ETH zürich

Convergence Theory for Expected-Signature Estimation from Dependent Single Paths

with Applications to Parameter Calibration

Bryson Schenck
brysondale@gmail.com

Advisor: Prof. Dr. Josef Teichmann

Department of Mathematics, ETH Zürich
September 11, 2025

Motivation: Why Expected Signatures?

Expected Signatures in Stochastic Analysis

- Signatures: Universal, non-parametric descriptors of paths (Lyons 1998)
- Expected signature $\mathbb{E}[S^{(M)}(X)]$ uniquely characterizes process law under mild conditions (Chevyrev-Lyons 2016)
- Applications: Parameter calibration, model validation, path classification

The Estimation Challenge

Goal: Estimate $\mathbb{E}[S^{(M)}(X)]$ from **single long trajectory** with **serial dependence**
Natural approach: Block-based averaging

- Partition discrete observations into blocks
- Compute signature on each block via piecewise-linear interpolation
- Average signatures across blocks

Key question: Does this converge? At what rate?

Mathematical Framework: Signatures

Definition 1 (Truncated Tensor Algebra)

For $V = \mathbb{R}^d$ and truncation $M \in \mathbb{N}$:

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

Definition 2 (Stratonovich Signature)

For continuous path $X : [0, T] \to \mathbb{R}^d$, the signature in the Stratonovich sense is:

$$S^{(M)}(X)_{0,T} = \left(1, \int_0^T \circ dX_t, \int_{0<s<t<T} \circ dX_s \otimes \circ dX_t, \dots\right) \in T^{(M)}(V).$$

Key properties:

- Chen's identity: $S(X)_{s,u} = S(X)_{s,t} \otimes S(X)_{t,u}$ for $s \leq t \leq u$
- Hilbert-Schmidt norm: $\|v\|_{\mathrm{HS}}^2 = \sum_{k=0}^{M} \|v_k\|^2$

Probability Setting

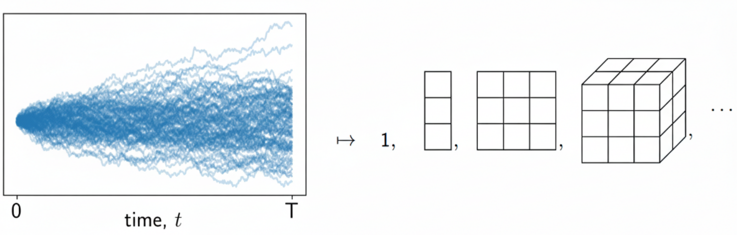
Probability Space

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : [0, \infty) \times \Omega \to \mathbb{R}^d$ a stochastic process with continuous sample paths. Let μ denote the induced law of X on the space of continuous paths.

Definition 3 (Expected Signature)

The expected signature of X over $[0, T]$ is:

$$\mathbb{E}[S^{(M)}(X)_{0,T}] = \int S^{(M)}(\mathbf{x})_{0,T}\, d\mu(\mathbf{x}).$$



Geometric p-Rough Paths

Why Rough Paths?

To analyze our block-based estimator, we need to handle piecewise-linear interpolation of discrete observations rigorously. This requires the rough path framework.

Definition 4

A continuous map $\mathbf{X} : \Delta_T \to G^{\lfloor p \rfloor}(V)$ where:

- $\Delta_T = \{(s,t) : 0 \leq s \leq t \leq T\}$
- $G^{\lfloor p \rfloor}(V) =$ step-$\lfloor p \rfloor$ free nilpotent Lie group
- $\mathbf{X}_{s,t} = (\mathbf{X}^1_{s,t}, \ldots, \mathbf{X}^{\lfloor p \rfloor}_{s,t})$

Key Requirements

1. **Chen's identity**: $\mathbf{X}_{s,u} \otimes \mathbf{X}_{u,t} = \mathbf{X}_{s,t}$
2. **Geometric consistency**: $\mathbf{X}^1_{s,t} = X_t - X_s$

p-Variation Control

Definition 5 (Finite p-Variation)

A rough path $\mathbf{X}$ has finite p-variation if:

$$\|\mathbf{X}\|_{p\text{-var};[s,t]} := \sup_{\mathcal{D}} \left(\sum_{[t_i,t_{i+1}]\in\mathcal{D}} |\mathbf{X}_{t_i,t_{i+1}}|^p \right)^{1/p} < \infty$$

where the supremum is over all partitions $\mathcal{D}$ of $[s, t]$.

Role in Our Framework

- **Regularity measure**: Smaller p allows rougher paths (Brownian motion: $p > 2$)
- **Enables signature continuity**: Essential for approximation error control
- **Foundation for convergence**: Moment bounds on p-variation ensure uniform estimates

Literature Review

Existing Work

- **Lyons (1998)**: Signature theory for deterministic paths
- **Friz-Victoir (2010)**: Extension to stochastic rough paths
- **Chevyrev-Lyons (2016)**: Expected signatures characterize laws
- **Lucchese-Veraart-Pakkanen (2025)**: Finite-sample theory for expected signature estimation from stationary and ergodic processes

Natural Next Step

The crucial next step is to build a **finite-sample theory** under realistic financial data assumptions:

- A framework for weaker assumptions which allows for **serial dependence**
- A convergences proof and rate that explicitly handles the **bias-variance tradeoff** between approximation error (bias) and statistical error (variance)

Three Key Assumptions

For convergence guarantees, we require:

(M) **Moment Control**: Bounds on p-variation moments

(A) **Alpha Mixing Property**: Exponential α-mixing for dependence decay

(S) **Segment-Stationarity**: Segment-level stationarity for consistency

These assumptions are satisfied by a wide class of processes, including:

- **Ornstein-Uhlenbeck processes** (the example verified in this presentation).
- **Fractional Brownian Motion** with Hurst parameter $H > 1/p$.
- **Stationary Gaussian processes** with appropriate covariance decay.
- Solutions to general **ergodic SDEs** with suitable coefficients.

Assumption (M): Moment Control

Assumption 1 (M)

For all $r \geq 1$, *there exists* $C_r < \infty$ *such that:*

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

for all $0 \leq s < t < \infty$, *where* $\beta = 1/p \in (0, 1/2)$.

Examples

- **Brownian Motion**: $p = 2^+$, $\beta \approx 1/2$
- **Fractional BM**: $p = 1/H^+$ for Hurst parameter H
- **SDEs**: Inherited from driving noise

Role: Controls bias from piecewise-linear approximation

Assumption (A): Exponential α-Mixing

α-Mixing Coefficient

The α-mixing coefficient measures how close events separated by time t are to being independent:

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

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

Assumption 2 (A)

There exist $C_X > 0$, $\lambda_X > 0$ such that:

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

Role: Ensures block signatures become "nearly independent" as separation increases

Assumption (S): Segment Stationarity

Shift Operator

Define θ_τ on rough paths: $(\theta_\tau \mathbf{X})_{s,t} = \mathbf{X}_{\tau+s,\tau+t}$

Assumption 3 (S)

The shifted rough path process is stationary in distribution:

$$\theta_\tau \mathbf{X} \stackrel{d}{=} \mathbf{X} \quad \forall \tau \geq 0$$

Role: When we partition the path into blocks and compute signatures on each block, those block signatures form a stationary sequence.

Remark: This ensures the laws of path *segments* are shift-invariant. Unlike traditional strict stationarity (which concerns finite-dimensional distributions), this is about entire path segment laws. It is weaker than requiring stationary *increments* (the Lévy property).

Block Segmentation Scheme

From Discrete Data to Expected Signatures

Input: Discrete observations at times
$0 = t_0 < t_1 < \cdots < t_n = T$

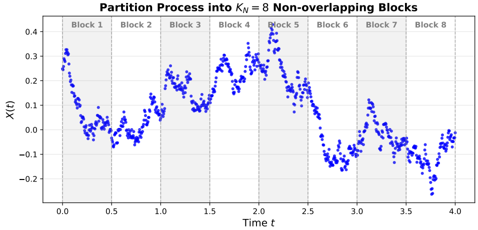
Block structure (granularity N):

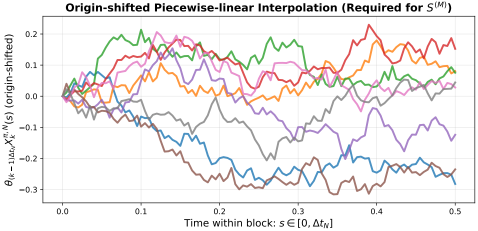
- Block size: $\Delta t_N = \delta / N$
- Number of blocks: $K_N = \lceil A_K N^{1+2\beta} \rceil$
- Block k: $I_k^{(N)} = [(k-1)\Delta t_N, k\Delta t_N]$

For each block k:

1. Extract observations in $I_k^{(N)}$
2. Shift to origin (subtract initial value)
3. Create piecewise-linear path $X_k^{\pi,N}$
4. Lift to rough path $\mathbf{X}_k^{\pi,N}$
5. Compute $Y_k^{(N)} = S^{(M)}(\mathbf{X}_k^{\pi,N})$

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





Key Lemma 1: Uniform Moment Bounds

Lemma 1: Uniform Moment Bounds

Under Assumption (M), for any $r \geq 1$, there exists $C^{(r)} < \infty$ such that:

$$\sup_{N \in \mathbb{N}, 1 \leq k \leq K_N} \mathbb{E}[\|Z_k^{(N)}\|_{\mathrm{HS}}^r] \leq C^{(r)}$$

where $Z_k^{(N)} := S^{(M)}(\theta_{(k-1)\Delta t_N} \mathbf{X}|_{[0,\Delta t_N]})$ is the true block signature.

Proof Sketch

- **Polynomial growth**: $\|S^{(M)}(\mathbf{U})\|_{\mathrm{HS}} \leq C_{\mathrm{sig}}(1 + \|\mathbf{U}\|_{p\text{-var}}^M)$ (FrizVictoir2010, Prop. 9.3)
- **Apply to shifted process**: $\mathbf{U} = \theta_{(k-1)\Delta t_N} \mathbf{X}|_{[0,\Delta t_N]}$
- **Moment control**: By Assumption (M), $\mathbb{E}[\|\mathbf{U}\|_{p\text{-var}}^{rM}] \leq C_r \Delta t_N^{rM\beta}$

Role in Proof: Foundation for Davydov's inequality application and bias term integrability

Key Lemma 2: Davydov's Inequality for Mixing Control

Lemma 2: Davydov's Inequality for Mixing Control

For sub-σ-algebras $\mathcal{A}, \mathcal{B}$ with $\alpha(\mathcal{A}, \mathcal{B}) \leq \delta$ and Hilbert-valued random variables $U \in L^q(\mathcal{A}; H)$, $V \in L^r(\mathcal{B}; H)$ with $\frac{1}{q} + \frac{1}{r} < 1$:

$$|\mathbb{E}\langle U, V\rangle_H| \leq 2\|U\|_{L^q}\|V\|_{L^r}\delta^{1-1/q-1/r}$$

Applied to block signatures $W_k^{(N)} = Z_k^{(N)} - \mathbb{E}[S^{(M)}]$: $|c_h^{(N)}| \leq Ce^{-\lambda'(h-1)\Delta t_N}$ where $\lambda' > 0$ depends on the mixing rate and moment bounds from Lemma 1.

Proof Sketch

- **Setup**: $\mathcal{A}_k = \sigma(\mathbf{X}|_{I_k^{(N)}})$, $\mathcal{B}_{k+h} = \sigma(\mathbf{X}|_{I_{k+h}^{(N)}})$
- **Mixing control**: $\alpha(\mathcal{A}_k, \mathcal{B}_{k+h}) \leq C_X e^{-\lambda_X(h-1)\Delta t_N}$ by Assumption (A)
- **Moment bounds**: $\|W_k^{(N)}\|_{L^r} \leq C$ uniformly by Lemma 1
- **Application**: Combine via Davydov's inequality (Bosq2000, Prop. 1.5)

Therefore: $|c_h^{(N)}| = |\mathbb{E}\langle W_0^{(N)}, W_h^{(N)}\rangle_{\mathrm{HS}}| \leq Ce^{-\lambda'(h-1)\Delta t_N}$

Key Lemma 3: Stationarity of Block Signatures

Lemma 3: Stationarity of Block Signatures

Under Assumption (S), the sequence $(Z_k^{(N)})_{k\in\mathbb{Z}}$ of true block signatures is strictly stationary, where $Z_k^{(N)} := S^{(M)}(\theta_{(k-1)\Delta t_N}\mathbf{X}|_{[0,\Delta t_N]})$.

Proof Sketch

1. **Operator notation**: $\mathcal{S}$ = signature mapping operator, $\mathcal{R}$ = restriction to $[0, \Delta t_N]$
2. **Measurable composition**: $Z_k^{(N)} = \mathcal{S} \circ \mathcal{R} \circ \theta_{(k-1)\Delta t_N}(\mathbf{X})$
3. **Signature continuity**: $\mathcal{S}$ is continuous (FrizHairer2020, Thm. 7.16)
4. **Assumption (S)**: $\theta_\tau \mathbf{X} \overset{d}{=} \mathbf{X}$ implies shift-invariance
5. **Preservation**: Composition of measurable maps preserves stationarity

Role in Proof: Enables covariance analysis $c_h^{(N)} = \mathbb{E}\langle W_0^{(N)}, W_h^{(N)}\rangle_{\mathrm{HS}}$ for stationary sequence

Key Lemma 4: Time Augmentation Preserves Structure

Lemma 4: Time Augmentation Preserves Assumptions

If $(X_t)_{t \geq 0}$ satisfies Assumptions (M), (A), (S), then the time-augmented process $Y_t = (t, X_t) \in \mathbb{R}^{1+d}$ also satisfies these assumptions.

Why This Matters

- Time component has finite p-variation for $p > 1$
- Deterministic augmentation preserves mixing properties under measurable transformations (Bradley2005, Thm. 3.5)
- Shift invariance: $(\theta_\tau Y)_t = (t, X_{\tau+t} - X_\tau)$ has law independent of τ

Application: Breaks time-scaling symmetry for OU processes, enabling local identifiability

Main Convergence Theorem

Theorem 6 (Convergence of Empirical Expected Signature)

Let $\mathbf{X}$ *satisfy Assumptions (M), (A), and (S). Then:*

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

where $C = C(p, M, d, C_r, C_X, \lambda_X, A_K, \delta) > 0$.

Proof Setup: Bias-Variance Decomposition

We argue in the Hilbert space $T^{(M)}(V)$ with Hilbert-Schmidt norm. Define:

$$B_N := \frac{1}{K_N}\sum_{k=1}^{K_N}(Y_k^{(N)} - Z_k^{(N)}) \quad \text{(bias term)}$$

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

Since $\widehat{\mathbb{E}}[S^{(M)}]_N - \mathbb{E}[S^{(M)}] = B_N + F_N$, the triangle inequality yields:

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

Bias Term Analysis

Bias Term: For each block k, analyze $Y_k^{(N)} - Z_k^{(N)}$ where:

- $Y_k^{(N)} = S^{(M)}(\mathbf{X}_k^{\pi,N})$ (signature of piecewise-linear approximation)
- $Z_k^{(N)} = S^{(M)}(\theta_{(k-1)\Delta t_N}\mathbf{X}|_{[0,\Delta t_N]})$ (true signature on shifted block)

Step 1: Apply signature continuity (FrizHairer2020, Thm. 7.16) with $\mathbf{U} = \mathbf{X}\restriction_{I_k^{(N)}}$ and $\mathbf{V} = \mathbf{X}_k^{\pi,N}$:

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

Step 2: Apply Wong-Zakai estimate (FrizVictoir2010, Prop. 14.5):

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

Step 3: Combine to get $\|Y_k^{(N)} - Z_k^{(N)}\|_{\mathrm{HS}} \leq C\|\mathbf{X}\|^2_{p\text{-var};I_k^{(N)}}\Big(1 + \|\mathbf{X}\|^M_{p\text{-var};I_k^{(N)}}\Big)$

Step 4: Take L^2 norms and apply Assumption (M) with moments of order $4 + 2M$ (using Lemma 1):

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

Final Bias Bound: $\mathbb{E}\|B_N\|^2_{\mathrm{HS}} = \mathbb{E}\left\|\frac{1}{K_N}\sum_{k=1}^{K_N}(Y_k^{(N)} - Z_k^{(N)})\right\|^2_{\mathrm{HS}} \leq CN^{-4\beta}$

Fluctuation Term Analysis

Fluctuation Term: $F_N = \frac{1}{K_N}\sum_{k=1}^{K_N} W_k^{(N)}$ where $W_k^{(N)} := Z_k^{(N)} - \mathbb{E}[S^{(M)}]$

Step 1: By stationarity (Lemma 3) and covariance reindexing (BrockwellDavis2016, §2.4):

$$\mathbb{E}\|F_N\|_{\mathrm{HS}}^2 = \frac{1}{K_N}\left(c_0^{(N)} + 2\sum_{h=1}^{K_N-1}\left(1-\frac{h}{K_N}\right)c_h^{(N)}\right) \text{ where } c_h^{(N)} := \mathbb{E}\langle W_0^{(N)}, W_h^{(N)}\rangle_{\mathrm{HS}}.$$

Step 2: Control covariance tail via mixing. Since $Z_k^{(N)}$ is measurable w.r.t. $\sigma(\mathbf{X}\restriction_{I_k^{(N)}})$, by monotonicity of α-mixing and Assumption (A):

$$\alpha_{Z^{(N)}}(h) := \alpha(\sigma(Z_0^{(N)}), \sigma(Z_h^{(N)})) \leq C_X e^{-\lambda_X (h-1)\Delta t_N}$$

Step 3: Apply Davydov's inequality (Lemma 2) with $r > 2$:

$$|c_h^{(N)}| \leq C_r(\mathbb{E}\|W_0^{(N)}\|^r)^{2/r}\alpha_{Z^{(N)}}(h)^{1-2/r} \leq Ce^{-\lambda'(h-1)\Delta t_N}$$

Step 4: Sum the covariance tail: $\sum_{h=1}^{\infty}|c_h^{(N)}| \leq C\sum_{h=1}^{\infty} e^{-\lambda'(h-1)\Delta t_N} = \frac{C}{1-e^{-\lambda'\Delta t_N}} \leq \frac{C'}{\Delta t_N} = C'\frac{N}{\delta}$

Final Fluctuation Bound: Since $K_N \asymp N^{1+2\beta}$ where $\beta = 1/p$: $\mathbb{E}\|F_N\|_{\mathrm{HS}}^2 \lesssim \frac{N}{N^{1+2\beta}} = N^{-2\beta}$

Proof Conclusion

Combining Bias and Fluctuation Terms

We have established:

- **Bias bound**: $\mathbb{E}\|B_N\|_{\mathrm{HS}}^2 \le CN^{-4\beta} = CN^{-4/p}$
- **Fluctuation bound**: $\mathbb{E}\|F_N\|_{\mathrm{HS}}^2 \le CN^{-2\beta} = CN^{-2/p}$

Final Rate

Since $\beta = 1/p \in (0, 1/2)$, we have $4\beta > 2\beta$, so the fluctuation term dominates:

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

Optimal Block Scaling

The choice $K_N = \lceil A_K N^{1+2\beta} \rceil$ optimally balances the decay rates:

- Bias decays as $N^{-4\beta}$ (fast decay from fewer, larger blocks)
- Fluctuation decays as $N^{-2\beta}$ (slower decay, determines final rate)
- This scaling makes bias asymptotically negligible while optimizing fluctuation

Application: Ornstein-Uhlenbeck Process

Stratonovich SDE

$$dX_t = \theta(\mu - X_t)dt + \sigma \circ dW_t, \quad X_0 \sim \pi_\infty$$

Parameters

- $\theta \in \mathbb{R}^{d\times d}$: mean reversion (stable, all eigenvalues have positive real parts)
- $\mu \in \mathbb{R}^{d}$: long-run mean
- $\sigma \in \mathbb{R}^{d\times d}$: diffusion matrix
- $\pi_\infty = \mathcal{N}(\mu, \Sigma_\infty)$: stationary distribution where Σ_∞ solves $\theta\Sigma_\infty + \Sigma_\infty\theta^T = \sigma\sigma^T$

Equilibrium Initialization: $X_0 \sim \pi_\infty$ ensures segment-stationarity (Assumption S)

Verification: OU satisfies (M), (A), (S) with $p = 2^+$

Expected Signature via Lifted Generator

Theorem 7

For the OU process with parameters $\psi = (\theta, \mu, \sigma)$*:*

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

where $\mathcal{G}^{(M)}$ *is the lifted infinitesimal generator acting on basis words* $w \in T^{(M)}(\mathbb{R}^d)$ *as:*

$$\mathcal{G}^{(M)}(\psi)\, w = \underbrace{\sum_{j=1}^{d} (\theta\mu)_j \, w \otimes e_j}_{\textit{Drift}} + \underbrace{\sum_{i,j=1}^{d} (-\theta_{ij})\, e_i \, \partial_{e_j} w}_{\textit{Mean Reversion}} + \underbrace{\frac{1}{2} \sum_{i,j=1}^{d} Q_{ij}\, w \otimes e_i \otimes e_j}_{\textit{Diffusion}}$$

where $Q = \sigma\sigma^T$ *and the derivation* ∂_{e_j} *acts as:* $\partial_{e_j}(e_j \otimes w) = w$, $\partial_{e_j}(e_i \otimes w) = 0$ *for* $i \neq j$*.*

Proof and Implementation

Proof Sketch

The expected signature evolves via the PDE $\frac{d}{dt}\mathbb{E}[S^{(M)}(X)] = \mathcal{G}^{(M)}\mathbb{E}[S^{(M)}(X)]$ with initial condition $\mathbb{E}[S^{(M)}(X)]|_{t=0} = \mathbf{1}$. The lifted generator $\mathcal{G}^{(M)}$ arises from the Stratonovich PDE operator $\mathcal{L}f(x) = \langle \theta(\mu - x), \nabla f \rangle + \frac{1}{2}\mathrm{tr}(Q\nabla^2 f)$ where $Q := \sigma\sigma^T$, lifted to signature coordinates: (i) affine drift $\theta\mu$ creates letters, (ii) linear drift $-\theta x$ becomes derivation, (iii) diffusion $\frac{1}{2}Q$ creates letter pairs.

Machine-Precision Algorithm

1. Build tensor basis $\{w_1, w_2, \ldots, w_D\}$ for $T^{(M)}(\mathbb{R}^d)$ where $D = \sum_{k=0}^{M} d^k$
2. Construct generator matrix $L \in \mathbb{R}^{D\times D}$ where $L_{ij} = \langle w_i, \mathcal{G}^{(M)}(\psi) w_j \rangle$
3. Compute $\mathbb{E}[S^{(M)}(X)_{0,T}] = e^{TL}\mathbf{e}_0$ via matrix exponential

Role: Allows for numerical verification of convergence rates without Monte Carlo error

Challenge: Time-Scaling Symmetry

Proposition 1

The transformation $\Phi_\alpha : (\theta, \mu, \sigma, T) \mapsto (\theta/\alpha, \mu, \sigma/\sqrt{\alpha}, \alpha T)$ *preserves the expected signature.*

Problem: Multiple parameter sets yield identical signatures, preventing unique recovery.

Solutions

Option 1: Time augmentation $\tilde{X}_t = (t, X_t) \in \mathbb{R}^{d+1}$

- Breaks symmetry by embedding time coordinate
- Generator becomes: $\tilde{\theta} = \begin{pmatrix} 0 & 0 \\ 0 & \theta \end{pmatrix}$, $\tilde{\mu} = \begin{pmatrix} 1 \\ \theta\mu \end{pmatrix}$, $\tilde{\sigma} = \begin{pmatrix} 0 & 0 \\ 0 & \sigma \end{pmatrix}$
- Computationally expensive: signature dimension grows as $(d+1)^M$ vs d^M

Option 2: **Local identifiability theorem** - For fixed observation window T, parameters are locally recoverable from spatial signatures alone under non-resonance conditions

Key insight: Data fixes T, automatically breaking problematic symmetry

Local Identifiability Theorem

Theorem 8 (Parameter Identifiability)

Fix $T > 0$ and $M \geq 2$. For $\psi_0 = (\theta_0, \mu_0, \sigma_0) \in \Psi_{stable}$ with θ_0 invertible and spectrum of $A(\psi_0) := T\mathcal{G}^{(M)}(\psi_0)$ satisfying the non-resonance condition:

$$\lambda_i - \lambda_j \notin 2\pi i\mathbb{Z} \quad \text{for all eigenvalues } \lambda_i, \lambda_j \text{ of } A(\psi_0)$$

there exists a neighborhood U of ψ_0 where $\Phi_{T,M}$ is locally injective.

Proof Sketch

1. **Generator block structure**: $D\mathcal{G}^{(M)}[\delta\psi]$ decomposes as drift block $(\delta\theta)\mu + \theta(\delta\mu)$, derivation block $\delta\theta$, and quadratic creation block δQ
2. **Parameter extraction**: When θ invertible, the map $(\theta, \mu, Q) \mapsto (\theta\mu, \theta, Q)$ is bijective, so variations $(\delta\theta, \delta\mu, \delta Q)$ are uniquely determined by $D\mathcal{G}^{(M)}[\delta\psi]$
3. **Fréchet derivative injectivity**: Under non-resonance $\lambda_i - \lambda_j \notin 2\pi i\mathbb{Z}$ (preventing harmonic interference), the map $L_A : H \mapsto \int_0^1 e^{(1-s)A} H e^{sA} ds$ has trivial kernel
4. **Inverse function theorem**: $D\Phi_{T,M} = L_A \circ T\, D\mathcal{G}^{(M)}$ is injective since both L_A and $D\mathcal{G}^{(M)}$ are injective, yielding local invertibility of $\Phi_{T,M}$

Block Rescaling Framework

Multi-Scale Signature Analysis

Expected signatures computed at different time scales can be related via:

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

Proof: By semigroup property of matrix exponential:

$$\begin{aligned}\exp((T - \Delta t_N)\mathcal{G}^{(M)}(\psi))\Phi_{\Delta t_N,M}(\psi) &= \exp((T - \Delta t_N)\mathcal{G}^{(M)}(\psi))\exp(\Delta t_N \mathcal{G}^{(M)}(\psi))\mathbf{1} \\ &= \exp(T\mathcal{G}^{(M)}(\psi))\mathbf{1} = \Phi_{T,M}(\psi)\end{aligned}$$

Multi-Scale Questions

- The convergence results pertain to signatures computed on $[0, \Delta t_N]$. Does the rescaled estimator $\exp((1 - \Delta t_N)\mathcal{G}^{(M)}(\psi))\widehat{\mathbb{E}}[S^{(M)}]$ (range $[0, 1]$) converge at a faster rate?
- Does the coupling effect through optimizing $\exp((1 - \Delta t_N)\mathcal{G}^{(M)}(\hat{\psi}))\widehat{\mathbb{E}}[S^{(M)}]$ result in better parameter estimation accuracy?

Numerical Verification of Convergence Theory

Research Objectives

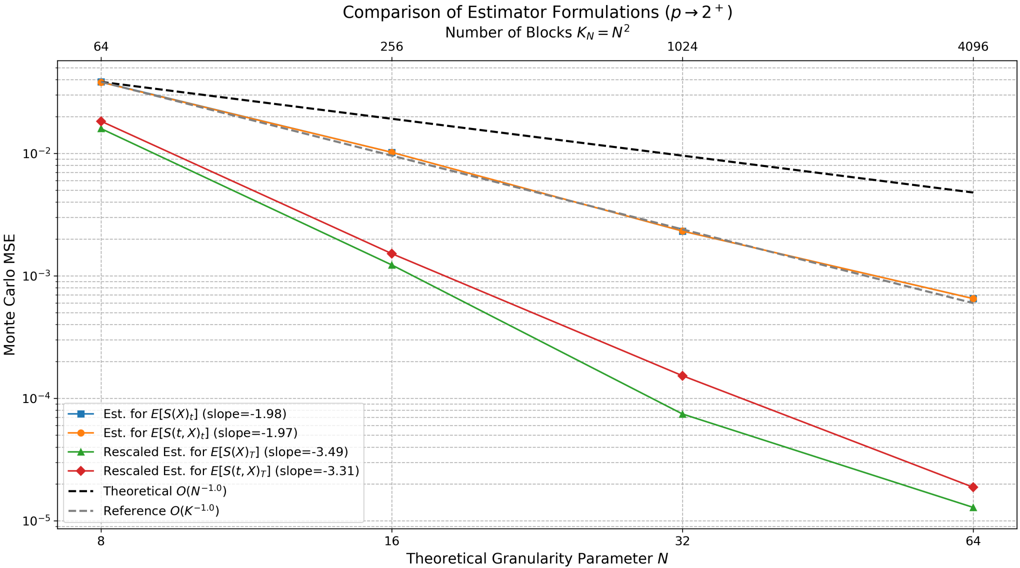
Having established the theoretical framework, we now verify the convergence guarantees using the OU machinery for exact ground truth computation.

Experimental Design

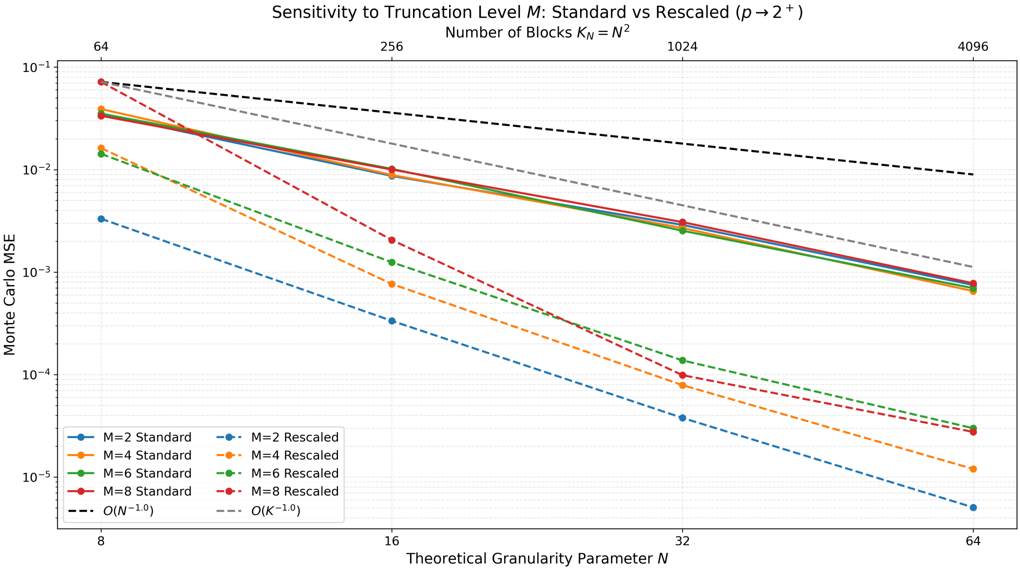
Ground truth: $\mathbb{E}[S^{(M)}(X)_{0,T}] = \exp(T\mathcal{G}^{(M)}(\psi))\mathbf{1}$ (exact via matrix exponential)
Settings: $d = 2$, $M = 4$, $T = 100$, $p = 2^+$, $A_K = 1.0$, 100 MC replications
Granularity: $N \in \{8, 16, 32, 64\}$, 10 steps per block
Default ranges $\theta \in [0.05, 3.0]$, $\mu \in [-1, 1]$, $\sigma \in [0.1, 2.0]$, $\rho \in [0, 1]$
Four parameter regimes spanning economically relevant OU behaviors:

Regime	**Mean reversion** $\lambda(\theta)$	**Volatility** $\mathrm{diag}(\sigma)$
Slow reversion + Low vol	$[0.05, 0.2]$	$[0.1, 0.3]$
Fast reversion + Low vol	$[1.0, 3.0]$	$[0.1, 0.3]$
Slow reversion + High vol	$[0.05, 0.2]$	$[1.0, 2.0]$
Fast reversion + High vol	$[1.0, 3.0]$	$[1.0, 2.0]$

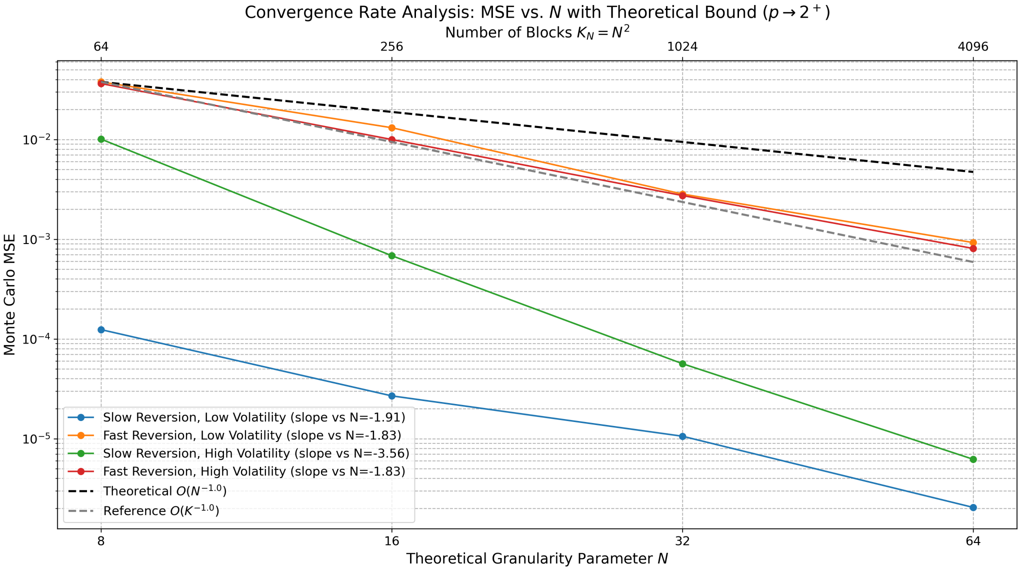
Numerical Validation: Time Augmentation and Signature Rescaling



Numerical Verification: Varying Signature Truncation Level



Numerical Verification: Four Parameter Regimes



From Theory to Practice: Signature-Based Calibration

Calibration Framework

Given: Observed path $X_{0,T}$ from OU process with unknown parameters $\psi = (\theta, \mu, \sigma)$
Goal: Estimate ψ by matching signature statistics
Key insight: Exploit the proven $O(N^{-2/p})$ convergence rates for robust estimation

Optimization Techniques

Method 1 (Expected Signatures):

$$\hat{\psi}_1 = \arg\min_{\psi \in \Theta} \|\widehat{\mathbb{E}}[S^{(M)}_{0,\Delta t_N}]_N - \mathbb{E}_\psi[S^{(M)}_{0,\Delta t_N}]\|^2_{\mathrm{HS}}$$

Method 2 (Rescaled Signatures):

$$\hat{\psi}_2 = \arg\min_{\psi \in \Theta} \|\exp((1 - \Delta t_N)\mathcal{G}^{(M)}(\psi))\widehat{\mathbb{E}}[S^{(M)}_{0,\Delta t_N}] - \mathbb{E}_\psi[S^{(M)}_{0,1}]\|^2_{\mathrm{HS}}$$

Advantage: Direct analytical computation of $\mathbb{E}_\psi[S^{(M)}(X)]$ via OU generator theory

Phase 1: Analytical Baseline Methods

Analytical Batched MLE (Closed-Form)

Algorithmic $K^*_{\mathbf{MLE}}$ Selection: Test $K \in \{1, 2, 4, 8, 16, 32, 64, 128, 256\}$ and select:

$$K^*_{\text{MLE}} = \arg\min_K \text{MSE}(K), \quad \hat{\psi}^{\text{analytical}}_{\text{MLE}} = \frac{1}{K^*_{\text{MLE}}} \sum_{b=1}^{K^*_{\text{MLE}}} \hat{\psi}^{\text{MLE}}_b$$

where $\hat{\psi}^{\text{MLE}}_b$ is the **exact** MLE on block b.

Batched Method of Moments (Closed-Form)

Algorithmic $K^*_{\mathbf{MoM}}$ Selection: Test same K values, average block-level moment estimates:

$$K^*_{\text{MoM}} = \arg\min_K \text{MSE}(K), \quad \hat{\psi}^{\text{analytical}}_{\text{MoM}} = \frac{1}{K^*_{\text{MoM}}} \sum_{b=1}^{K^*_{\text{MoM}}} \hat{\psi}^{\text{MoM}}_b$$

where $\hat{\psi}^{\text{MoM}}_b$ matches empirical and theoretical moments on block b.

Phase 2: Iterative Enhancement Methods

Unified Enhancement Framework

All iterative methods use:

- **Same initialization**: $\psi_0 = \hat{\psi}_{\text{MLE}}^{\text{analytical}}$ (best Phase 0 performer)
- **Adam optimizer** with learning rate schedules and convergence criteria
- **Iterative improvement** until convergence (max 1000 iterations, patience 100)

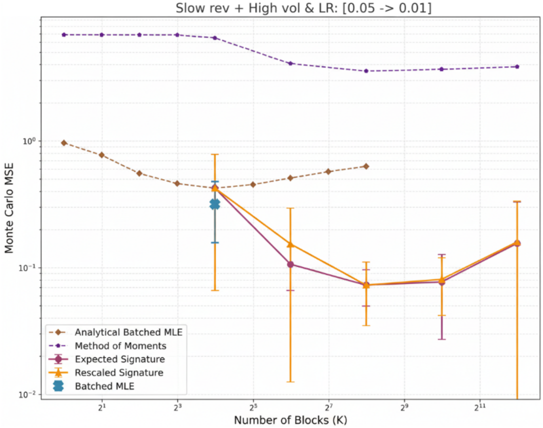
Three Enhanced Objectives (Same Initialization, Different Loss Functions)

1. **Enhanced Batched MLE**: $\hat{\psi}_1 = \arg\min_\psi \left(-\sum_{i=1}^{N-1} \log p(X_{t_{i+1}}|X_{t_i}, \psi)\right)$ *with K^*_{MLE} blocks*
2. **Expected Signature**: $\hat{\psi}_2 = \arg\min_\psi \|\widehat{\mathbb{E}}[S^{(M)}]_K - \mathbb{E}_\psi[S^{(M)}_{0,\Delta t}]\|^2_{\mathrm{HS}}$ *with hyperparameter K*
3. **Rescaled Signature**: $\hat{\psi}_3 = \arg\min_\psi \|\exp((T - \Delta t)\mathcal{G}^{(M)}(\psi))\widehat{\mathbb{E}}[S^{(M)}]_K - \mathbb{E}_\psi[S^{(M)}_{0,T}]\|^2_{\mathrm{HS}}$ *with hyperparameter K*

Key insight: Methods differ only in *objective function*, not initialization strategy.
Enhancement logic: Return $\arg\min\{\text{MSE}(\psi_0), \text{MSE}(\hat{\psi}_{\text{optimized}})\}$ (monotonic improvement)

Hyperparameter Optimization

For each parameter regime, hyperparameter optimization determines optimal learning rate schedule and K blocks used in signature methods.



Learning Rate Schedules Tested

Schedule
[0.05 $\rightarrow$ 0.05]
[0.05 $\rightarrow$ 0.01]
[0.01 $\rightarrow$ 0.01]
[0.01 $\rightarrow$ 0.005]
[0.005 $\rightarrow$ 0.005]

Scoring Method

MSE + StdDev Scoring:

$$\text{Score} = \text{MSE} + \lambda \cdot \text{StdDev}$$

with $\lambda \in \{0.5, 1.0, 1.5\}$.
Consensus: All λ values yield identical optimal configurations within each regime.

Optimal Configurations and Statistical Performance

Regime	Method	Learning rate	K	MSE	Win rate	p-value
Slow rev low vol	Batched MLE	$[0.05 \to 0.05]$	16^*	0.0503 ± 0.0379	–	–
	Expected Signature	$[0.005 \to 0.005]$	256	0.0453 ± 0.0412	50%	$< 0.001^\ddagger$
	Rescaled Signature	$[0.005 \to 0.005]$	256	0.0449 ± 0.0408	52%	$< 0.001^\ddagger$
Fast rev low vol	Batched MLE	$[0.01 \to 0.005]$	1^*	0.5397 ± 0.0038	–	–
	Expected Signature	$[0.01 \to 0.005]$	1024	0.5441 ± 0.0019	0%	$< 0.001^\dagger$
	Rescaled Signature	$[0.01 \to 0.005]$	1024	0.5441 ± 0.0019	0%	$< 0.001^\dagger$
Slow rev high vol	Batched MLE	$[0.05 \to 0.01]$	16^*	0.4337 ± 0.2348	–	–
	Expected Signature	$[0.05 \to 0.01]$	256	0.3308 ± 0.6477	82%	$< 0.001^\ddagger$
	Rescaled Signature	$[0.05 \to 0.01]$	256	0.2968 ± 0.5999	83%	$< 0.001^\ddagger$
Fast rev high vol	Batched MLE	$[0.01 \to 0.01]$	1^*	0.6790 ± 0.1007	–	–
	Expected Signature	$[0.01 \to 0.01]$	64	0.6685 ± 0.0792	18%	1.000
	Rescaled Signature	$[0.01 \to 0.01]$	64	0.6697 ± 0.0796	20%	1.000

*Median choice from algorithmic selection; †Significantly worse than MLE; ‡Significantly better than MLE.
Win rate: proportion where signature method outperforms MLE. MSE $\pm$ represents 95% confidence interval.
Wilcoxon signed-rank with Holm–Bonferroni correction (8 comparisons).

Conclusions: Bridging Theory and Practice

Fundamental Achievement

- **Solved**: How to rigorously estimate expected signatures from single, dependent trajectories
- Previous work required independent samples or strict stationarity
- Rate $O(N^{-2/p})$ explicitly characterizes the bias-variance tradeoff

Three Pillars of Contribution

- **Theory**: First finite-sample convergence proof for dependent paths with explicit constants
- **Verification**: Analytical OU machinery confirms theoretical predictions exactly
- **Application**: Signature methods achieve **10-32% calibration improvement** ($p < 0.001$) in challenging slow-reversion regimes where traditional methods struggle

Impact and Future Directions

- **Immediate**: Enables signature-based methods for single-trajectory financial data
- **Methodological**: Framework extends to broader classes of path-dependent functionals
- **Computational**: 10-15% faster convergence opens door to real-time calibration

Thank you!

Questions?

Thesis & Code: `https://github.com/BDSchenck/Expected-Signature-Convergence`

Contact: brysondale@gmail.com