Signature Methods in Variance Swap Pricing

Zuyuan Han (zyhan@umich.edu)

December 16, 2023

Supervisor:
Prof. Edwards Ionides (ionides@umich.edu)

Co-Supervisor:
Prof. Bingyan Han (byhan@umich.edu)

Abstract

This thesis addresses the pricing challenge of variance swap through the application of signature methods, a novel addition to quantitative finance area in recent years. Under certain assumptions, we build a signature model on volatility and we conduct a special case of the general underlying process, where the model is driven by a one-dimensional underlying Brownian motion. To compute the expected signatures, we refer to a matrix-form operator which extracts signature values at the initial level. By modelling volatility, we construct variance swap strike price as a function of signature values. Finally we derive a parameter-dense model for the final calibration, a task characterized as a high-dimensional regression. We discuss the calibration results on two cases. In the context of the 'static' case, our approach yields precise calibration results. However, in the 'sampled' case, our calibration accuracy is not as high. This discrepancy can be attributed to the current limitations of our method, which include a lack of incorporation of additional parameters and the absence of more complex processes.

Acknowledgement. *The work on this project was supported by the Summer 2023 REU (Research Experience for Undergraduates) program at the Department of Mathematics at the University of Michigan. The REU report could be found here Han et al. (2023)[10]. The main theorem (Theorem 4.1) in this thesis is jointly finished by Prof. Qi Feng, Prof. Bingyan Han, Haisu Ding and Zuyuan Han during the REU program.*[1] *I would like to thank them for*

[1] Theoretical part collaborators: Prof. Qi Feng (qfeng2@fsu.edu), Prof. Bingyan Han(byhan@umich.edu), Haisu Ding (hsding@umich.edu).

their contribution to this work. I would also like to thank Jesse Wheeler[2] *for his help in my code implementation.*

Contents

[2]Jesse Wheeler, jeswheel@umich.edu

1 Introduction

1.1 Background Information For Variance Swap

A variance swap is a type of financial derivative enabling investors to trade or hedge against the volatility of an underlying asset. It is a contract where two parties agree to exchange the realized variance of the underlying asset for a predetermined fixed payment. Typically, the underlying asset of a variance swap is an equity index, such as the S&P 500, or a specific stock. Variance swaps provide a unique benefit: they offer direct exposure to an asset's volatility, in contrast to call and put options, which may involve directional risk. The gains and losses associated with a variance swap are determined by the difference between the realized and implied volatility, as illustrated in Figure 1.

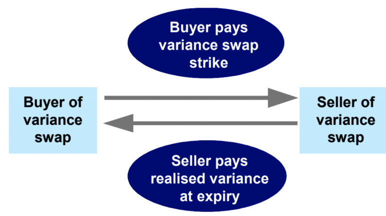


Figure 1: Variance Swap Cash Flows [*Allen et al.,* 2006[3]]

Variance swaps are commonly used by investors and traders to hedge against, or speculate on, changes in volatility. By entering into a variance swap, market participants can effectively isolate and trade the volatility component of an asset's price movement, independently of the asset's direction. This enables investors to manage their exposure to market volatility separately from their exposure to the asset's returns. Variance swaps offer a flexible and efficient means to gain exposure to volatility and are often utilized by institutional investors, hedge funds, and other sophisticated market participants. Such exposure to volatility can be particularly advantageous for investors during periods of significant stock market turbulence, as seen during the 2008 financial crisis and the COVID-19 pandemic.

1.2 Literature Review

Recent developments in machine learning and big-data analysis for finance applications have given rise to novel machine learning based approaches to solve challenging asset pricing problems. For example, neural networks were used by *Ferguson & Green (2019)*[6] to approximate pricing functions of derivatives. Similarly, *Buehler et al. (2019)* focuses on optimal hedging also with neural networks. Although the learning methods have strong ability to adapt to complex patterns in data, model interpretability can be an issue. In contrast, model-based methods can capture volatility dynamics by using mathematical models such as the Heston, Ornstein-Uhlenbeck process, etc., which could take more detailed features of rough paths into consideration.

Cuchiero et al. (2022)[1] attempt to use mathematical objects, *signatures*, to assist the incorporation of stochastic process in pricing problems. They formulate the time extended signature as the linear regression basis of continuous path functionals, aiming at applying data-driven, parameter-dense and tractable signature-based model in achieving outstanding calibration results in their pricing challenge. This modeling approach is supported by previous theory on rough paths like *Lyons (1998)*[9]. We refer to asset price modeling process as a *signature method* or a *signature-based model.*

Furthermore, in their most recent work, *Cuchiero et al. (2023)*[2] achieves highly accurate results in joint calibration to S&P500(SPX) and Volatility Index(VIX) options with signature-based model. The section 4 "*Expected signature of polynomial diffusion processes*" in *Cuchiero et al. (2023)*[2] also provides a direct inspiration of how signature method could be incorporated into describing polynomial processes and deriving close forms of expectation terms. Our work of taking signature model into computation starts from the pricing formula of variance swap and a small portion of our formula have a similar form with VIX index presented as Theorem 5.1 in *Cuchiero et al. (2023)*[2]. Nevertheless, our model for Variance Swap would be much more complex than their model for VIX index and could involve calibration work on higher-order term.

1.3 Model Choice

Signature-based models can be used to encode path information of underlying processes under various modeling structures. For example, a signature-based model can incorporate different mathematical processes such as Brownian Motion (in our project) or the Ornstein-Uhlenbeck process (in *Cuchiero et al. (2023)*[2]'s case) to approximate the volatility process.

Furthermore, an additional advantage of signature-based models is that they can be implemented using both discrete and continuous time models for the underlying processes. Variance swap pricing are usually based on discrete monitoring (i.e., a discrete time signal process), while traditional pricing methods tend to have assumptions on continuous monitoring (such as the Heston model). The signature method could be incorporated with discrete price points, potentially reducing the gap between the theoretical model and the real-world contract term.

In this project, we use signature model in modeling stochastic volatility. This is compared to traditional models for Volatility (e.g. Heston Model). The signature model is based on the theory of rough paths and focuses on the path properties of stochastic processes, thus we expect to have a better calibration results by capturing more path details. Inspired by *Cuchiero et al. (2023)*[2], we build a model for the volatility process σ_t from the price process $S = (S_t)_{t\geq 0}$ with dynamics given by

$$dS_t = rS_t dt \ + \ \sigma_t^S(\ell)S_t dB_t, \quad S_0 \in \mathbb{R}_+.$$

Here, r is the risk-free interest rate, and B_t is a one dimensional Brownian motion process. The process B_t is correlated with σ_t^S with correlated coefficient ρ. Then we model σ_t^S as,

$$\sigma_t^S(\ell) = \ell_\emptyset + \sum_{\ell_I} \ell_I \langle e_I, \widehat{\mathbb{Z}}_t \rangle,$$

where $\mathbb{Z}_t$ is the *signature* of process $\widehat{Z}_t = (t, W_t, B_t)$. Here, W_t is the underlying Brownian motion that drives the volatility, and is referred to the primary process. In this model, $\widehat{\mathbb{Z}}_t$ serves as a linear regression basis for the volatility process, $\ell = (\ell_\emptyset, \ell_I)$ are the parameters learned from the observed price data. Our goal is to calibrate on these parameters ℓ to derive accurate calibration results.

In comparing our study with the work of *Cuchiero et al. (2023)*[2], here are two key differences. First, *Cuchiero et al. (2023)* [2] use a complex multidimensional process X_t (see Definition3.1), while our study focuses on a simpler case using a one-dimensional Brownian motion (W_t) as the primary process. This approach simplifies our methodology to concentrate on specific aspects of the model.

Second, the methods we use for pricing VIX options and futures are more complex than those of *Cuchiero et al. (2023)*[2] While their pricing formula partly matches the variance swap pricing model, we need more complex components for variance swap. It requires higher-level computational methods. This highlights a significant contrast: *Cuchiero et al. (2023)*[2]'s work features a simpler pricing approach within a broader model, whereas our study involves more complex calculations within a more streamlined model of volatility.

2 Signature Model

The concepts and notions of signature and its application in rough path theory originally goes back to *Chen (1957)*[8] and *Lyons (1998)*[9]. We consider the *time extended signature* of an $\mathbb{R}^d$ path to serve as the linear regression basis for continuous path functionals. *Cuchiero et al. (2022)*[1] introduce the Universal Approximation Theorem, which describes the fact that the continuous path functionals on compact sets can be uniformly approximated by a linear function of the time extended signature. Specifically, the signature is an object associated with a path which captures many of the path's important analytic and geometric properties. The detailed theoretical properties and some numerical applications are discussed in *Chevyrev & Kormilitzin (2016)*[7].

2.1 Mathematical Foundations

In this section we present the necessary mathematical foundations that we use in constructing our model.

For $n, d \in \mathbb{N}$, the n-fold tensor product of $\mathbb{R}^d$ is given by

$$(\mathbb{R}^d)^{\otimes n} = \underbrace{(\mathbb{R}^d)\otimes(\mathbb{R}^d)\otimes ... \otimes (\mathbb{R}^d)}_{\text{n}},$$

where we construct an nd-dimensional space out of n vectors from d-dimensional space.

For $d \in \mathbb{N}$, the extended tensor algebra on $\mathbb{R}^d$ is given by

$$T(\mathbb{R}^d) = \{\mathbf{a} = (a_0, a_1, ..., a_n) : a_i \in (\mathbb{R}^d)^{\otimes i},\ i = 0, 1, 2, ..., n\}.$$

Suppose $\mathbf{a} = (a_i)_{i=0}^{\infty}$, $\mathbf{b} = (b_i)_{i=0}^{\infty} \in T((\mathbb{R}^d))$, define the sum $+$ and product $\otimes$ by

$$\mathbf{a}\ +\ \mathbf{b}\ := (a_i + b_i)_{i=0}^{\infty},$$

$$\mathbf{a}\ \otimes\ \mathbf{b}\ := (\sum_{k=0}^{i} a_k \otimes b_{i-k})_{i=0}^{\infty}.$$

Suppose a multi-index $I\ :=\ (i_1, i_2, ..., i_n)$, then we set $|I| := n$. Remark that we also set $I' := (i_1, ..., i_{n-1})$ and $I'' := (i_1, ..., i_{n-2})$ whenever needed.
We also have the notation $\{I : |I| = n\}\ :=\ \{1, 2, ..., d\}^n$.
Now we combine the multi-index with the tensor basis:

$$e_I\ =\ e_{i_1} \otimes e_{i_2} \otimes\ ...\ e_{i_n},$$

where $e_{i_1}, e_{i_2}, ..., e_{i_n}$ denotes the canonical basis vectors of $\mathbb{R}^d$. Denoting that $e_\emptyset$ is the basis element corresponding to $(\mathbb{R}^d)^{\otimes 0}$.

Suppose we have $e_I : |I| = N$ to be an orthonormal basis of $(\mathbb{R}^d)^{\otimes N}$. Then for any $\mathbf{a} \in T((\mathbb{R}^d))$, it can be written as

$$\mathbf{a} = \sum_{|I|>=0} a_I e_I,$$

where $a_I \in \mathbb{R}$ is a coefficient for a vector basis on multi-index I. Equivalently,

$$a_I = \langle e_I,\ \mathbf{a} \rangle.$$

2.2 Definitions And Theorems

In probability theory, a semimartingale could be understood as a model of a fair game where knowledge of past events never helps to predict future winnings. In financial mathematics context, the semimartingales are widely used to model asset prices since the financial instruments not only describe trends and patterns of the market, but also have a certain level of unpredictability.

We consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{Q})$, where $(\mathcal{F}_t)_{t\geq 0}$ is a filtration representing all available information and $\mathbb{Q}$ is a risk-neutral probability measure. Here we present the definition of the signature of a continuous semimartingale.

Definition 2.1 (Definition 2.1 in *Cuchiero et al. (2022)*[1]). *Let $(Y_t)_{t\in[0,T]}$ be a continuous $\mathbb{R}^d$-valued semimartingale. The signature of X is the $T((\mathbb{R}^d))$-valued process $(s,t) \mapsto \mathbb{Y}_{s,t}$ whose components are recursively defined as*

$$\langle e_\emptyset, \mathbb{Y}_{s,t} \rangle = 1,$$

$$\langle e_I, \mathbb{Y}_{s,t} \rangle = \int_s^t \langle e_{I'}, \mathbb{Y}_{s,t} \rangle \circ dX_r^{i_n},$$

for each $I = (i_1, i_2, ..., i_n)$, $I' = (i_1, i_2, ..., I_{n-1})$ and $0 \leqslant s \leqslant t \leqslant T$, where $\circ$ denotes the Stratonovich integral. Its projection $\mathbb{Y}^N$ on $T^{(N)}(\mathbb{R}^d)$ is given by

$$\mathbb{Y}_{s,t}^N = \sum_{|I|\leqslant N} \langle e_I, \mathbb{Y}_{s,t} \rangle e_I,$$

and is called signature of Y truncated at level N.
The equivalent notation:

$$\mathbb{Y}_t = (1,\ \int_0^t 1 \circ dY_s^1, ...,\ \int_0^t 1 \circ dY_s^d,\ \int_0^t (\int_0^t 1 \circ dY_s^1) \circ dY_s^1,\ ...\ ,\ \int_0^t (\int_0^t 1 \circ dY_s^d) \circ dY_s^d,\ ...\ ).$$

Signature is a collection of iterated integrals of the given multidimensional path. The integrals are listed in the collection in a strict order. Similar to 1.2.1 presented in *Chevyrev & Kormilitzin (2016)*[7], here is an example for a single signature term.

Example.

$$\langle e_1, \mathbb{Y}_{s,t}\rangle = \int_0^t 1 \circ dY_s^1 = Y_t^1 - Y_s^1,$$

Note that $\langle e_1 \otimes e_2, \mathbb{Y}_{s,t}\rangle \neq \langle e_2 \otimes e_1, \mathbb{Y}_{s,t}\rangle$ *as* $\langle e_1 \otimes e_2, \mathbb{Y}_{s,t}\rangle = \int_0^t (\int_0^t 1 \circ dY_s^1) \circ dY_s^2$ *while* $\langle e_2 \otimes e_1, \mathbb{Y}_{s,t}\rangle = \int_0^t (\int_0^t 1 \circ dY_s^2) \circ dY_s^1$.

Definition 2.2 and Proposition 2.1 together articulate a conclusion that is of significant importance in the computation of expected values of signatures, particularly when the terms are subjected to multiplication. To be more specific, every polynomial on the signature may be realized as a linear function via shuffle product.

Definition 2.2 (Definition 2.2 in *Cuchiero et al. (2022)* [1])**.** *(Shuffle Product) Given two multi-indices* $I = (i_1, i_2, ..., i_n)$ *and* $J = (j_1, j_2, ..., j_m)$*, the* ***shuffle product*** *is defined recursively as*

$$e_I \sqcup\!\sqcup\, e_J = (e_{I'} \sqcup\!\sqcup\, e_J) \otimes e_{i_n} + (e_I \sqcup\!\sqcup\, e_{J'}) \otimes e_{j_m},$$

where $e_I \sqcup\!\sqcup\, e_\emptyset \ = \ e_\emptyset \sqcup\!\sqcup\, e_I \ = \ e_I$.

Proposition 2.1. *(Shuffle Property)[Proposition 2.3 in Cuchiero et al. (2022)[1]] Let* $((X_t))t \in [0,T]$ *be a continuous* $\mathbb{R}^d$*-valued semimartingale and I, J be two multi-indices. Then*

$$\langle e_I, \mathbb{Y}_{s,t}\rangle\langle e_J, \mathbb{Y}_{s,t}\rangle \ = \ \langle e_I \sqcup\!\sqcup\, e_J, \mathbb{Y}_{s,t}\rangle,$$

Proof.

We prove this proposition by induction with Stratonovich integrals. See appendix A.

With the motivation from *Cuchiero et al. (2023)*[2] and *Cuchiero et al. (2022)*[1], we model the volatility process σ_t in our pricing formula for variance swap. The main idea of this modelling choice is to track the volatility process with the linear combinations of the signature of the underlying process.

Definition 2.3. *Our goal is to parameterize the volatility process* σ_t *as a signature model (Sig-SDE), which is, applying signature model to describe the dynamics of* S_t *as*

$$\sigma_t(\ell) = \ell_\emptyset + \sum_{|I| \leq n} \ell_I \langle e_I, \widehat{\mathbb{Z}}_t\rangle,$$

where $\widehat{\mathbb{Z}}_t$ *is the signature of* $\widehat{Z}_t = (t, W_t, B_t)$ *with the primary process* W_t *and Brownian motion* B_t *from pricing form. Furthermore, by Proposition 2.1,*

$$(\sigma_t(\ell))^2 = \ell_\emptyset + \sum_{|I|,|J| \leq n} \ell_I \ell_J \langle e_I \sqcup\!\sqcup\, e_J, \widehat{\mathbb{Z}}_t\rangle.$$

3 Expected Signature

In order to incorporate concept of signature into the pricing formula, it is essential to compute the expected values for signature terms. We use a very important definition in *Cuchiero et al. (2023)*[2] that introduces a matrix operator G that helps us to approximate expected signature values. As mentioned in Definition 2.1, to derive signature values, we need to have a underlying process who satisfies certain properties (namely a semimartingale). First we introduce a general form of the primary process X. *Cuchiero et al. (2023)*[2] incorporated multidimensional Ornstein–Uhlenbeck processes (OU processes) as their primary process. In our case, we adjust the process so that the primary process is Brownian motion. Also notice that the general primary process X_t is partly driven by the underlying Brownian motion W_t.

Definition 3.1. *A d-dimensional real-valued process Y is called a d-dimensional polynomial diffusion process if it is a weak solution of*

$$dY_t = b(Y_t)dt + \sqrt{a(Y_t)}dW_t, \quad Y_0 = y_0,$$

where maps $a : \mathbb{R}^d \mapsto \mathbb{S}_+^d$ and $b : \mathbb{R}^d \mapsto \mathbb{R}^d$ such that each a_{ij} is a polynomial of degree at most 2 and b_j is a polynomial of degree at most 1 for each $i,\ j = 1, ..., d+1$.

Let $\mathbb{Y}_t$ denote the signature of Y_t. Lemma 4.1 in *Cuchiero et al. (2023)*[2] shows specific details in drift coefficient a and diffusion coefficient b. Specifically, the signature terms are mapped to these coefficients by certain multi-indices $\mathbf{b}_j$ and $\mathbf{a}_{ij}$, where $i, j \in \{1, ..., d\}$,

$$b_j(Y_t) = \langle \mathbf{b}_j, \mathbb{Y}_t^1 \rangle,$$
$$a_{ij}(Y_t) = \langle \mathbf{a}_{ij}, \mathbb{Y}_t^2 \rangle.$$

Lemma 4.2 in *Cuchiero et al. (2023)*[2] indicates as the truncated signature $(\mathbb{Y}_t^n)_{t \geq 0}$ also admits a polynomial representation similar to form given in Definition 3.1, then for each $|I| \leq n$, we have

$$\langle e_I, \mathbb{Y}_t^n \rangle = \int_0^t \langle Le_I, \mathbb{Y}_s^n \rangle ds \ + \ \int_0^t \langle e_{I'}, \mathbb{Y}_s^n \rangle \sigma_{i_{|I|}}(Y_s) dW_s,$$

where the operator $L : T(\mathbb{R}^d) \to T(\mathbb{R}^d)$ satisfies $L(T^{(n)}(\mathbb{R}^d)) \subseteq T^{(n)}(\mathbb{R}^d)$ and is given by

$$Le_I = e_{I'} \sqcup\!\sqcup\, \mathbf{b}_{i_{|I|}} \ + \ \frac{1}{2} e_{I''} \sqcup\!\sqcup\, \mathbf{a}_{i_{|I|-1} i_{|I|}},$$

where $\mathbf{b}_{i_{|I|}}$ and $\mathbf{a}_{i_{|I|-1} i_{|I|}}$ are the multi-indices used in drift and diffusion coefficients' mappings.

Notice that the L-operator maps from $T(\mathbb{R}^d)$ to $T(\mathbb{R}^d)$, then it is reasonable to write a matrix representation on it.

Definition 3.2. *[Definition 4.3 in Cuchiero et al. (2023)[2].] For each $|I| \leq n$ set, then $\eta_{IJ} \in \mathbb{R}$ such that*

$$He_I = \sum_{|J| \leq n} \eta_{IJ} e_J.$$

Then we fix a labelling injective function $\mathcal{H} : \{I : |I| \leq n\} \to \{1, 2, ..., d_n\}$. We then call the matrix $G \in \mathbb{R}^{d_n \times d_n}$ where

$$G_{\mathcal{H}(I)\mathcal{H}(J)} = \eta_{IJ},$$

the d_n dimensional matrix representative of H.

Now we have everything to construct a conclusion that we could use to compute the expected signature. Given the information up to time T, $\mathcal{F}_T$, we would like to calculate the expected value for $\langle e_I, \mathbb{Y}_{T+t} \rangle$.

Definition 3.3 (Theorem 4.4 in *Cuchiero et al.(2023)*[2])**.** *Let $(Y_t)_{t \geq 0}$ be the polynomial process given by Theorem 3.1. Let G be the d_n-dimensional matrix representative of the L-operator corresponding to $\mathbb{Y}$. Then for each $T, t \geq 0, |I| \leq n$ we introduce a lifting operator $\mathbf{P}$ which is defined as:*

$$\mathbf{P}_t^I(\mathbb{Y}_T) \;=\; \mathbb{E}\left[\langle e_I, \mathbb{Y}_{T+t} \rangle \middle| \mathcal{F}_T \right] = \sum_{|J| \leq n} (e^{tG^\intercal})_{\mathcal{L}(I)\mathcal{L}(J)} \langle e_J, \mathbb{Y}_T \rangle.$$

By conditioning on $\mathcal{F}_T$, we're taking the conditional expectation as a function of the signature terms of the underlying process at time T. $\mathcal{F}_T$, the filtration up to time T, is the acquisition of all relevant information up to time T. Indeed, the signature values of the underlying process at time T is the information we have at time T.

In their work *Cuchiero et al. (2023)*[2], to incorporate signature-based model in explicit VIX pricing expression, operator $\mathbf{P}$ is enough for computation. In our case, the variance swap pricing form, it involves higher order terms that could not be elegantly presented by $\mathbf{P}$. Therefore, we define more complex operator for our variance swap pricing formula.

Lemma 3.1. *In variance swap pricing form, it involves terms in the form of*

$$\mathbb{E}\left[\langle e_I, \mathbb{Y}_{t_2} \rangle \langle e_H, \mathbb{Y}_{t_1} \rangle \middle| \mathcal{F}_{t_0} \right],$$

thus we introduce operator $\mathbf{\Phi}$, which could be seen as a function of operator $\mathbf{P}$. Specifically, when $0 \leq t_0 < t_1 < t_2$, by Tower's property,

$$
\begin{aligned}
\mathbf{\Phi}^{(I,t_2),(H,t_1)}(\mathbb{Y}_{t_0}) &= \mathbb{E}\Big[\ \langle e_I, \mathbb{Y}_{t_2}\rangle\langle e_H, \mathbb{Y}_{t_1}\rangle \ \Big| \mathcal{F}_{t_0}\Big] \\
&= \ \mathbb{E} \ \Big[\ \mathbb{E}\big[\langle e_I, \mathbb{Y}_{t_2}\rangle \big| \mathcal{F}_{t_1}\big] \ \langle e_H, \mathbb{Y}_{t_1}\rangle \ \Big| \ \mathcal{F}_{t_0}\Big] \\
&= \mathbb{E}\Big[\ \sum_{|J|\leq n} (e^{(t_2-t_1)G^\intercal})_{\mathcal{L}(I)\mathcal{L}(J)}\langle e_J, \mathbb{Y}_{t_1}\rangle\langle e_H, \mathbb{Y}_{t_1}\rangle \ \Big| \ \mathcal{F}_{t_0}\Big] \\
&= \mathbb{E}\Big[\ \sum_{|J|\leq n} (e^{(t_2-t_1)G^\intercal})_{\mathcal{L}(I)\mathcal{L}(J)}\langle e_J \sqcup e_H, \mathbb{Y}_{t_1}\rangle \ \Big| \ \mathcal{F}_{t_0}\Big] \\
&= \sum_{|J|\leq n} (e^{(t_2-t_1)G^\intercal})_{\mathcal{L}(I)\mathcal{L}(J)}\mathbb{E}\Big[\ \langle e_J \sqcup e_H, \mathbb{Y}_{t_1}\rangle \ \Big| \ \mathcal{F}_{t_0}\Big] \\
&= \sum_{|J|\leq n} (e^{(t_2-t_1)G^\intercal})_{\mathcal{L}(I)\mathcal{L}(J)}\mathbf{P}^{J\sqcup H}_{t_1-t_0}(\mathbb{Y}_{t_0}).
\end{aligned}
$$

4 Calibration Of Variance Swap Price With Signature Model

In section 3, the G matrix operator and the lifting operator $\mathbf{P}$ and $\mathbf{\Phi}$ are defined on general primary process X_t, whose dynamics were given in Definition 3.1. However, we need to revisit the fact in our project that our primary process is one-dimensional Brownian motion W_t. To be more specific, given the problem setup in *Cuchiero et al. (2023)*[2], we could choose the parameters in H-operator to formulate X_t as W_t.

Given Lemma 4.1. in *Cuchiero et al. (2023)*[2], if we let $b_j^c, b_j^k, a_{ij}^k, a_{ij}^{kh} = 0$ and $a_{ij}^c = 1$, then we have $dX_t = dW_t$. Now we've formulated the general polynomial diffusion process X_t as the Brownian motion W_t.

Therefore here we use the calibration on the most W_t. $\widehat{W_t} = (t, W_t)$ is the time-extended process where $\widehat{\mathbb{W}}_t$ being its signature. Notice that for the last term in the formula, namely the K_3 in Theorem 4.1, involves the brownian motion B from the pricing formula. So for K_3 computation, we use $\widehat{Z}_t = (t, W_t, B_t)$ as the underlying process we're calibrating on. $\widehat{\mathbb{Z}}_t$ is the signature of $\widehat{Z}_t$. Given a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{Q})$ on which we define the stochastic process

$$dS_t = rS_t dt \ + \ \sigma_t S_t dB_t, \quad S_0 \in \mathbb{R}_+,$$

where S_t is the pricing process, σ_t is the volatility process represented by signature model. B_t here is a Brownian motion which is correlated with σ_t.

The variance swap contract is structured to ensure that losses are minimized when market volatility exceeds the agreed strike price. Thus in the following theorem, we formulate the closed form for our objective strike price $\hat{K}$. To make the market free of arbitrage, the fair strike price $\hat{K}$ under risk-free measure $\mathbb{Q}$ is given by

$$\hat{K} = \mathbb{E}^{\mathbb{Q}}\left[\sum_{k=1}^{N} \left[\log(\frac{S_{t_{k+1}}}{S_{t_k}})\right]^2 \middle| \mathcal{F}_0\right].$$

Then from SDE formulated earlier, we get

$$S_{t_{k+1}} = S_{t_k} e^{r(t_{k+1}-t_k) - \int_{t_k}^{t_{k+1}} \frac{\sigma_s^2}{2} ds + \int_{t_k}^{t_{k+1}} \sigma_s dB_s}.$$

Then $\hat{K}$ is given by (see proof in appendix)

$$\hat{K} = \mathbb{E}\Big[\sum_{k=0}^{N}\Big(r^2(t_{k+1}-t_k)^2 + \big(1 - r(t_{k+1}-t_k)\big)\int_{t_k}^{t_{k+1}}\sigma_s^2 ds + \frac{1}{4}(\int_{t_k}^{t_{k+1}}\sigma_s^2 ds)^2 - \int_{t_k}^{t_{k+1}}\sigma_s^2 ds\int_{t_k}^{t_{k+1}}\sigma_s dB_s\Big)\Big|\mathcal{F}_0\Big].$$

Theorem 4.1. *[3] *Assume Definition 2.3, Definition 3.3, Definition 3.1 and the linear signature presentation for dynamic process, the risk-neutral strike $\widehat{K}$ for variance swap is given by*

$$\hat{K} = \sum_{k=0}^{N}\Big(r^2(t_{k+1}-t_k)^2 + \big(1 - r(t_{k+1}-t_k)\big)\mathcal{K}^1_{t_k,t_{k+1}} + \frac{1}{4}\mathcal{K}^2_{t_k,t_{k+1}} - \mathcal{K}^3_{t_k,t_{k+1}}\Big),$$

where

(1) $\mathcal{K}^1_{t_k,t_{k+1}} = \mathbb{E}\Big[\int_{t_k}^{t_{k+1}}\sigma_s^2 ds\Big|\mathcal{F}_0\Big] = \sum_{|I|,|J|\le n}\ell_I\ell_J\Big(\mathbf{P}^{I\sqcup J\otimes\mathbf{0}}_{t_{k+1}}(\widehat{\mathbb{W}}_0) - \mathbf{P}^{I\sqcup J\otimes\mathbf{0}}_{t_k}(\widehat{\mathbb{W}}_0)\Big),$

(2) $\mathcal{K}^2_{t_k,t_{k+1}} = \mathbb{E}\Big[(\int_{t_k}^{t_{k+1}}\sigma_s^2 ds)^2\Big|\mathcal{F}_0\Big] = \sum_{|I|,|J|,|M|,|N|\le n}\ell_I\ell_J\ell_M\ell_N\Big(\mathbf{P}^{(I\sqcup J\otimes\mathbf{0})\sqcup(M\sqcup N\otimes\mathbf{0})}_{t_{k+1}}(\widehat{\mathbb{W}}_0)$
$+ \mathbf{P}^{(I\sqcup J\otimes\mathbf{0})\sqcup(M\sqcup N\otimes\mathbf{0})}_{t_k}(\widehat{\mathbb{W}}_0) - 2\mathbf{\Phi}^{(I\sqcup J\otimes\mathbf{0},t_{k+1}),(M\sqcup N\otimes\mathbf{0},t_k)}(\widehat{\mathbb{W}}_0)\Big),$

(3) $\mathcal{K}^3_{t_k,t_{k+1}} = \mathbb{E}\Big[\int_{t_k}^{t_{k+1}}\sigma_s^2 ds\int_{t_k}^{t_{k+1}}\sigma_s dB_s\Big|\mathcal{F}_0\Big] = \sum_{|I|,|J|,|M|\le n}\ell_I\ell_J\ell_M\Big(\mathbf{P}^{(I\sqcup J\otimes\mathbf{0})\sqcup\tilde{M}^{d+1}}_{t_{k+1}}(\widehat{\mathbb{Z}}_0)$
$+ \mathbf{P}^{(I\sqcup J\otimes\mathbf{0})\sqcup\tilde{M}^{d+1}}_{t_k}(\widehat{\mathbb{Z}}_0) - \mathbf{\Phi}^{(I\sqcup J\otimes\mathbf{0},t_{k+1}),(\tilde{M}^{d+1},t_k)}(\widehat{\mathbb{Z}}_0) - \mathbf{\Phi}^{(\tilde{M}^{d+1},t_{k+1}),(I\sqcup J\otimes\mathbf{0},t_k)}(\widehat{\mathbb{Z}}_0)\Big).$

Proof. See appendix A.

Notice that the $\mathcal{K}_1$ is similar to the computation of the VIX squared introduced in Theorem 5.1 from *Cuchiero et al. (2023)*[2]. However when incorporated the signature models, we use different notations specified in Definition 3.3. When constructing $\mathcal{K}_1$, $\mathcal{K}_2$ and $\mathcal{K}_3$, we organized the expectation terms with similar structures and summarized them into Definition 3.3 and Definition 3.1, which makes our complete formula more orderly and easier to interpret.

[3] Theorem 4.1.* is a joint effort with my collaborators, Prof. Qi Feng, Prof. Bingyan Han and Haisu Ding in *Han et al. (2023)*[10].

5 Data Analysis

5.1 Data Description

We utilize a dataset comprising annualized variance swap prices on the S&P 500 Index (SPX), sourced from the Bloomberg Terminal at Ross School of Business, University of Michigan. The dataset spans from 2008 to 2023, and we are particularly interested in modeling periods during abnormal price behavior, such as the 2008 financial crisis and the economic impacts of the COVID-19 pandemic during 2019-2020.

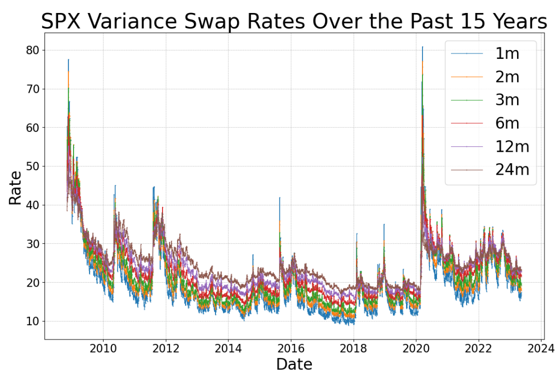


Figure 2: SPX Variance Swap Annualised Prices 2008 - 2023

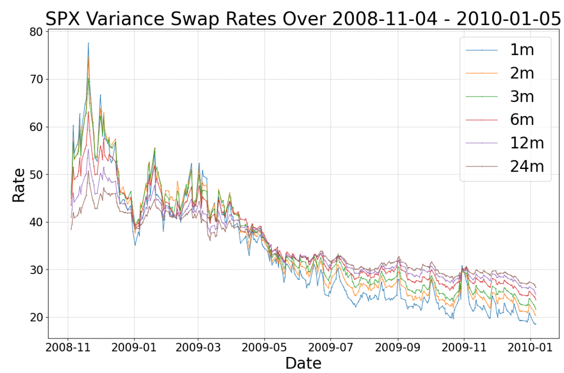


Figure 3: SPX Variance Swap Annualised Prices during the 2008 Financial Crisis

In convention, the volatility is calculated using *daily* returns, where the annualised volatility is presented as,

$$\sigma^2_{\text{realized}} = \frac{252}{T} \sum_{i=1}^{T} \left[\log(\frac{S_i}{S_{i-1}}) \right]^2, \tag{1}$$

where S_i is the underlying price on day i, and T is the maturities in number of days. In daily setup, the calculation is conducted for each trading day. To reduce computational intensity in this project, we switch to modeling *monthly* returns. Therefore we conduct calculation once for each month. Then in our case, we have

$$\sigma^2_{\text{realized}} = \frac{12}{M} \sum_{i=1}^{M} \left[\log(\frac{S_i}{S_{i-1}}) \right]^2 \approx \frac{12}{M} K, \tag{2}$$

where S_i is the underlying price on month i, M is the maturities in number of months and $K = \mathbb{E}[\sum_{i=1}^{M} [\log(\frac{S_i}{S_{i-1}})]^2]$ is the risk-neutral strike price.

5.2 Method Specification

Observe Definition 3.3, Definition 3.1 and the proof for Theorem 4.1, for efficiency purpose, the $e^{tG^{\intercal}}$ matrices should be stored and loaded when needed as the matrix exponential computation in such case is time consuming and hard to approximate.

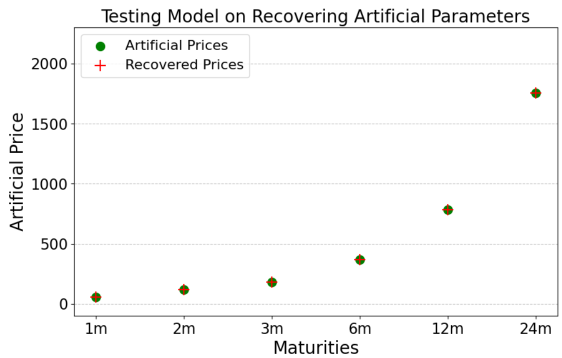
Notice that G matrix is determined by the primary process of different choices. In their work in *Cuchiero et al. (2023)*[2], they choose multidimensional OU-processes X which differs from our case, one-dimensional Brownian motion W, then the G matrices are not align with each other in our cases consequently. We're not using consistent G matrices in this project either. For computation of $\mathcal{K}_1$ and $\mathcal{K}_2$, G matrices are generated on $\widehat{W} = (t, W)$. For $\mathcal{K}_3$, which involves the pricing form Brownian motion B, G matrices are generated on $\widehat{Z} = (t, W, B)$. The size of G matrices also differ for different index lengths. (In Definition 3.2, $|I|$ determines different size of G matrices.) In $\mathcal{K}_3$ case, W and B represents two correlated Brownian motions that form a 2-dimensional diffusion process given in Definition 3.1. In this case, the diffusion coefficients a_{ij} are actually the correlation coefficients ρ_{ij} of different processes. Namely in our current computation, we assume the correlation coefficient between W and B to be -0.2. Specifically, our input ρ-matrix into Gs' computation for $\mathcal{K}_3$ is given by

$$\rho = \begin{pmatrix} 1 & -0.2 \\ -0.2 & 1 \end{pmatrix}.$$

Also, we assume interest rate r to be 0.05.

5.3 Model Validation Test

We conduct a validation test on our existing model by manually defining the parameters and generating artificial prices using these parameters. The plot below presents the fact that our model is able to recover the parameters we previously set, as all newly calibrated prices recover the artificial prices.



5.4 Static Case

On each trading day, we could observe the variance swap price on 1 month, 2 months, 3 months, 6 months, 12 months, 24 months maturities. We pick a single trading day and calibrate the 6 pricing points together. In this case we choose the truncation level to be 1 and the number of parameters is 4. On trading day $\mathcal{T}$, the calibration on variance swap prices on different maturities $\mathcal{M} = \{1m, 2m, 3m, 6m, 12m, 24m\}$ consists in minimizing the functional

$$\mathsf{L}(\ell) = \sum_{m \in \mathcal{M}} (\hat{K}_m^{\mathcal{T}} - K_m^{*\mathcal{T}})^2,$$

where L denotes the real value loss function. $\hat{K}_m^{\mathcal{T}}$ and $K_m^{*\mathcal{T}}$ are the calibrated and observed strike prices respectively.

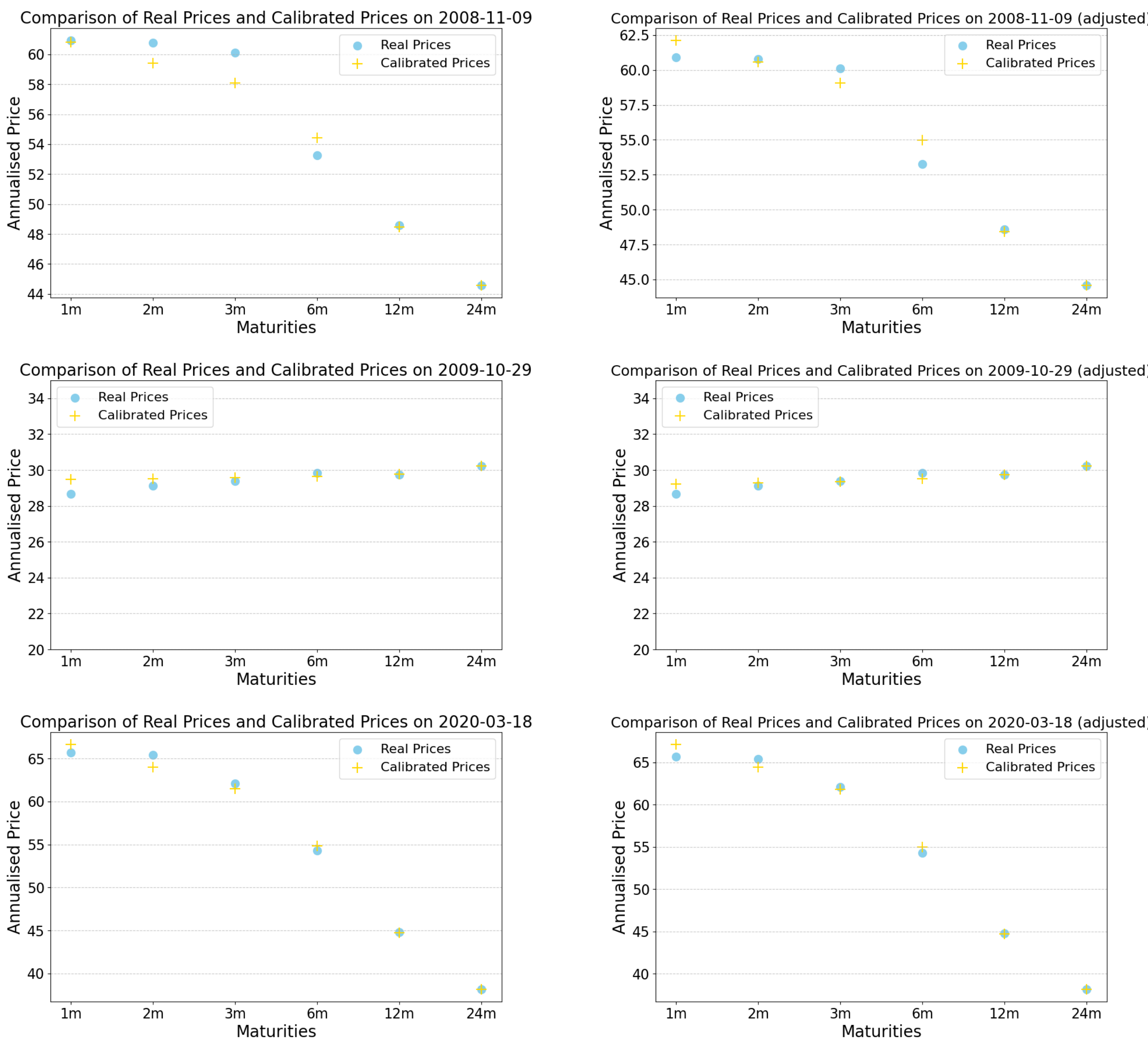


Figure 4: Variance Swap prices calibration (Static Case)

The left column in Figure 4 represents the original calibration results. The three date we picked: 2008-11-09 has a high variance swap price level during the 2008 financial crisis era, 2009-10-29 is an ‘ordinary’ day where the price level are comparatively low, 2020-03-18 has a high price level during COVID-19 pandemic period.

Our model shows good calibration results on most price points, except for the first three maturities, the results are slightly off. We manually add penalties to the first three maturities in the cost function. Then we derived the ‘adjusted’ calibration on the right. We observe much better calibrated prices on 2-months and 3-months maturities while failed to improve on 1-month maturity.

5.5 Use Static Setup To Test Model's Consistency Within A Short Period

Within a short period of time (a series of sequential trading days), the calibrated parameters should be considered as the estimates $\ell^* = (\ell^*_\emptyset, \ell^*_0, \ell^*_1, \ell^*_2)$ of the consistent 'real' parameters $\ell = (\ell_\emptyset, \ell_0, \ell_1, \ell_2)$. As the model is tracking the volatility within the short period, the model results should be consistent during the period. From Table 1, we could observe that the calibrated parameters have lower variance than the calibrated parameters in Table 2 where we chose distant trading days. Parameters ℓ should not be stationary in a long term as the market evolves, so it should only be consistent within a specified time window.

Dates \Parameters	$\ell^*_\emptyset$	ℓ^*_0	ℓ^*_1	ℓ^*_2
2008-11-10	1.63652	-0.07049	0.30366	-0.07712
2008-11-11	1.64862	-0.07365	0.3089	-0.08266
2008-11-12	1.68402	-0.07924	0.31386	-0.10649
2008-11-13	1.63192	-0.0651	0.29517	-0.07174
2008-11-14	1.67081	-0.0718	0.30424	-0.09012

Table 1: Calibrated parameters from 2008.11.10 to 2008.11.14

Dates \Parameters	$\ell^*_\emptyset$	ℓ^*_0	ℓ^*_1	ℓ^*_2
2008-11-10	1.63652	-0.07049	0.30366	-0.07712
2009-04-06	1.52544	-0.0379	0.23967	-0.00226
2009-08-27	1.38714	-0.02374	0.20564	0.09479
2010-01-21	1.3193	-0.02288	0.19698	0.13739
2008-06-21	1.37418	-0.0233	0.20951	0.10141

Table 2: Calibrated parameters from 2008.11.10 with 100 trading days' gaps

The purpose of calibration is to adjust the parameters of a model so that its output aligns with real-world prices. Apart from providing some evidence on the model's accuracy, we can offer reference opinions for customized orders using an approach similar to interpolation. Notice that now we're calibrating on 6 standard maturities variance swap. What if there are some taylor-made contracts that have non-standard maturities like 17 days? What we could do is once we acquire the 6 standard maturities variance swap prices, we conduct calibration on these 6 pricing points. Once we derived the calibrated parameters, with these parameters and the customized maturity, we could offer a variance swap price to this taylor-made contract.

5.6 Sampled Case

Now we pick one maturity, sample a path with different number of days to derive signature values. Then we calibrate these number of days together. This sampled case could be a good model performance test on series of prices. We pick maturity $\mathcal{M}$, the calibration for each day in $\mathcal{T}$ consists in minimizing the functional

$$\mathsf{L}(\ell) = \sum_{t \in \mathcal{T}} (\hat{K}^t_{\mathcal{M}} - K^{*t}_{\mathcal{M}})^2,$$

where L denotes the real value loss function. $\hat{K}^{\mathcal{T}}_m$ and $K^{*\mathcal{T}}_m$ are the calibrated and observed strike prices respectively.

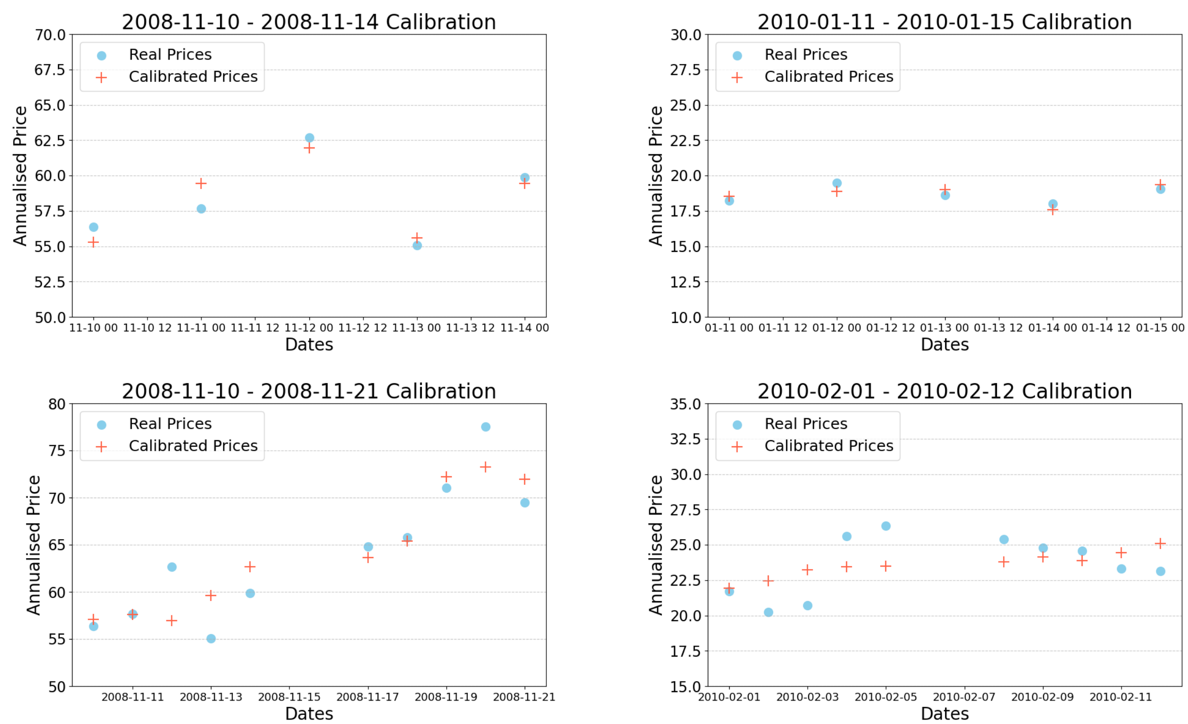


Figure 5: Variance Swap prices calibration (Sampled Case)

In this case we still choose the truncation level to be 1 and the number of parameters is 4. The four calibration presented above are all on 1-month maturity variance swap price. As we're using annualized prices, and with observations from Figure 2 and 3, on different maturities, the model's performance should be similar (see Figure 6). Calibration results are mostly good for 4 parameters versus 5 data points (two plots on the first row), while the calibration is poor when our model's 4 parameters are trained on 10 price points (two plots on the second row). Other than the number of parameters, another factor that could influence our model performance is the primary process we chose. We incorporated correlated

Brownian motions as our primary process. However, correlated Brownian motions may not provide advanced physical features which could potentially improve fitting. For example, the next stage of this project could be incorporating mean-reverting process like OU-process instead of correlated Brownian motions. Moreover, in this case, we choose $d = 1$, which means our underlying Brownian Motion is one-dimensional. One-dimensional Brownian motion may not be sufficient to capture all the dynamics of the pricing process; exploring higher-dimensional models could yield more accurate and comprehensive results.

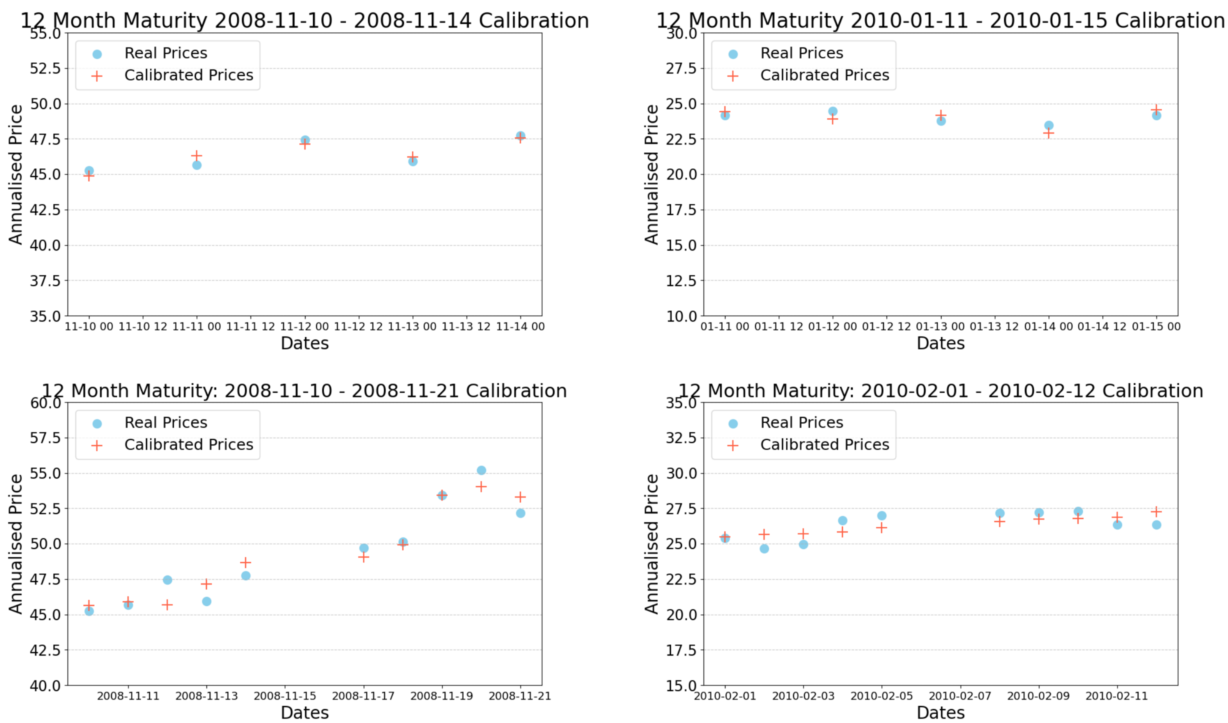


Figure 6: 12 Month Maturity Variance Swap prices calibration (Sampled Case). We observe a similar result with 1 Month Maturity case in Figure 5.

6 Conclusion

In this thesis, we explored a new approach for pricing variance swaps using signature methods, a recent development in quantitative finance. Our focus was to build a model that accurately captures the behavior of volatility in financial markets. This approach is based on the concept of rough path theory, which provides a detailed understanding of stochastic processes. Our signature model, designed to represent volatility, has shown its effectiveness in handling complex financial instruments. The model's performance was tested against real market data, particularly during periods of significant market fluctuations like the 2008 financial crisis and the COVID-19 pandemic. The results were encouraging, demonstrating the model's potential in capturing market dynamics.

We applied our model in two scenarios: static and sampled cases. The static case showed the model's accuracy in predicting prices for standard variance swaps. The sampled case, although more challenging, pointed out areas for further improvement. This includes exploring more complex mathematical processes or considering models with more parameters.

In summary, the findings from this study are valuable for professionals in finance, offering a new tool for the understanding and managing volatility. The signature model's ability to summarize key information from price movements can lead to better risk management and strategic decisions in financial markets. Our signature-based model is a significant addition to financial modeling techniques. It stands out in its ability to handle complex market conditions and provides detailed insights into volatility. Future research could enhance this model further, applying it to other financial instruments and improving its predictive power. This thesis lays a foundation for future advancements in financial modeling, promising to be of great benefit both in theory and practice.

A Appendix: Proof for Propostion 2.1

We prove this proposition by induction with Stratonovich integrals.

Induction Basis. $\langle e_\emptyset, \mathbb{X}_{s,t}\rangle\langle e_J, \mathbb{X}_{s,t}\rangle = 1\ \langle e_J, \mathbb{X}\rangle = \langle e_\emptyset \sqcup e_J, \mathbb{X}_{s,t}\rangle$, which agrees with the proposition.

Induction Steps. Assumed that for arbitrary subsets S_1 and S_2 of multi indices I and J, we have

$$\langle e_{S_1}, \mathbb{X}_{s,t}\rangle\langle e_{S_2}, \mathbb{X}_{s,t}\rangle \ = \ \langle e_{S_1} \sqcup e_{S_2}, \mathbb{X}_{s,t}\rangle$$

After that,

$$\begin{aligned}
\langle e_I, \mathbb{X}_{s,t}\rangle\langle e_J, \mathbb{X}_{s,t}\rangle &= \int_s^t \langle e_J, \mathbb{X}_{s,u}\rangle \circ d\langle e_I, \mathbb{X}_{s,u}\rangle + \int_s^t \langle e_I, \mathbb{X}_{s,u}\rangle \circ d\langle e_J, \mathbb{X}_{s,u}\rangle \\
&= \int_s^t \langle e_{I'}, \mathbb{X}_{s,r}\rangle\langle e_J, \mathbb{X}_{s,r}\rangle \circ d\mathbb{X}_r^{i_n} + \int_s^t \langle e_I, \mathbb{X}_{s,r}\rangle\langle e_{J'}, \mathbb{X}_{s,r}\rangle \circ d\mathbb{X}_r^{j_m} \\
&= \int_s^t \langle e_{I'} \sqcup e_J, \mathbb{X}_{s,r}\rangle \circ d\mathbb{X}_r^{i_n} + \int_s^t \langle e_I \sqcup e_{J'}, \mathbb{X}_{s,r}\rangle \circ d\mathbb{X}_r^{j_m} \\
&= \langle e_{I'} \sqcup e_J \otimes e_{i_n}, \mathbb{X}_{s,t}\rangle + \langle e_I \sqcup e_{J'} \otimes e_{j_m}, \mathbb{X}_{s,t}\rangle \\
&= \langle e_{I'} \sqcup e_J \otimes e_{i_n} + e_I \sqcup e_{J'} \otimes e_{j_m}, \mathbb{X}_{s,t}\rangle \\
&= \langle e_I \sqcup e_J, \mathbb{X}_{s,t}\rangle
\end{aligned}$$

B Appendix: Proof for Theorem 4.1

Given

$$\hat{K} = \mathbb{E}^{\mathbb{Q}}\Big[\sum_{k=1}^{N} \big[\ln(\frac{S_{t_{k+1}}}{S_{t_k}})\big]^2 \Big| \mathcal{F}_0\Big]$$

and

$$S_{t_{k+1}} = S_{t_k} e^{r(t_{k+1}-t_k) - \int_{t_k}^{t_{k+1}} \frac{\sigma_s^2}{2} ds + \int_{t_k}^{t_{k+1}} \sigma_s dB_s},$$

we derive

$$
\begin{aligned}
\hat{K} &= \mathbb{E}\Big[\sum_{k=0}^{N}\Big(r(t_{k+1}-t_k) - \int_{t_k}^{t_{k+1}} \frac{\sigma_s^2}{2} ds + \int_{t_k}^{t_{k+1}} \sigma_s dB_s\Big)^2 \Big| \mathcal{F}_0\Big] \\
&= \mathbb{E}\Big[\sum_{k=0}^{N}\Big(r^2(t_{k+1}-t_k)^2 \; - \; r(t_{k+1}-t_k)\int_{t_k}^{t_{k+1}} \sigma_s^2 ds \; + \; 2r(t_{k+1}-t_k)\int_{t_k}^{t_{k+1}} \sigma_s dB_s \\
&\qquad + \; \frac{1}{4}(\int_{t_k}^{t_{k+1}} \sigma_s^2 ds)^2 \; - \; \int_{t_k}^{t_{k+1}} \sigma_s^2 ds \int_{t_k}^{t_{k+1}} \sigma_s dB_s \; + \; (\int_{t_k}^{t_{k+1}} \sigma_s dB_s)^2\Big)\Big| \mathcal{F}_0\Big] \\
&= \mathbb{E}\Big[\sum_{k=0}^{N}\Big(r^2(t_{k+1}-t_k)^2 \; - \; r(t_{k+1}-t_k)\int_{t_k}^{t_{k+1}} \sigma_s^2 ds \\
&\qquad + \; \frac{1}{4}(\int_{t_k}^{t_{k+1}} \sigma_s^2 ds)^2 \; - \; \int_{t_k}^{t_{k+1}} \sigma_s^2 ds \int_{t_k}^{t_{k+1}} \sigma_s dB_s \; + \; \int_{t_k}^{t_{k+1}} \sigma_s^2 ds\Big)\Big| \mathcal{F}_0\Big] \\
&= \mathbb{E}\Big[\sum_{k=0}^{N}\Big(\; r^2(t_{k+1}-t_k)^2 \; + \; \big(1 \; - \; r(t_{k+1}-t_k)\big)\int_{t_k}^{t_{k+1}} \sigma_s^2 ds + \; \frac{1}{4}(\int_{t_k}^{t_{k+1}} \sigma_s^2 ds)^2 \\
&\qquad - \; \int_{t_k}^{t_{k+1}} \sigma_s^2 ds \int_{t_k}^{t_{k+1}} \sigma_s dB_s\Big)\Big| \mathcal{F}_0\Big]
\end{aligned}
$$

Then with Definition 3.3 and Lemma 3.1, we express the formula for K in signature terms.

step **1**. We take $\mathcal{K}^1_{t_k,t_{k+1}} = \mathbb{E}\Big[\int_{t_k}^{t_{k+1}} \sigma_s^2 ds | \mathcal{F}_0\Big]$. With Remark **??** where we apply signature formula in term σ and Theorem 3.3

$$
\begin{aligned}
\mathcal{K}^1_{t_k,t_{k+1}} &= \mathbb{E}\Big[\int_{t_k}^{t_{k+1}} \sigma_s^2 ds\Big|\mathcal{F}_0\Big] \\
&= \mathbb{E}\Big[\int_{t_k}^{t_{k+1}} \sum_{|I|,|J|\le n} \ell_I\ell_J\langle e_I \sqcup e_J, \widehat{\mathbb{W}}_s\rangle ds\Big|\mathcal{F}_0\Big] \\
&= \mathbb{E}\Big[\int_{0}^{t_{k+1}} \sum_{|I|,|J|\le n} \ell_I\ell_J\langle e_I \sqcup e_J, \widehat{\mathbb{W}}_s\rangle ds\Big|\mathcal{F}_0\Big] \;-\; \mathbb{E}\Big[\int_{0}^{t_k} \sum_{|I|,|J|\le n} \ell_I\ell_J\langle e_I \sqcup e_J, \widehat{\mathbb{W}}_s\rangle ds\Big|\mathcal{F}_0\Big] \\
&= \mathbb{E}\Big[\sum_{|I|,|J|\le n} \ell_I\ell_J\langle e_I \sqcup e_J\otimes e_0, \widehat{\mathbb{W}}_{t_{k+1}}\rangle\Big|\mathcal{F}_0\Big] \;-\; \mathbb{E}\Big[\sum_{|I|,|J|\le n} \ell_I\ell_J\langle e_I \sqcup e_J\otimes e_0, \widehat{\mathbb{W}}_{t_k}\rangle\Big|\mathcal{F}_0\Big] \\
&= \sum_{|I|,|J|\le n} \ell_I\ell_J\mathbb{E}\Big[\langle e_I \sqcup e_J\otimes e_0, \widehat{\mathbb{W}}_{t_{k+1}}\rangle\Big|\mathcal{F}_0\Big] \;-\; \sum_{|I|,|J|\le n} \ell_I\ell_J\mathbb{E}\Big[\langle e_I \sqcup e_J\otimes e_0, \widehat{\mathbb{W}}_{t_k}\rangle\Big|\mathcal{F}_0\Big] \\
&= \sum_{|I|,|J|\le n} \ell_I\ell_J\mathbf{P}^{I\sqcup J\otimes\mathbf{0}}_{t_{k+1}}(\widehat{\mathbb{W}}_0) \;-\; \sum_{|I|,|J|\le n} \ell_I\ell_J\mathbf{P}^{I\sqcup J\otimes\mathbf{0}}_{t_k}(\widehat{\mathbb{W}}_0) \\
&= \sum_{|I|,|J|\le n} \ell_I\ell_J\Big(\mathbf{P}^{I\sqcup J\otimes\mathbf{0}}_{t_{k+1}}(\widehat{\mathbb{W}}_0) \;-\; \mathbf{P}^{I\sqcup J\otimes\mathbf{0}}_{t_k}(\widehat{\mathbb{W}}_0)\Big)
\end{aligned}
$$

step **2**. $\mathcal{K}^2_{t_k,t_{k+1}} \;=\; \mathbb{E}\Big[(\int_{t_k}^{t_{k+1}} \sigma_s^2 ds)^2|\mathcal{F}_0\Big]$. With Theorem 3.3 and Remark 3.1

$$
\begin{aligned}
\mathcal{K}^2_{t_k,t_{k+1}} &= \mathbb{E}\Big[(\int_{t_k}^{t_{k+1}} \sigma_s^2 ds)^2|\mathcal{F}_0\Big] \\
&= \mathbb{E}\Big[(\int_{0}^{t_{k+1}} \sigma_s^2 ds \;-\; \int_0^{t_k} \sigma_s^2 ds)^2\Big|\mathcal{F}_0\Big] \\
&= \mathbb{E}\Big[(\int_{0}^{t_{k+1}} \sigma_s^2 ds)^2 + (\int_0^{t_k} \sigma_s^2 ds)^2 - 2(\int_{0}^{t_{k+1}} \sigma_s^2 ds \int_0^{t_k} \sigma_s^2 ds)\Big|\mathcal{F}_0\Big] \\
&= \sum_{|I|,|J|,|M|,|N|\le n} \ell_I\ell_J\ell_M\ell_N\Bigg(\mathbf{P}^{(I\sqcup J\otimes\mathbf{0})\sqcup(M\sqcup N\otimes\mathbf{0})}_{t_{k+1}}(\widehat{\mathbb{W}}_0) \;+\; \mathbf{P}^{(I\sqcup J\otimes\mathbf{0})\sqcup(M\sqcup N\otimes\mathbf{0})}_{t_k}(\widehat{\mathbb{W}}_0) \\
&\quad -\; 2\Phi^{(I\sqcup J\otimes\mathbf{0},t_{k+1}),(M\sqcup N\otimes\mathbf{0},t_k)}(\widehat{\mathbb{W}}_0)\Bigg)
\end{aligned}
$$

step **3**. Before we dive into the actual computation of the cross term, we introduce the special case when we take the extra correlated Brownian process B into the picture. If we take W as our multidimensional brownian process and the underlying primary process under the signature model of $\sigma = \sum_{|I|\le n} \ell_I\langle e_I, \widehat{\mathbb{W}}\rangle$, here we need to define the process $Z = (W, B)$ and the time-extended version $\widehat{Z} = (t, W, B)$ as well as its signature $\widehat{\mathbb{Z}}$.

Assumption B.1. *Cuchiero et al. (2022) [1] For all $i \in 1, 2, ..., d$, we have*

$$d[W^i, Z^{d+1}]_t = \sum_{|J| \leq m} a^J_{i(d+1)} \langle e_J, \widehat{\mathbb{Z}}_t \rangle dt$$

for some $m \in \mathbb{N}$ and $a \in \mathbb{R}$ where $\widehat{Z} = (t, W, B)$.

Now we compute the special case correlated with B.

$$\begin{aligned}
\int_0^T \sigma_s dZ_s^{d+1} &= \int_0^T \sigma_s \circ dZ_s^{d+1} - \frac{1}{2}[\sigma, Z^{d+1}]_T \\
&= \int_0^T \sum_{|I| \leq n} \ell_I \langle e_I, \widehat{\mathbb{Z}}_s \rangle \circ dZ_s^{d+1} - \frac{1}{2}[\sum_{|I| \leq n} \ell_I \langle e_I, \widehat{\mathbb{Z}} \rangle, Z^{d+1}]_T \\
&= \sum_{|I| \leq n} \ell_I \langle e_I \otimes e_{d+1}, \widehat{\mathbb{Z}}_T \rangle - \frac{1}{2} \int_0^T \sum_{|I| \leq n} \ell_I \langle e_{I'}, \widehat{\mathbb{Z}}_s \rangle d[Z^{i_{|I|}}, Z^{d+1}]_s \\
&= \sum_{|I| \leq n} \ell_I \Big(\langle e_I \otimes e_{d+1}, \widehat{\mathbb{Z}}_T \rangle - \frac{1}{2} \int_0^T \langle e_{I'}, \widehat{\mathbb{Z}}_s \rangle d[Z^{i_{|I|}}, Z^{d+1}]_s \Big) \qquad (3) \\
&= \sum_{|I| \leq n} \ell_I \Big(\langle e_I \otimes e_{d+1}, \widehat{\mathbb{Z}}_T \rangle - \frac{1}{2} \sum_{|J| \leq m} a^J_{i_{|I|}(d+1)} \langle e_{I'} \sqcup e_J \otimes e_0, \widehat{\mathbb{Z}}_T \rangle \Big) \qquad (4) \\
&= \sum_{|I| \leq n} \ell_I \Big(\langle e_I \otimes e_{d+1} - \frac{1}{2} \sum_{|J| \leq m} a^J_{i_{|I|}d+1} \; e_{I'} \sqcup e_J \otimes e_0, \widehat{\mathbb{Z}}_T \rangle \Big) \\
&= \sum_{|I| \leq n} \ell_I \; \langle \tilde{e}_I^{d+1}, \widehat{\mathbb{Z}}_T \rangle
\end{aligned}$$

where $\tilde{e}_I^{d+1} = e_I \otimes e_{d+1} - \frac{1}{2} \sum_{|J| \leq m} a^J_{i_{|I|}d+1} \; e_{I'} \sqcup e_J \otimes e_0$.
(Note that the highest order of $\tilde{e}_I^{d+1}$ is $n + m$)
From (3) to (4) with assumption B.1 :

$$\begin{aligned}
\int_0^T \langle e_{I'}, \widehat{\mathbb{Z}}_s \rangle d[Z^{i_{|I|}}, Z^{d+1}]_s &= \int_0^T \langle e_{I'}, \widehat{\mathbb{Z}}_s \rangle \sum_{|J| \leq m} a^J_{i_{|I|}d+1} \langle e_J, \widehat{\mathbb{Z}}_s \rangle ds \\
&= \int_0^T \sum_{|J| \leq m} a^J_{i_{|I|}d+1} \langle e_{I'}, \widehat{\mathbb{Z}}_s \rangle \langle e_J, \widehat{\mathbb{Z}}_s \rangle ds \\
&= \sum_{|J| \leq m} a^J_{i_{|I|}d+1} \langle e_{I'} \sqcup e_J \otimes e_0, \widehat{\mathbb{Z}}_T \rangle
\end{aligned}$$

Remark. When we are not incorporate any process form (like OU, Heston) into our model and we are calibrating directly on the underlying Brownian motions, the term could be written as

$$\tilde{e}_I^{d+1} = e_I \otimes e_{d+1} - \frac{1}{2}\sigma^{i_{|I|}}\rho_{i_{|I|}d+1}\ e_{I'} \otimes e_0$$

Now we compute the explicit formula of the cross term.

$$\mathcal{K}^3_{t_k,t_{k+1}} \ = \ \mathbb{E}\Big[\int_{t_k}^{t_{k+1}} \sigma_s^2 ds \int_{t_k}^{t_{k+1}} \sigma_s dZ_s^{d+1}\Big|\mathcal{F}_0\Big]$$

$$\begin{aligned}
&= \mathbb{E}\Big[\int_{t_k}^{t_{k+1}} \sigma_s^2 ds \int_{t_k}^{t_{k+1}} \sigma_s dZ_s^{d+1}\Big|\mathcal{F}_0\Big] \\
&= \mathbb{E}\Big[(\int_0^{t_{k+1}} \sigma_s^2 ds - \int_0^{t_k} \sigma_s^2 ds)(\int_0^{t_{k+1}} \sigma_s dZ^{d+1}s - \int_0^{t_k} \sigma_s dZ^{d+1}s)\Big|\mathcal{F}_0\Big] \\
&= \mathbb{E}\Big[\int_0^{t_{k+1}} \sigma_s^2 ds \int_0^{t_{k+1}} \sigma_s dZ^{d+1}s + \int_0^{t_k} \sigma_s^2 ds \int_0^{t_k} \sigma_s dZ^{d+1}s \\
&- \int_0^{t_{k+1}} \sigma_s^2 ds \int_0^{t_k} \sigma_s dZ^{d+1}s - \int_0^{t_{k+1}} \sigma_s dZ^{d+1}s \int_0^{t_k} \sigma_s^2 ds\Big|\mathcal{F}_0\Big] \\
&= \sum_{|I|,|J|,|M|\le n} \ell_I \ell_J \ell_M \Big(\mathbf{P}_{t_{k+1}}^{(I\sqcup J\otimes \mathbf{0})\sqcup \tilde{M}^{d+1}}(\widehat{\mathbb{Z}}_0) \ + \ \mathbf{P}_{t_k}^{(I\sqcup J\otimes \mathbf{0})\sqcup \tilde{M}^{d+1}}(\widehat{\mathbb{Z}}_0) \\
&- \ \Phi^{(I\sqcup J\otimes \mathbf{0},t_{k+1}),(\tilde{M}^{d+1},t_k)}(\widehat{\mathbb{Z}}_0) \ - \ \Phi^{(\tilde{M}^{d+1},t_{k+1}),(I\sqcup J\otimes \mathbf{0},t_k)}(\widehat{\mathbb{Z}}_0)\Big)
\end{aligned}$$

References

[1] Cuchiero,C., Gazzani,G., Ferro,S.S.(2022). Signature-Based Models: Theory and Calibration. SIAM Journal on Financial Mathematics, 14(3), 910-957. https://doi.org/10.1137/22M1512338

[2] Cuchiero,C., Gazzani,G., Moller,J., Ferro,S.S.(2023). Joint calibration to SPX and VIX options with signature-based models. arXiv. https://arxiv.org/pdf/2301.13235.pdf

[3] Allen, P., Einchcomb, S., Granger, N. (2006). *[JP Morgan] Variance Swap*. European Equity Derivatives Research, J.P. Morgan Securities Ltd., London, 17 November 2006. Retrieved from `http://quantlabs.net/academy/download/free_quant_instituitional_books_/%5BJP%20Morgan%5D%20Variance%20Swaps.pdf`

[4] Yahoo Finance. (2023). *CBOE Volatility Index (VIX) historical data*. Retrieved from `https://finance.yahoo.com/quote/^VIX/history/`

[5] Buehler, H., Gonon, L., Teichmann, J., Wood, B. (2019). Deep Hedging. Quantitative Finance, 19(8), 1271-1291. `https://arxiv.org/pdf/1802.03042.pdf`

[6] Ferguson, R., Green, A. (2019). Deep Learning Derivatives. arXiv. `https://arxiv.org/pdf/1809.02233.pdf`

[7] Chevyrev, I., Kormilitzin, A. (2016). A Primer on the Signature Method in Machine Learning. arXiv. `https://arxiv.org/pdf/1603.03788.pdf`

[8] Chen, K.-T. (1957). Integration of Paths, Geometric Invariants and a Generalized Baker-Hausdorff Formula. Annals of Mathematics, 65(1), 163–178. `https://doi.org/10.2307/1969671`

[9] Lyons, T. J. (1998). Differential equations driven by rough signals. Revista Matemática Iberoamericana, 14(2), 215-310. Retrieved from `http://eudml.org/doc/39555`

[10] Han,Z., Ding,H., Feng,Q., Han, B., (2023) Signature Methods in Variance Swap. Research Experiences for Undergraduate, Department of Mathematics, University of Michigan. Retrieved from: `https://lsa.umich.edu/content/dam/math-assets/reu-su22/reu-2023/REU_report_zyhan.pdf`