \partial_{e_j} \partial_{e_i} w,\, S_t \rangle \right] = \frac{1}{2}\sum_{i,j} Q_{ij} \langle \partial_{e_j} \partial_{e_i} w,\, \mathbb{E}[S_t] \rangle
$$

$$
= \left\langle w,\, \mathbb{E}[S_t] \otimes \left(\frac{1}{2}\sum_{i,j} Q_{ij} e_i \otimes e_j \right) \right\rangle.
$$

This corresponds to the action of the degree-+2 creation operator $w \mapsto w \otimes (\frac{1}{2}\sum Q_{ij} e_i \otimes e_j)$.

Conclusion. Collecting the three operator actions on the test word w and using linearity of the inner product, we find

$$
\frac{d}{dt}\mathbb{E}\langle w, S_t \rangle = \left\langle \mathcal{G}^{(M)}(\psi)\, w,\, \mathbb{E}[S_t] \right\rangle,
$$

with $\mathcal{G}^{(M)}(\psi)$ given by $(*)$. Since the linear span of evaluation functionals $y \mapsto \langle w, y \rangle$ separates points in $T^{(M)}(\mathbb{R}^d)$, this is equivalent to the vector ODE $\frac{d}{dt}E^{(M)}(t) = \mathcal{G}^{(M)}(\psi)\, E^{(M)}(t)$, $E^{(M)}(0) = \mathbf{1}$. and therefore $E^{(M)}(t) = \exp\big(t\,\mathcal{G}^{(M)}(\psi)\big)\,\mathbf{1}$, as stated in Section 4.2.

Block triangularity. By construction: the $-\theta$ block preserves tensor degree (derivation), the $\theta\mu$ block raises degree by 1, and the Q block raises degree by 2, so $\mathcal{G}^{(M)}(\psi)$ is block lower-triangular in the grading by degree, exactly as remarked after the proposition, with the projection convention at degree M. □

Analytical Calibration Methods

This appendix provides the mathematical details of the analytical calibration methods used as baselines in Chapter 6. These methods provide closed-form parameter estimates without iterative optimization, serving as classical benchmarks for comparison with signature-based approaches.

Analytical Maximum Likelihood Estimation

For the Ornstein-Uhlenbeck process

$$dX_t = \theta(\mu - X_t)dt + \sigma dW_t, \tag{.0}$$

the transition density between observations is known exactly:

$$X_{t+\Delta t} \mid X_t \sim \mathcal{N}\left(\mu + (X_t - \mu)e^{-\theta\Delta t}, \frac{\sigma^2}{2\theta}(1 - e^{-2\theta\Delta t})\right). \tag{.0}$$

Parameter Estimation

Given a discrete path $\{X_{t_0}, X_{t_1}, \ldots, X_{t_N}\}$ with uniform time step Δt, the maximum likelihood estimators are derived from the log-likelihood function:

$$\ell(\theta, \mu, \sigma) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log\left(\frac{\sigma^2}{2\theta}(1 - e^{-2\theta\Delta t})\right)$$
$$- \sum_{i=1}^{N} \frac{(X_{t_i} - \mu - (X_{t_{i-1}} - \mu)e^{-\theta\Delta t})^2}{2\sigma^2(1 - e^{-2\theta\Delta t})/(2\theta)}.$$

The analytical MLE procedure:

1. **Mean estimation**: The long-run mean is estimated as

$$\hat{\mu} = \frac{1}{N+1}\sum_{i=0}^{N} X_{t_i}.$$

2. **Mean reversion estimation**: Using the lag-1 autocorrelation structure,

$$\hat{\theta} = -\frac{1}{\Delta t} \log \left(\frac{\mathrm{Cov}(X_{t+\Delta t}, X_t)}{\mathrm{Var}(X_t)} \right),$$

where the covariance and variance are computed from the centered observations.

3. **Volatility estimation**: From the stationary variance relationship,

$$\hat{\sigma}^2 = 2\hat{\theta} \cdot \mathrm{Var}(X_t).$$

K-Block Batching

In the experimental framework, the analytical MLE is applied with K-block batching:

1. The full path is divided into K non-overlapping blocks of equal length

2. Analytical MLE is performed independently on each block, yielding $\{\hat{\theta}_k, \hat{\mu}_k, \hat{\sigma}_k\}_{k=1}^{K}$

3. Final estimates are obtained by averaging: $\hat{\theta} = \frac{1}{K} \sum_{k=1}^{K} \hat{\theta}_k$

The optimal K^* is selected as:

$$K^* = \arg \min_{K \in \{1,2,4,8,16,32,64,128,256\}} \mathrm{MSE}(K),$$

where MSE is evaluated against the true parameters in simulation studies.

Method of Moments Estimation

The Method of Moments approach matches empirical moments to their theoretical counterparts under the OU model.

Theoretical Moments

For the stationary OU process, the theoretical moments are:

$$\mathbb{E}[X_t] = \mu$$

$$\mathrm{Var}(X_t) = \frac{\sigma^2}{2\theta}$$

$$\mathrm{Cov}(X_t, X_{t+\tau}) = \frac{\sigma^2}{2\theta} e^{-\theta\tau}$$

Moment-Based Estimators

The Method of Moments estimators are:

1. **Mean**: $\hat{\mu}_{\mathrm{MoM}} = \bar{X} = \frac{1}{N+1} \sum_{i=0}^{N} X_{t_i}$

2. **Variance**: $\hat{V} = \frac{1}{N} \sum_{i=0}^{N} (X_{t_i} - \bar{X})^2$

3. **Lag-1 autocovariance**: $\hat{C}_1 = \frac{1}{N-1} \sum_{i=0}^{N-1} (X_{t_i} - \bar{X})(X_{t_{i+1}} - \bar{X})$

4. **Mean reversion from autocorrelation**:

$$\hat{\theta}_{\mathrm{MoM}} = -\frac{1}{\Delta t} \log\left(\frac{\hat{C}_1}{\hat{V}}\right)$$

5. **Volatility from stationary variance**:

$$\hat{\sigma}_{\mathrm{MoM}} = \sqrt{2\hat{\theta}_{\mathrm{MoM}} \cdot \hat{V}}$$

Robustness Considerations

The Method of Moments estimator is particularly robust in high-volatility regimes where the autocorrelation structure remains identifiable even when likelihood-based methods struggle with numerical optimization. The method applies the same K-block batching strategy as the analytical MLE for consistency.

Initialization Strategy for Iterative Methods

Both Batched MLE and signature-based methods use smart initialization derived from the analytical methods:

1. **Phase 0**: Analytical OU is computed for $K \in \{1, 2, 4, \ldots, 256\}$

2. **K* Selection**: $K^* = \arg\min_K \mathrm{MSE}_{\mathrm{Analytical}}(K)$

3. **Smart Initialization**: Parameters $\{\hat{\theta}_{K^*}, \hat{\mu}_{K^*}, \hat{\sigma}_{K^*}\}$ from the best analytical result

4. **Batched MLE**: Uses K^* as its block count (algorithmic choice)

5. **Signature Methods**: Test $K \in \{16, 64, 256, 1024, 4096\}$ as hyperparameters

This ensures all iterative methods start from the same high-quality initialization point, enabling fair comparison of their optimization capabilities.

Hyperparameter Optimization Results

This appendix presents detailed results from the hyperparameter optimization study described in Chapter 6. The analysis comprises 200 experiments examining five learning rate configurations across four parameter regimes with 10 Monte Carlo replications each, conducted to identify optimal hyperparameters for signature-based calibration methods.

Learning Rate Configuration Specifications

Table 1 presents the five learning rate schedules evaluated in the hyperparameter optimization study.

Table 1: Learning Rate Configuration Specifications

Learning Rate	Start	End
$0.05 \rightarrow 0.05$	0.05	0.05
$0.05 \rightarrow 0.01$	0.05	0.01
$0.01 \rightarrow 0.01$	0.01	0.01
$0.01 \rightarrow 0.005$	0.01	0.005
$0.005 \rightarrow 0.005$	0.005	0.005

Cross-Configuration Performance Analysis

Table 2 presents the aggregated results across all parameter regimes, showing the average performance of each learning rate configuration.

Regime-Specific Optimal K Selection

The hyperparameter discovery phase systematically evaluated five learning rate configurations across four parameter regimes to identify optimal settings

Table 2: Cross-Configuration Performance Summary (Aggregated Across All Regimes)

LR	Method	Conv%	MSE	Std	K
	Batched MLE	87.5%	0.447	0.299	–
0.05 → 0.05	Expected Signature	97.5%	0.343	0.265	256
	Rescaled Signature	100.0%	0.343	0.271	256
	Batched MLE	82.5%	0.396	0.262	–
0.05 → 0.01	Expected Signature	90.0%	0.336	0.278	256
	Rescaled Signature	92.5%	0.337	0.279	256
	Batched MLE	72.5%	0.410	0.272	–
0.01 → 0.01	Expected Signature	90.0%	0.380	0.297	256
	Rescaled Signature	92.5%	0.380	0.291	256
	Batched MLE	67.5%	0.415	0.231	–
0.01 → 0.005	Expected Signature	97.5%	0.452	0.242	1024
	Rescaled Signature	97.5%	0.450	0.242	1024
	Batched MLE	60.0%	0.419	0.282	–
0.005 → 0.005	Expected Signature	85.0%	0.440	0.407	256
	Rescaled Signature	85.0%	0.441	0.409	256

for each regime-method combination.

Optimal Configurations by Regime

Table 3: Optimal Configuration Discovery by Regime

Regime	Method	Optimal LR Config	Optimal K	MSE
Slow Rev Low Vol	Batched MLE	[0.05→0.05]	16*	0.048
	Expected Sig	[0.005→0.005]	256	0.033
	Rescaled Sig	[0.005→0.005]	256	0.033
Fast Rev Low Vol	Batched MLE	[0.01→0.005]	1*	0.540
	Expected Sig	[0.01→0.005]	1024	0.544
	Rescaled Sig	[0.01→0.005]	1024	0.544
Slow Rev High Vol	Batched MLE	[0.05→0.01]	16*	0.396
	Expected Sig	[0.05→0.01]	256	0.336
	Rescaled Sig	[0.05→0.01]	256	0.337
Fast Rev High Vol	Batched MLE	[0.01→0.01]	1*	0.539
	Expected Sig	[0.01→0.01]	64	0.544
	Rescaled Sig	[0.01→0.01]	64	0.544

The results demonstrate that different parameter regimes require fundamentally different learning rate schedules and K values. Slow reversion regimes benefit from conservative learning rates but higher K values, while fast reversion regimes use more aggressive learning rates with moderate K values.

The hyperparameter discovery phase used 10 Monte Carlo runs per configuration-regime combination. The apparent disadvantages for signature methods in

certain regimes were resolved in Phase 2 validation with 100 Monte Carlo runs, which provided sufficient statistical power for hypothesis testing.

*Batched MLE uses algorithmic K^* selection rather than hyperparameter optimization; the modal K^* value chosen is presented.

MSE + Standard Deviation Scoring Impact

Robustness to Weight Parameter

The MSE + StdDev scoring function Score = MSE + $\lambda \cdot$ StdDev was tested with three weight values to verify robustness. All values $\lambda \in \{0.5, 1.0, 1.5\}$ produced identical optimal configurations across all four parameter regimes and both signature methods. This remarkable stability demonstrates that the scoring innovation captures a fundamental trade-off between accuracy and stability that is not sensitive to the precise weight value. We adopt $\lambda = 1.0$ as the balanced choice for all reported results.

Comparison with MSE-Only Optimization

To demonstrate the alignment advantage of MSE + StdDev scoring, Table 4 presents what the optimal learning rate and K configurations would have been under pure MSE optimization versus the MSE + StdDev approach actually employed. This comparison reveals how the variance penalty term improves methodological consensus between signature methods.

Table 4: MSE-Only vs MSE+StdDev Optimal Configuration Comparison

Regime	Method	MSE-Only Optimal		MSE+StdDev Optimal	
		Learning Rate	K	Learning Rate	K
Slow Rev Low Vol	Expected Sig	[0.005→0.005]	256	[0.005→0.005]	256
	Rescaled Sig	[0.005→0.005]	256	[0.005→0.005]	256
Fast Rev Low Vol	Expected Sig	[0.01→0.005]	4096	[0.01→0.005]	1024
	Rescaled Sig	[0.05→0.05]	16	[0.01→0.005]	1024
Slow Rev High Vol	Expected Sig	[0.05→0.01]	256	[0.05→0.01]	256
	Rescaled Sig	[0.05→0.01]	256	[0.05→0.01]	256
Fast Rev High Vol	Expected Sig	[0.05→0.05]	256	[0.01→0.01]	64
	Rescaled Sig	[0.05→0.05]	64	[0.01→0.01]	64

The comparison reveals three key advantages of MSE + StdDev scoring:

Enhanced Inter-Method Consensus: Under MSE-only optimization, Expected and Rescaled signature methods would select different configurations in fast reversion regimes (particularly Fast Rev, Low Vol where they choose completely different learning rates and K values). The MSE + StdDev approach

achieves 100% consensus between methods, with both signature methods selecting identical learning rate and K configurations across all four regimes.

Improved Regime Alignment: In slow reversion regimes, both scoring approaches yield identical results, confirming that the methods naturally converge when signal quality is high. However, in challenging fast reversion regimes, MSE + StdDev stabilizes the selection process by penalizing high-variance configurations that may appear optimal under limited sampling but lack robustness.

Practical Deployment Benefits: The alignment created by variance-aware scoring enables practitioners to apply either signature method with consistent hyperparameter recommendations across market conditions, eliminating the need for method-specific parameter tuning that pure MSE optimization would require.

Convergence Rate Analysis by Configuration

Table 5 presents convergence rates for different learning rate schedules across parameter regimes.

Table 5: Convergence Rates by Configuration and Regime

Learning Rate	Convergence Rate by Regime			
	Slow/Low	**Fast/Low**	**Slow/High**	**Fast/High**
Batched MLE				
$0.05 \rightarrow 0.05$	100%	100%	50%	100%
$0.05 \rightarrow 0.01$	90%	100%	40%	100%
$0.01 \rightarrow 0.01$	80%	90%	30%	90%
$0.01 \rightarrow 0.005$	70%	90%	20%	90%
$0.005 \rightarrow 0.005$	60%	80%	10%	90%
Expected Signature				
$0.05 \rightarrow 0.05$	100%	100%	90%	100%
$0.05 \rightarrow 0.01$	90%	100%	80%	90%
$0.01 \rightarrow 0.01$	90%	100%	70%	100%
$0.01 \rightarrow 0.005$	100%	100%	90%	100%
$0.005 \rightarrow 0.005$	80%	90%	80%	90%
Rescaled Signature				
$0.05 \rightarrow 0.05$	100%	100%	100%	100%
$0.05 \rightarrow 0.01$	90%	100%	90%	100%
$0.01 \rightarrow 0.01$	90%	100%	80%	100%
$0.01 \rightarrow 0.005$	100%	100%	90%	100%
$0.005 \rightarrow 0.005$	80%	90%	80%	90%

Batched MLE exhibits significant convergence issues in the Slow Reversion, High Volatility regime across all learning rate configurations, while signature methods maintain robust convergence rates above 80% in most scenarios.

Statutory Declaration / Affidavit

I hereby declare that the thesis with title

Convergence Theory for Expected Signature Estimation from Dependent Single Paths with Applications to Parameter Calibration

has been composed by myself autonomously and that no means other than those declared were used. In every single case, I have marked parts that were taken out of published or unpublished work, either verbatim or in a paraphrased manner, as such through a quotation.

This thesis has not been handed in or published before in the same or similar form.

Zurich, Wednesday 15$^{\text{th}}$ October, 2025