Functional Stochastic Volatility

Phillip A. Jang*
Center for Applied Mathematics, Cornell University
and
Michael Jauch†
Center for Applied Mathematics, Cornell University
and
David S. Matteson‡
Department of Statistics and Data Science, Cornell University

October 21, 2021

Abstract

This paper investigates the notion of stochastic volatility in the time series of functions. A functional analogue for stochastic volatility can be built from first setting a time-invariant basis and then modelling randomness through the vector time series of random coefficients, which follow the usual stochastic volatility model. An application on daily SPX option surfaces is used to demonstrate the value of this approach, as the empirical behavior of such data are well-characterized by such a model and the functional time series approach can naturally facilitate daily changes in observation locations. The resulting methodology provides a description of joint movements of option prices which accounts for heteroscedasticity, and hence provides a more realistic characterization of risk for option portfolios.

Keywords: Functional Time Series, Stochastic Volatility, SPX Option Surfaces, Linear Mixed Model

*Cornell University, Center for Applied Mathematics, 657 Frank H.T. Rhodes Hall, Ithaca, NY 14853
†Cornell University, Center for Applied Mathematics, 657 Frank H.T. Rhodes Hall, Ithaca, NY 14853
‡Cornell University, Department of Statistics and Data Science, 1196 Comstock Hall, Ithaca, NY 14853

1 Introduction

A *functional time series* $\{Y_t(\tau)\}$ is a time-indexed sequence ($t = 1, 2, \ldots$) of stochastic processes used to model the time dynamics of random phenomena which live on a continuum ($\tau \in \mathcal{T}$) and are best understood as random functions. The topic has become increasingly relevant as the collection of data in high resolution or frequency has become commonplace.

By modelling randomness on a whole continuum instead of a fixed or discrete set of domain points, a functional time series presents some key advantages over multivariate or vector time series. Firstly, a multivariate time series requires measurements on a regular unchanging grid and special handling of missing data. However, in real world applications the measurement locations of the data may be changing with time, sparse, or irregularly spaced. Furthermore, the parameter space of a vector time series model grows in size with respect to the dimension of the vector. By adopting a function space perspective, a functional time series bypasses these problems because its parameterization is independent of the set of measurement points.

Functional time series have found application in a wide variety of domains such as demographic forecasting (Hyndman and Booth, 2008; Shang et al., 2011; Hyndman et al., 2013; Li et al., 2020), spatiotemporal modelling (Besse et al., 2000; Ruiz-Medina et al., 2013; Cressie and Wikle, 2015; Jang and Matteson, 2018; King et al., 2018), yield curve forecasting (Kowal et al., 2017; Sen and Klüppelberg, 2019; Kowal et al., 2019), high frequency finance (Hörmann et al., 2013; Aue et al., 2017; Li et al., 2020; Huang et al., 2020), traffic flow modelling (Klepsch et al., 2017), and electricity demand forecasting (Shang, 2013; Yasmeen and Sharif, 2015; Chen and Li, 2017).

Results of Karhunen (1947) and Loève (1945) imply that functional principal component analysis (FPCA) provides a mean-square optimal finite dimensional representation of a functional time series $\{Y_t(\tau)\}$ if Y_t are i.i.d. for all t and have a defined second moment. However, when the functional observations exhibit time dependence, mean-square optimality no longer holds and FPCA on its own cannot provide a description of serial dependence. Further background on the foundations of functional data analysis and FPCA can be found

in Ramsay et al. (2005), Horváth and Kokoszka (2012), and Hsing and Eubank (2015).

Many important multivariate time series models have been generalized to a functional setting. A functional analogue of autoregressive integrated moving average (ARIMA) has been developed with theory of the functional autoregression (FAR) model studied by Bosq (2000) on Hilbert and Banach spaces. Further developments in the prediction of functional time series are in Hyndman and Shahid Ullah (2007), Bosq (2014) and Aue et al. (2015). A functional moving average (FMA) model was further studied in Aue and Klepsch (2017). Sen and Klüppelberg (2019) studied how a functional ARMA model leads to a vector ARMA structure on the principal component scores. Li et al. (2020) studied long memory in functional time series through building analogous fractional differenced ARIMA models. However, the study of time-varying volatility in functional time series is a relatively new area and this paper presents an approach to modelling heteroscedasticity in functional time series.

In time series analysis, accounting for heteroscedasticity or time-varying volatility is critical to accurate and reliable uncertainty quantification in forecasting problems. Heteroscedasticity is commonly addressed through autoregressive conditional heteroscedasticity (ARCH) models, generalized autoregressive conditional heteroscedasticity (GARCH) models, or stochastic volatility (SV) models. In ARCH/GARCH models, the conditional volatility process is a deterministic function of past data, while in SV models, the conditional volatility is driven by its own stochastic process and is not measurable with respect to past observable information.

These models are well-studied for univariate and multivariate time series. In univariate time series, the ARCH model was proposed by Engle (1982) and the GARCH model by Bollerslev (1986), while the univariate SV model was proposed by Taylor (1982) and Taylor (1986), with further investigations and comparisons by Taylor (1994), Shephard (1996), and Ghysels et al. (1996) for example. Multivariate versions of GARCH were examined in Bollerslev et al. (1988) and Bollerslev (1990) while multivariate SV was examined in Harvey et al. (1994), Daniélsson (1998), and Asai et al. (2006). Bayesian estimation of multivariate GARCH and SV were examined in Vrontos et al. (2003) and Yu and Meyer

(2006) respectively.

In recent developments of heteroscedastic functional time series models, Hörmann and Kokoszka (2010) formalized the notion of weakly dependent functional time series, while Hörmann et al. (2014) addresseed serial correlation with a dynamic Karhunen-Loève expansion, a functional analogue of the dynamic principal component analysis for multivariate time series in Brillinger (1981). This approach eliminates cross-correlations in principal component scores at all leads and lags. Jang et al. (2017) developed Bayesian inference for stationary covariance functions using Lévy process priors. Hörmann et al. (2013) and Aue et al. (2017) approached heteroscedasticity with functional autoregressive conditional heteroscedasticity (fARCH) and functional generalized autoregressive conditional heteroscedasticity (fGARCH) models respectively.

This paper introduces the notion of stochastic volatility to the functional time series setting and demonstrates through an application where such a model could be of interest. As SV models lack a closed-form likelihood, they have found relatively less popularity in empirical applications compared to ARCH/GARCH type models. However, the randomness in the volatility process of an SV model offers a higher degree of flexibility and allows them to fit to data as well as more heavily parameterized GARCH models (Daniélsson, 1998; Kim et al., 1998). SV models also provide a more natural discrete analogue to the stochastic differential equations used in option pricing such as the stochastic volatility diffusion of Hull and White (1987).

The challenge of SV model estimation is effectively addressed through the Bayesian approach pioneered by Jacquier et al. (1994), which recasts the SV model into linear state space form and then employs the Kalman filter. This approach was refined in the mixture sampler of Kim et al. (1998) and the Ancillary-Sufficiency Interweaving Strategy of Kastner and Frühwirth-Schnatter (2014). These methods lead to highly efficient Bayesian inference for the SV model and can be readily adapted to more complex hierarchical models.

Kowal et al. (2017) and Kowal et al. (2019) introduced a Bayesian framework for functional time series with a Bayesian hierarchical specification, and this paper adds one more level to

the hierarchical model of Kowal et al. (2017) to build a fully Bayesian model for functional stochastic volatility.

As stochastic volatility is ubiquitous in the empirical movements of financial prices and returns (Hull and White, 1987; Heston, 1993; Heston and Nandi, 2000; Gatheral, 2006; Shephard and Andersen, 2009), we apply the functional stochastic volatility model in a financial setting, modelling the daily movements of SPX option surfaces. For each day in the time period from January 1, 2010 to December 31, 2017 we observe hundreds to thousands of option contracts (differentiated by their strike price and time to maturity) as a set of discrete points on an underlying continuous option surface. The empirical movements of the principal component scores of these surfaces clearly exhibit stochastic volatility in an exploratory analysis. By accounting for the empirically observed stochastic volatility in SPX option surfaces, we show through an experiment that a functional stochastic volatility model is capable of accurately estimating the Value-at-Risk of SPX option portfolios.

We now outline the remainder of the paper. Section 2 describes the functional stochastic volatility model and justifies consideration of a finite-dimensional form. Section 3 addresses Bayesian inference for the FSV model. Section 4 presents an application of the FSV model to SPX option portfolio risk management.

2 Methodology

We will now formulate the functional stochastic volatility model. In Subsection 2.1 we define a functional stochastic volatility model. Subsection 2.2 will set up the notation of a functional time series and describe the functional analogue of the ARMA model. Then in Subsection 2.3 we describe the reduction of the functional specification to a finite-dimensional vector specification suited for inference. Lastly, Subsection 2.4 describes the basis functions used for dimension reduction.

2.1 Functional Stochastic Volatility

Let $\{\eta_t\}$ be a sequence of random functions in $L^2(\mathcal{T})$ for some compact domain $\mathcal{T} \subseteq \mathbb{R}^d$. Suppose $\{\eta_t\}$ consists of the following heteroscedastic innovations

$$\eta_t(\tau) = v_t(\tau) z_t(\tau), \qquad\qquad v_t \geq 0,$$

where $\{z_t\}$ is a sequence of independent and identically distributed mean zero random functions on $\mathcal{T}$ with known covariance function $Cov(z_t(\tau), z_t(u)) = k_z(\tau, u)\ \forall t$, $k_z(\tau, \tau) = 1\ \forall \tau$, and $\{z_t\}$ is independent of volatility process $\{v_t\}$.

Then $\{\eta_t\}$ is a *functional stochastic volatility* model if for any t, v_t is a random function on $\mathcal{T}$ and is not measurable with respect to the sigma field generated by $\{v_s\}_{s=1}^{t-1}$ and $\{\eta_s\}_{s=1}^{t-1}$.

2.2 Functional ARMA

Assume the data arise from an underlying sequence of time-ordered mean-zero real-valued random functions $\{Y_t\}$ in $L^2(\mathcal{T})$ for some compact domain $\mathcal{T} \subseteq \mathbb{R}^d$. We will suppose for each time t that Y_t is observed at $n_t \geq 0$ locations with measurement error as $y_{t,i} = m(\tau_{t,i}) + Y_t(\tau_{t,i}) + \varepsilon_{t,i}$ where $\tau_{t,1}, \ldots, \tau_{t,n_t} \in \mathcal{T}$ are the observation locations, $m(\tau)$ is the mean function, and $\varepsilon_{t,i} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$ is the observation error.

Then we define a functional ARMA(p,q) model for $\{Y_t\}$ with the following form:

$$Y_t(\tau) = \eta_t(\tau) + \sum_{i=1}^{p} \int_{\mathcal{T}} \psi_i(\tau, u) Y_{t-i}(u) du - \sum_{j=1}^{q} \int_{\mathcal{T}} \theta_j(\tau, u) \eta_{t-j}(u) du$$

where $\{\eta_t\}$ are mean zero random functions with $\mathbb{E}\left\|\eta_t\right\|^2 \equiv \mathbb{E}\left[\int_{\mathcal{T}} \eta_t(\tau)^2 d\tau\right] < \infty$ and $\mathbb{E}[\eta_t(\tau)\eta_s(\tau)] = 0\ \forall t \neq s$ and $\forall \tau \in \mathcal{T}$ so that $\{\eta_t\}$ are uncorrelated in t but not necessarily independent. These conditions are satisfied, for example, by the functional stochastic volatility model in Section 2.1. The parameter kernel functions $\{\psi_i\}$ and $\{\theta_j\}$ are analogous to the parameter matrices in an multivariate ARMA model and control time dependence in the conditional mean. Note that the functional ARMA model can be defined more generally

with bounded linear operators on Hilbert spaces (see Bosq (2000)). For this paper we will work with the functional AR(1) process and drop the index on the integral kernel function ψ_1:

$$Y_t(\tau) = \int_{\mathcal{T}} \psi(\tau, u) Y_{t-1}(u) du + \eta_t(\tau)$$

2.3 Reducing a Functional Time Series to a Vector Time Series

The functional ARMA model becomes more tractable if we make certain assumptions. In particular, we assume $\mathcal{T} \subset \mathbb{R}^d$ is compact, $f_1, \ldots, f_K \in L^2(\mathcal{T})$ are orthonormal deterministic functions, and that

1. $\{Y_t\} \subseteq \text{Span}\,(f_1, \ldots, f_K)$, so there exist random coefficient vectors $\boldsymbol{\beta}_t = (\beta_{t1}, \ldots, \beta_{tK})^\top$ and $\boldsymbol{\gamma}_t = (\gamma_{t1}, \ldots, \gamma_{tK})^\top$ such that

$$Y_t = \sum_{k=1}^{K} \beta_{tk} f_k, \quad \text{and} \quad \eta_t = \sum_{k=1}^{K} \gamma_{tk} f_k.$$

   where $\beta_{tk} =< Y_t, f_k >$ and $\gamma_{tk} =< \eta_t, f_k >$. Since f_k's are orthogonal, for any fixed t the entries of $\boldsymbol{\beta}_t$ are uncorrelated, as are the entries of $\boldsymbol{\gamma}_t$.

2. $\psi(\tau, u) \in \text{Span}\,(f_1, \ldots, f_K) \otimes \text{Span}\,(f_1, \ldots, f_K)$, so there exists a $K \times K$ coefficient matrix $\Psi = [\psi_{kj}]$ such that

$$\psi(\tau, u) = \sum_{k=1}^{K} \sum_{j=1}^{K} \psi_{kj} f_k(\tau) f_j(u)$$

   where $\psi_{kj} =<< \psi(\star, \cdot), f_j(\cdot) >, f_k(\star) >= \int_{\mathcal{T}} \int_{\mathcal{T}} \psi(\tau, u) f_j(u) f_k(\tau) du d\tau$.

3. The coefficient vectors $\boldsymbol{\gamma}_t$ are a time-dependent sequence of random vectors with

$$\boldsymbol{\gamma}_t | \mathcal{F}_{t-1} \sim N_K(\mathbf{0}, \Sigma_{\gamma_t})$$

   and the conditional covariance function of $\eta_t(\tau)$ is defined and satisfies

$$Cov(\eta_t(\tau), \eta_t(u) | \mathcal{F}_{t-1}) \equiv k_{\eta_t}(\tau, u) = \sum_{k=1}^{K} e^{h_{tk}} f_k(\tau) f_k(u)$$

   where $\mathcal{F}_t$ is the sigma field generated by $\{\boldsymbol{\gamma}_s\}_{s=1}^t$ and $\{\boldsymbol{h}_s\}_{s=1}^t$ with $\boldsymbol{h}_t = (h_{t1}, \ldots, h_{tK})^\top$.

Under assumptions 1-3 above, the coefficients $\{\boldsymbol{\beta}_t\}_{t=1}^T$ form a K-dimensional VAR(1) process

$$\begin{aligned}\boldsymbol{\beta}_t &= \Psi\boldsymbol{\beta}_{t-1} + \boldsymbol{\gamma}_t \\ \boldsymbol{\gamma}_t|\mathcal{F}_{t-1} &\sim N_K(\mathbf{0}, \Sigma_{\gamma_t})\end{aligned}$$

where Σ_{γ_t} is assumed diagonal and equal to $\text{diag}\{e^{\boldsymbol{h}_t}\}$. $\{\boldsymbol{\beta}_t\}_{t=1}^T$ is a VAR(1) process with stochastic volatility as the conditional covariance matrix of $\boldsymbol{\gamma}_t$ depends on the random vector $\boldsymbol{h}_t$ and is hence not $\mathcal{F}_{t-1}$-measurable.

In this paper, we use VAR(1) dynamics for $\boldsymbol{h}_t$ which is well-suited for capturing a volatility clustering effect.

$$\begin{aligned}\boldsymbol{h}_t &= \boldsymbol{\mu} + \Phi(\boldsymbol{h}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\eta}_t \\ \boldsymbol{\eta}_t &\overset{iid}{\sim} N_K(\mathbf{0}, \Sigma_\eta)\end{aligned}$$

For parsimony we will assume the matrices Φ and Σ_η are diagonal and that each $\{h_{tk}\}_{t=1}^T$ is independent of all other $\{h_{tj}\}_{t=1}^T$'s for any $j \neq k$. Hence for K-dimensional parameter vectors $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_\eta$,

$$\begin{aligned}\boldsymbol{h}_t &= \boldsymbol{\mu} + \boldsymbol{\phi} \cdot (\boldsymbol{h}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\eta}_t \\ \boldsymbol{\eta}_t &\overset{iid}{\sim} N_K(\mathbf{0}, \text{diag}\{\boldsymbol{\sigma}_\eta^2\})\end{aligned}$$

Letting $\boldsymbol{y}_t = [y_{t,1}, \ldots, y_{t,n_t}]^\top$, $\boldsymbol{m}_t = [m(\tau_{t,1}), \ldots, m(\tau_{t,n_t})]^\top$, $[F_t]_{i,k} = f_k(\tau_{t,i})$, $\boldsymbol{\beta}_t = [\beta_{t1}, \ldots, \beta_{tK}]^\top$, $\boldsymbol{\varepsilon}_t = [\varepsilon_{t,1}, \ldots, \varepsilon_{t,n_t}]^\top$, $[\Psi]_{k,j} = \psi_{kj}$, $\boldsymbol{h}_t = [h_{t1}, \ldots, h_{tK}]^\top$, $\boldsymbol{\gamma}_t = [\gamma_{t1}, \ldots, \gamma_{tK}]^\top$, $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_K]^\top$, $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_K]^\top$, $\boldsymbol{\sigma}_\eta = [\sigma_1, \ldots, \sigma_K]^\top$, and $\boldsymbol{\eta}_t = [\eta_{t1}, \ldots, \eta_{tK}]^\top$ we can succinctly summarize our hierarchical model as follows:

$$\boldsymbol{y}_t = \boldsymbol{m}_t + F_t\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim N_{n_t}(\mathbf{0}, \sigma_\varepsilon^2 I) \qquad \forall t \geq 1, \quad (1)$$

$$\boldsymbol{\beta}_t = \Psi\boldsymbol{\beta}_{t-1} + \boldsymbol{\gamma}_t, \qquad \boldsymbol{\gamma}_t|\boldsymbol{h}_t \sim N_K(\mathbf{0}, \mathrm{diag}\{e^{\boldsymbol{h}_t}\}) \qquad \forall t \geq 1, \quad (2)$$

$$\boldsymbol{h}_t = \boldsymbol{\mu} + \boldsymbol{\phi} \cdot (\boldsymbol{h}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N_K(\mathbf{0}, \mathrm{diag}\{\boldsymbol{\sigma}_\eta^2\}) \qquad \forall t > 1, \quad (3)$$

$$\boldsymbol{h}_1 \sim N_K\left(\boldsymbol{\mu}, \mathrm{diag}\left\{\frac{\boldsymbol{\sigma}_\eta^2}{1-\boldsymbol{\phi}}\right\}\right) \quad (4)$$

where

1. $\{\boldsymbol{\varepsilon}_t\}$ and $\{\boldsymbol{\eta}_t\}$ are independent across t and with each other.
2. $\{\boldsymbol{\varepsilon}_t\}$ is independent of $\{\boldsymbol{\gamma}_t\}$.
3. $\boldsymbol{\gamma}_t$ is conditionally independent of $\{\boldsymbol{\gamma}_s\}_{s=1}^{t-1}$ given $\boldsymbol{h}_t$.
4. $\boldsymbol{h}_1$ is independent of everything else.

2.4 Basis Function Selection

A functional time series represents an infinite-dimensional random process, but in practical implementations we seek a finite-dimensional representation which under suitable regularity conditions can best approximate the underlying process with the least information loss. In our applications we use the first K functional principal components for $f_1, \ldots, f_K$ and estimate them from the spline-smoothed functional data. Refer to Appendix Section 6.1 for further justification.

3 Inference

This section describes Bayesian inference for the functional stochastic volatility model. Subsection 3.1 briefly introduces Gibbs sampling, while Subsection 3.2 presents the prior specification and resulting full conditional distributions.

3.1 Markov Chain Monte Carlo and Gibbs Sampling

Suppose our observed data $y_1, \ldots, y_n$ are sampled from some joint density $p(y_1, \ldots, y_n|\boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \Theta$ represents the parameters of the model. Let $p(\boldsymbol{\theta})$ be the prior density on $\boldsymbol{\theta}$. The posterior density after observing $y_1, \ldots, y_n$ is given by

$$p(\boldsymbol{\theta}|y_1, \ldots, y_n) = \frac{p(y_1, \ldots, y_n|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(y_1, \ldots, y_n|\boldsymbol{\theta'})p(\boldsymbol{\theta'})d\boldsymbol{\theta'}} \propto p(y_1, \ldots, y_n|\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

For many complex models, the posterior density cannot be evaluated analytically because the integral in the denominator is intractable. To address this problem, Markov Chain Monte Carlo (MCMC) methods construct a Markov chain on the parameter space that has the posterior as its stationary distribution. The sample path of the Markov chain can then be used to approximate posterior summaries, functionals, or other quantities of interest.

The Gibbs sampling algorithm proceeds by partitioning the parameter space into disjoint blocks such that $\Theta = \Theta_1 \times \ldots \times \Theta_M$ and then iteratively sampling from the full conditional distributions $p(\boldsymbol{\theta}_1|\{y_i\}_{i=1}^n, \{\boldsymbol{\theta}_m\}_{m\neq 1}), \ldots, p(\boldsymbol{\theta}_M|\{y_i\}_{i=1}^n, \{\boldsymbol{\theta}_m\}_{m\neq M})$ in sequential order. An advantage of the Gibbs sampling approach is that the full conditional distributions are often analytically tractable even when the joint posterior distribution is not. Also the Gibbs sampling algorithm is modular, allowing one to plug in existing techniques to sample from the full conditional distributions when they are available. Refer to Gelman et al. (2013) for an overview of MCMC, Gibbs sampling, and Bayesian data analysis.

3.2 Prior Specification and Full Conditional Distributions

This section presents the prior specification for the functional stochastic volatility model and the resulting full conditional distributions. In order to write these succinctly, we introduce matrix variate notation to simplify Equations 1-4 in Section 2.3.

From the terms in Equations 2, 3 and 4 define the following KT-dimension random vectors $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_T^\top]^\top$, $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_T^\top]$, and $\boldsymbol{h} = [\boldsymbol{h}_1^\top, \ldots, \boldsymbol{h}_T^\top]$. Then the random vectors

satisfy

$$P\boldsymbol{\beta} = \boldsymbol{\gamma} \tag{5}$$

where

$$P = \begin{bmatrix} I_K & 0_K & 0_K & ... & 0_K \\ -\Psi & I_K & 0_K & ... & 0_K \\ 0_K & -\Psi & I_K & ... & 0_K \\ ... & ... & ... & \ddots & ... \\ 0_K & 0_K & 0_K & ... & I_K \end{bmatrix}.$$

From the terms in Equation 1 with $n_y \equiv \sum_{t=1}^{T} n_t$, define the n_y-dimensional random vectors $\boldsymbol{y} = [\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_T^\top]^\top$, $\boldsymbol{m} = [\boldsymbol{m}_1^\top, \ldots, \boldsymbol{m}_T^\top]^\top$ and $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_1^\top, \ldots, \boldsymbol{\varepsilon}_T^\top]^\top$ and the block diagonal matrix $F = \text{diag}\{F_1, \ldots, F_T\}$. Then

$$\boldsymbol{y} = \boldsymbol{m} + F\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Writing the SV parameters as $\boldsymbol{\theta}_h = (\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}_\eta)$, it follows from Equation 1 that

$$\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{h}, \boldsymbol{\theta}_h, \Psi, \sigma_\varepsilon^2 \sim N_{n_y}(\boldsymbol{m} + F\boldsymbol{\beta}, \sigma_\varepsilon^2 I),$$

so the likelihood function is

$$p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{h}, \boldsymbol{\theta}_h, \Psi, \sigma_\varepsilon^2) = (2\pi\sigma_\varepsilon^2)^{-n_y/2} \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{m} - F\boldsymbol{\beta}\|^2}{2\sigma_\varepsilon^2}\right).$$

For the log-variance process $\boldsymbol{h}$ and its associated SV parameters $\boldsymbol{\theta}_h$, the priors are assigned as in Kim et al. (1998) and Kastner and Frühwirth-Schnatter (2014). More specifically, the conditional prior of $\boldsymbol{h}$ given $\boldsymbol{\theta}_h$ is

$$p(\boldsymbol{h}|\boldsymbol{\theta}_h) = p(\boldsymbol{h}_1|\boldsymbol{\theta}_h) \prod_{t=2}^{T} p(\boldsymbol{h}_t|\boldsymbol{h}_{t-1}, \boldsymbol{\theta}_h)$$

$$\boldsymbol{h}_t|\boldsymbol{h}_{t-1}, \boldsymbol{\theta}_h \sim N_K\left(\boldsymbol{\mu} + \boldsymbol{\phi} \cdot (\boldsymbol{h}_{t-1} - \boldsymbol{\mu}), \text{diag}\{\boldsymbol{\sigma}_\eta^2\}\right)$$

$$\boldsymbol{h}_1|\boldsymbol{\theta}_h \sim N_K\left(\boldsymbol{\mu}, \text{diag}\left\{\frac{\boldsymbol{\sigma}_\eta^2}{1 - \boldsymbol{\phi}}\right\}\right)$$

and the priors of the SV parameters $\boldsymbol{\theta}_h$ are

$$
\begin{aligned}
p(\boldsymbol{\theta}_h) &= \prod_{k=1}^{K} p(\mu_k)p(\phi_k)p(\sigma_k^2) \\
\mu_k &\sim N(b_\mu, B_\mu) && \forall k \\
(1+\phi_k)/2 &\sim \text{Beta}(a_\phi, b_\phi) && \forall k \\
\sigma_k^2 &\sim B_\sigma \cdot \chi_1^2 \equiv B_\sigma \cdot \text{Gamma}\left(\frac{1}{2}, \frac{1}{2B_\sigma}\right) && \forall k
\end{aligned}
$$

For the remaining parameters, we assign the following priors which yield closed form full conditionals:

$$
\begin{aligned}
\boldsymbol{\beta}|\boldsymbol{h}, \Psi &\sim N_{TK}(\mathbf{0}_{TK}, P^{-1}\text{diag}\{\exp(\boldsymbol{h})\}P^{-\top}), \\
\Psi &\sim \text{Matrix Normal}_{K,K}(M, U, V), \\
\sigma_\varepsilon^2 &\sim \text{InvGamma}(a_\varepsilon, b_\varepsilon),
\end{aligned}
$$

where Matrix $\text{Normal}_{K,K}(M, U, V)$ indicates a $K \times K$ matrix normal distribution with mean matrix M and scale matrices U and V. The density function is shown in Appendix Section 6.3. $\text{InvGamma}(a_\varepsilon, b_\varepsilon)$ indicates an inverse gamma distribution with shape parameter a_ε and scale parameter b_ε. The full set of prior hyperparameters is $(b_\mu, B_\mu, a_\phi, b_\phi, M, U, V, a_\varepsilon, b_\varepsilon, B_\sigma)$

Hence all full conditionals will be proportional to the joint density of $(\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{h}, \boldsymbol{\theta}_h, \Psi, \sigma_\varepsilon^2)$, defined by the following product:

$$
p(\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{h}, \boldsymbol{\theta}_h, \Psi, \sigma_\varepsilon^2) = p(\boldsymbol{y}|\boldsymbol{\beta}, \sigma_\varepsilon^2)p(\boldsymbol{\beta}|\boldsymbol{h}, \Psi)p(\boldsymbol{h}|\boldsymbol{\theta}_h)p(\boldsymbol{\theta}_h)p(\Psi)p(\sigma_\varepsilon^2).
$$

The full conditionals of $\boldsymbol{h}$ and $\boldsymbol{\theta}_h$ and their samplers are derived in Kim et al. (1998) and Kastner and Frühwirth-Schnatter (2014). After Gibbs updates to $\boldsymbol{\beta}$ and Ψ, $\boldsymbol{\gamma}$ is recalculated through Equation 5. As each univariate series $\{\gamma_{tk}\}_{t=1}^T$ is assumed independent of the others for different k, each $\{\gamma_{tk}\}_{t=1}^T$ is a univariate time series with an associated stochastic volatility process $\{h_{tk}\}_{t=1}^T$ as in Kim et al. (1998), where Bayesian inference of $\{h_{tk}\}_{t=1}^T$ and its three associated parameters $\{\mu_k, \phi_k, \sigma_k\}$ is achieved through a mixture sampler applied to $\{\gamma_{tk}\}_{t=1}^T$. The R package `stochvol` developed by Kastner (2016) employs this

mixture sampler and uses the Ancillarity-Sufficiency Interweaving Strategy in Kastner and Frühwirth-Schnatter (2014) for highly efficient Bayesian inference.

The remaining full conditional distributions are as follows:

$$\begin{aligned}
\boldsymbol{\beta}|\boldsymbol{y},\boldsymbol{h},\boldsymbol{\theta}_h,\Psi,\sigma_\varepsilon^2 &\sim N_{TK}\left(\boldsymbol{\mu}_\beta,\sigma_\varepsilon^2\Lambda_\beta^{-1}\right),\\
\sigma_\varepsilon^2|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{h},\boldsymbol{\theta}_h,\Psi &\sim \text{InvGamma}\left(a_\varepsilon+\frac{n_y}{2},b_\varepsilon+\frac{1}{2}\|\boldsymbol{y}-\boldsymbol{m}-F\boldsymbol{\beta}\|^2\right),\\
vec(\Psi)|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{h},\sigma_\varepsilon^2 &\sim N_{K^2}(\boldsymbol{\mu}_\Psi,\Sigma_\Psi),
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\mu}_\beta &= \Lambda_\beta^{-1}F^\top(\boldsymbol{y}-\boldsymbol{m}),\\
\Lambda_\beta &= F^\top F+\sigma_\varepsilon^2 P^\top\text{diag}\{\exp(-\boldsymbol{h})\}P,\\
\Sigma_t &= \text{diag}\{\exp(\boldsymbol{h}_t)\},\\
\boldsymbol{\mu}_\Psi &= \Sigma_\Psi vec\left(\sum_{t=2}^{T}\Sigma_t^{-1}\boldsymbol{\beta}_t\boldsymbol{\beta}_{t-1}^\top+U^{-1}MV^{-1}\right),\\
\Sigma_\Psi &= \left[\sum_{t=2}^{T}(\boldsymbol{\beta}_{t-1}\boldsymbol{\beta}_{t-1}^\top\otimes\Sigma_t^{-1})+(V^{-1}\otimes U^{-1})\right]^{-1}.
\end{aligned}$$

The full conditionals for $\boldsymbol{\beta}$ and σ_ε^2 follow from the well-known Bayesian linear regression with the Normal-Inverse-Gamma conjugate prior, and a derivation can be found in Gelman et al. (2013) for example. Though the dimension of $\boldsymbol{\beta}$ is potentially very large, the block structure of its precision matrix can be exploited for efficient Monte Carlo sampling. Refer to Appendix Section 6.2 for how to sample the full conditional for $\boldsymbol{\beta}$. Refer to Appendix Section 6.3 for derivation of the full conditional for Ψ.

4 Value-at-Risk Estimation for Option Portfolios

In this section we apply the functional stochastic volatility model to SPX option surface data in order to estimate Value-at-Risk for option portfolios. The results indicate that the

stochastic volatility aspect can improve quantile estimates in this setting. Subsection 4.1 provides an overview of the application. Subsection 4.2 describes in detail the SPX option surface data set and its representation as a functional time series $\{Y_t(\tau)\}$. Subsection 4.3 presents a justification for the application of a functional stochastic volatility model through an exploratory analysis and visualizes the basis functions chosen for the application. Subsection 4.4 discusses the construction of a forecast distribution for $Y_t(\tau)$ from posterior samples and describes how the forecast distribution is used to estimate quantiles. Subsection 4.5 evaluates the quality of the forecast against a benchmark constant volatility model.

4.1 Overview of Application

The field of financial risk management is concerned with the estimation of worst case outcomes (rather than mean outcomes) in order to quantify the magnitude of potential losses due to adverse movements in market prices or other risk factors. Consequently, the problem of quantile estimation is of key importance. Stochastic volatility models are often employed in the field because they better represent the empirical movements of financial time series, and as a result, improve upon quantile estimates for univariate and multivariate time series (Sadorsky, 2005; Han et al., 2014; Huang, 2015; Bui Quang et al., 2018). This application tests an analogous problem for a functional time series.

In the present application, the functional time series process $\{Y_t(\tau)\}_{t=1}^{T}$ represents a surface of option price quotes and is modelled with the proposed functional stochastic volatility model. An option gives the owner the choice to buy (if a call option) or sell (if a put option) an underlying asset S_t at some fixed strike price K at some future maturity date $u > t$. For any single underlying asset (such as the S&P 500 Index) there exists an entire surface of options for that asset, because options can differ by their strike price K and maturity date u. These two variables define the dimensions of the domain variable $\tau \in \mathcal{T} \subseteq \mathbb{R}^2$. The range variable $Y_t(\tau)$ can be thought of as a proxy for the price of the option with contract terms τ, as there is a one-to-one relationship between $Y_t(\tau)$ and the option price. These

variables are described in full detail in Subsection 4.2.

Consider a loss random variable $L_{t+1} = \Pi_t(Y_{t+1})$ for some functional $\Pi_t : L^2(\mathcal{T}) \to \mathbb{R}$. The loss random variable L_{t+1} represents the decrease in value of a portfolio of options from time t to time $t+1$. The functional Π_t is determined by the composition of the option portfolio held at time t. L_{t+1} and Π_t are constructed explicitly in Subsection 4.4.

The next-day $(1-\alpha)100\%$ Value-at-Risk for L_{t+1} is defined as

$$\begin{aligned} \mathrm{VaR}(L_{t+1}, 1-\alpha) &= \sup\{x \in \mathbb{R} : \mathbb{P}(L_{t+1} \leq x) < 1-\alpha\} && (6) \\ &= F^{-1}_{L_{t+1}}(1-\alpha) \text{ if } L_{t+1} \text{ is a continuous random variable.} && (7) \end{aligned}$$

In other words, Value-at-Risk is the quantile function for the loss random variable L_{t+1}. Estimating Value-at-Risk is a key problem in financial risk management as it identifies the potential magnitude of loss L_t due to the random market price movements of Y_t.

A natural approach for evaluating the quality of an estimator $\widehat{\mathrm{VaR}}(L_{t+1}, 1-\alpha)$ is to compare the proportion of observed exceedences of the quantile estimate to the true level of the quantile. Define the exceedence indicator variables $X_t = \mathbb{1}_{\left\{L_{t+1} > \widehat{\mathrm{VaR}}(L_{t+1}, 1-\alpha)\right\}}$ for $t = 1, \ldots, T-1$ and set $N_{T-1} = \sum_{t=1}^{T-1} X_t$. If we assume the X_t variables are independent and identically distributed, then $N_{T-1} \equiv \sum_{t=1}^{T-1} X_t \sim \mathrm{Bin}(T-1, \alpha_N)$. We can perform a Binomial hypothesis test of $H_0 : \alpha_N = \alpha$ against $H_1 : \alpha_N \neq \alpha$ in order to evaluate the whether the estimator $\widehat{\mathrm{VaR}}(L_{t+1}, 1-\alpha)$ has the correct exceedence rate. This is done on a year-by-year basis in Subsection 4.5.

4.2 SPX Option Surface Data Set as a Functional Time Series

This section describes the SPX option data set and its representation as a functional time series. It will provide full descriptions of the domain and range variables, and motivate the use of a functional time series model.

A European option is a financial contract that, at some fixed maturity date in the future, gives the owner the option to purchase (if a call option) or sell (if a put option) a unit

of an underlying asset at a fixed strike price, regardless of the actual market price of said asset. In the case of an SPX option, the underlying asset is the Standard & Poors (S&P) 500 Index.

Our methodology is applied to daily SPX option surfaces sourced from OptionMetrics's IvyDB database (2017) from January 1, 2010 to December 31, 2017. There are $T = 2013$ days in total. Each daily option surface includes between 1400 and 10000 option contracts, which are differentiated by their strike price, date of maturity, and whether they are call options or put options. Option prices are quoted based on their Black-Scholes implied volatility, the log of which is the variable of interest $Y_t(\tau)$. There is a one-to-one relationship between the option price and $Y_t(\tau)$ (see Equation 9). The unfamiliar reader can regard $Y_t(\tau)$ as a proxy for option price.

The domain variable $\tau = (\tau^1, \tau^2)$ is two-dimensional with $\tau^1 \in [1, \sqrt{1095}]$ being the square root of number of days to maturity, and $\tau^2 \in [0, 1]$ being the call option delta. τ^1 is set as such because the implied volatility of an option is a function of its remaining time to maturity instead of the preset date of maturity. τ^2 is set as such because implied volatility depends on the ratio of the underlying asset price to the strike price (or equivalently, the option delta) instead of the absolute strike price. The functional domain $\mathcal{T}$ is the rectangle determined by the two univariate domains

$$\tau \equiv (\tau^1, \tau^2) \qquad \in \qquad \mathcal{T} \equiv [1, \sqrt{1095}] \times [0, 1].$$

We include both call and put options in our analysis. They share the same implied volatility if their strike price and maturity match, but their option deltas differ by a deterministic relationship. In order to keep domain variables equivalent for the two types of options, we convert a put option's delta to its equivalent call option delta by adding $\exp\{-\text{Dividend Yield} \times \text{Time To Maturity}\}$ to its value. Refer to Hull (2018) for an introduction to option pricing. Since illiquid option prices have high measurement error, options whose bid-ask spread is larger than 10% of the premium are excluded from the application.

On any given day t, the observation points $\tau_{t,1}, \ldots, \tau_{t,n_t} \in \mathcal{T}$ are determined by the options traded on the market that day. The set of observation points changes every day, motivating

the use of a functional time series model for the SPX option data set. The observation set changes each day because the option contracts in the market are issued with a fixed date of maturity. Hence, the remaining time to maturity τ^1 decreases with each passing day and all observation points shift towards the short maturity end of the domain with observations dropping out the domain when contracts mature ($\tau^1 = 0$). Furthermore, new observations enter the domain whenever new sets of option contracts are issued. Lastly, movements in the underlying S&P500 Index change the delta τ^2 of all options simultaneously, resulting in daily lateral shifts in observation points.

4.3 Basis Function Estimation and Exploratory Analysis

This section motivates the use of a functional stochastic volatility model. We estimate the functional basis $\{f_k\}_{k=1}^K$ that will be passed into subsequent Bayesian analysis using FPCA, and present an initial exploratory analysis. The sample time courses and principal component scores that suggest the functional time series exhibits stochastic volatility.

Based on empirical quantiles of the observation points, the following knot sequences were used in each dimension of the domain to define the cubic tensor splines:

$$\mathcal{K}^1 = [1.4, 1.4, 1.4, 1.4, 8.1, 12.7, 19.3, 33.1, 33.1, 33.1, 33.1], \text{ and}$$
$$\mathcal{K}^2 = [0.10, 0.10, 0.10, 0.10, 0.37, 0.57, 0.71, 0.79, 0.86, 0.90, 0.90, 0.90, 0.90].$$

The R package `mgcv` is used to perform penalized spline smoothing in order estimate the mean function $m(\tau)$ and functional principal components $\{f_1, \ldots, f_K\}$. These are shown in Figure 1.

The mean function $m(\tau)$ demonstrates the characteristic "volatility skew" endemic to equity option markets. The Black-Scholes model assumes a Gaussian distribution for market returns. If such an assumption were true, the observed volatility surface would be flat. However, we see that the S&P500 returns exhibit both heavy tails and a negative skewness, leading to convexity and skewness in the mean function respectively.

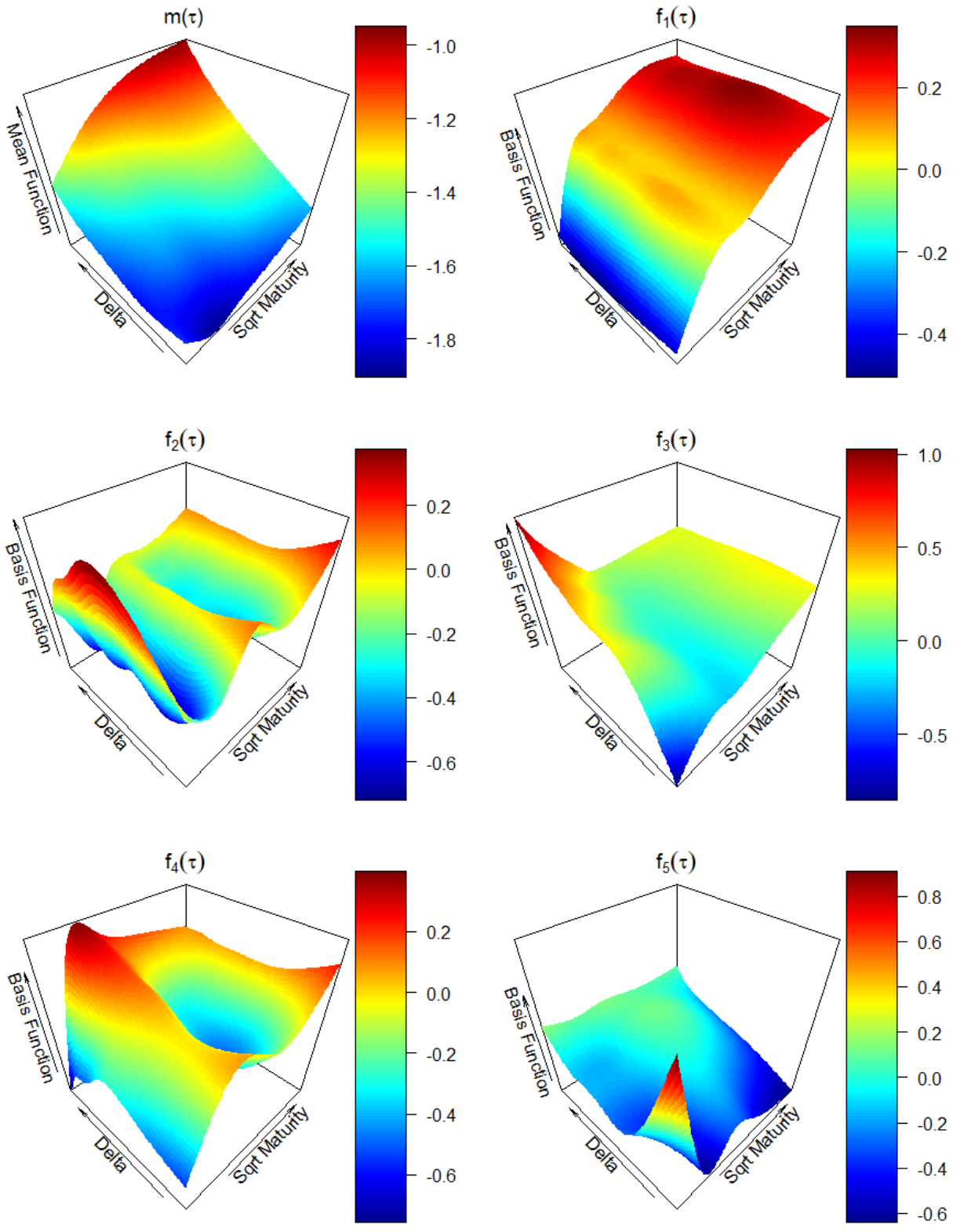

Figure 1: Mean Function $m(\tau)$ and Functional Principal Components $f_1(\tau), \ldots, f_5(\tau)$

The first principal component $f_1(\tau)$ is negative in the short end while positive on the medium to long end of the surface, resulting in a tilt of the surface. The second principal component $f_2(\tau)$ results in a bending effect with respect to maturity. The third principal component $f_3(\tau)$ tilts the short end of the surface. The fourth principal component $f_4(\tau)$ has a similar bending effect as the second but pushes the edges upward in general. The fifth principal component $f_5(\tau)$ pushes up one of the corners. As the amount of stochastic volatility observed in the previous section varies year-by-year, the FSV model is fit separately for each of the 8 years of data.

In an exploratory analysis, we see evidence for stochastic volatility. For five fixed sample domain points $\tau_1, \ldots, \tau_5$, their differenced time courses $\{Y_{t+1}(\tau_j) - Y_t(\tau_j)\}_{t=1}^{T-1}$ are shown in Figure 2. The time-differenced principal component scores shown in Figure 3. Both sets of series exhibit stochastic volatility and motivate the application of our model.

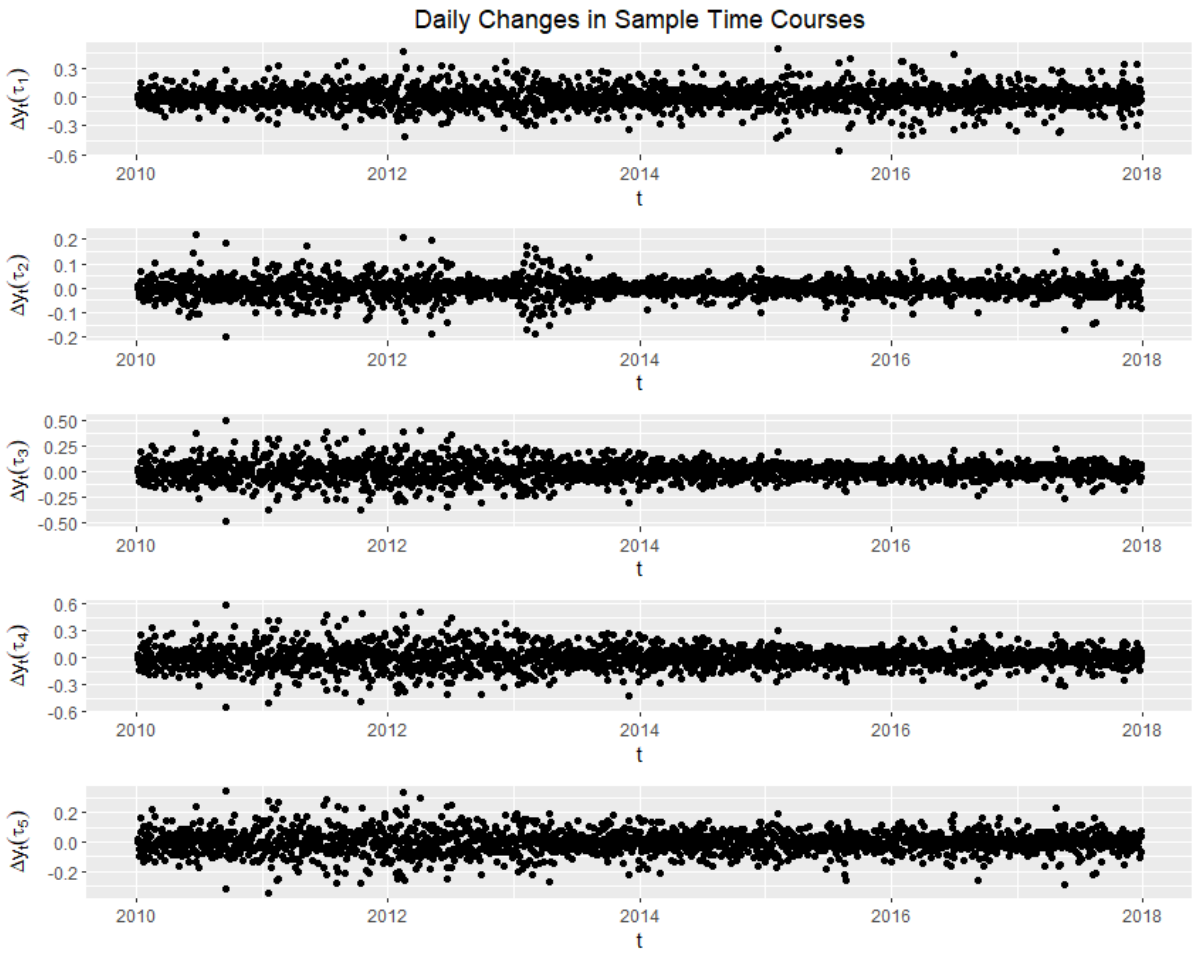

Figure 2: Differenced time courses $\{Y_{t+1}(\tau_j) - Y_t(\tau_j)\}_{t=1}^{T-1}$ for five sample locations $\tau_1, \ldots, \tau_5$, which exhibit stochastic volatility.

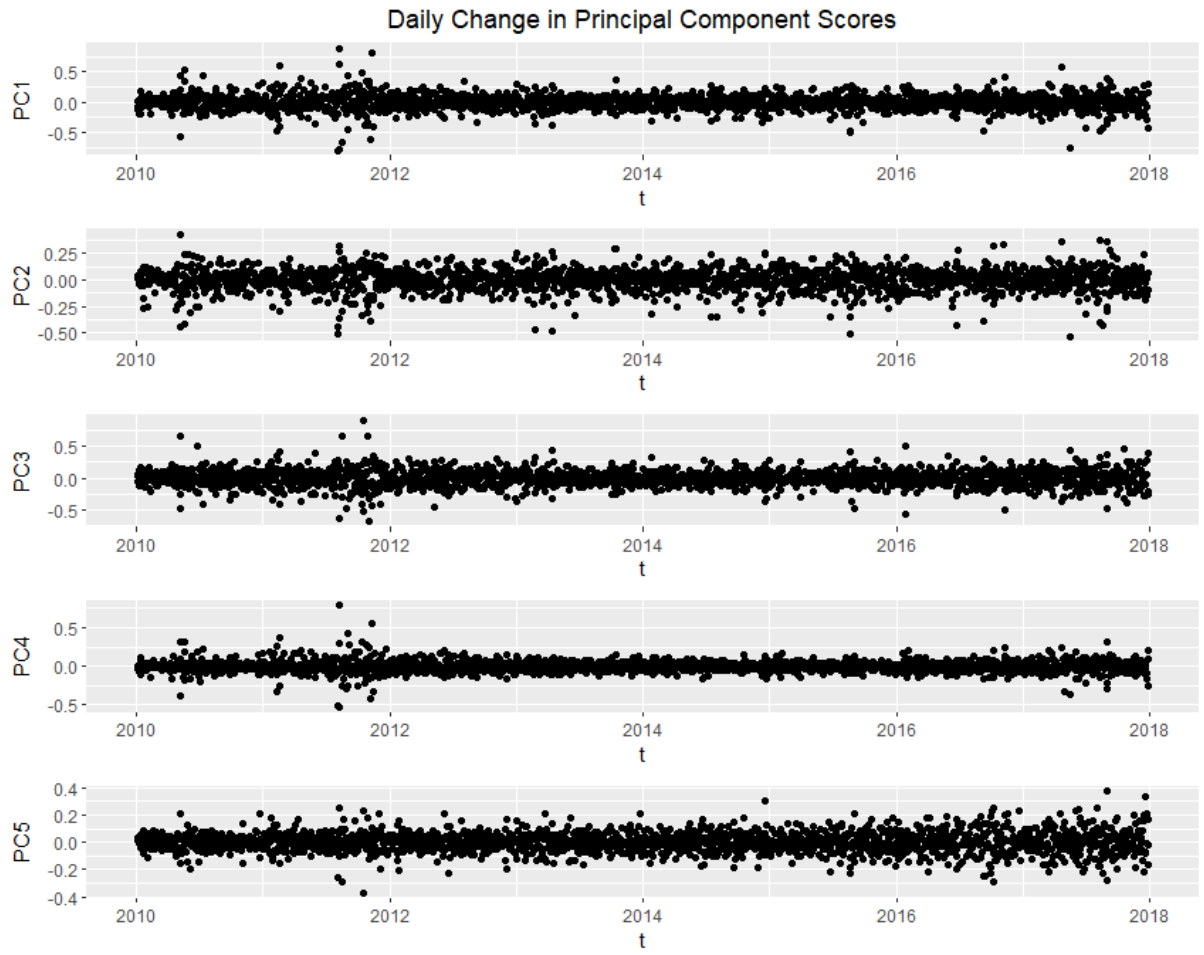

Figure 3: Differenced Principal Component Scores $\beta_{t+1,k} - \beta_{tk}$, which exhibit stochastic volatility.

4.4 Value-at-Risk Estimation for Option Portfolios

A forecast distribution for portfolio loss is required to estimate quantiles and thus Value-at-Risk. This section describes how to use the posterior samples $\{\boldsymbol{\theta}^j\}_{j=1}^J$ produced by our MCMC algorithm to build a conditional forecast distribution $\{L_{t+1}^{*j}\}_{j=1}^J$ for a next-day option portfolio loss random variable $L_{t+1} = \Pi_t(Y_{t+1})$ where $\Pi_t : L^2(\mathcal{T}) \to \mathbb{R}$. The option portfolio pricing functional Π_t is explicitly described by Equations 9–12. The forecast distribution $\{\boldsymbol{\theta}^j\}_{j=1}^J$ is constructed from Equations 8–11. The key quantity of interest in financial risk management is the Value-at-Risk (VaR) which is defined by the quantile of a loss distribution for a given time horizon and is defined in Equation 6.

For this application, the following priors are used for inference:

$$\begin{aligned}
\mu_k &\sim N(0,1) && \forall k,\\
(\phi_k + 1)/2 &\sim \text{Beta}(20, 1.5) && \forall k,\\
\sigma_k^2 &\sim \chi_1^2 && \forall k,\\
\Psi &\sim \text{Matrix Normal}_{K\times K}(0_{K\times K}, 10^6 I_K, I_K),\\
\sigma_\varepsilon^2 &\sim \text{InvGamma}(0.001, 0.001)
\end{aligned}$$

The prior for ϕ_k follows the example of Kim et al. (1998) and implies a prior mean of 0.86 to reflect the volatility clustering endemic to financial time series, while the priors for μ_k and σ_k do not make a difference with sufficient burn-in. In the matrix normal prior for Ψ, the scale matrix $U = 10^6 I_K$ results in a vague prior. Similarly, the parameters chosen for σ_ε^2 result in a vague prior. Since the amount of stochastic volatility varies for each year as observed in Figures 2 and 3, the FSV model is fit separately each year.

Figure 4 presents the 95% pointwise credible intervals for each of the log-volatility processes h_{tk} against t. Years 2010-2011 and 2015-2017 exhibit a higher degree of roughness in the random variation, indicating a higher presence of stochastic volatility for these years. This is further illustrated in Figure 5, which shows the posterior histograms of the stochastic volatility parameter σ_1, as the histograms are bounded away from 0 particularly in the years 2010-2011 and 2015-2017. The other σ_k's show similar results.

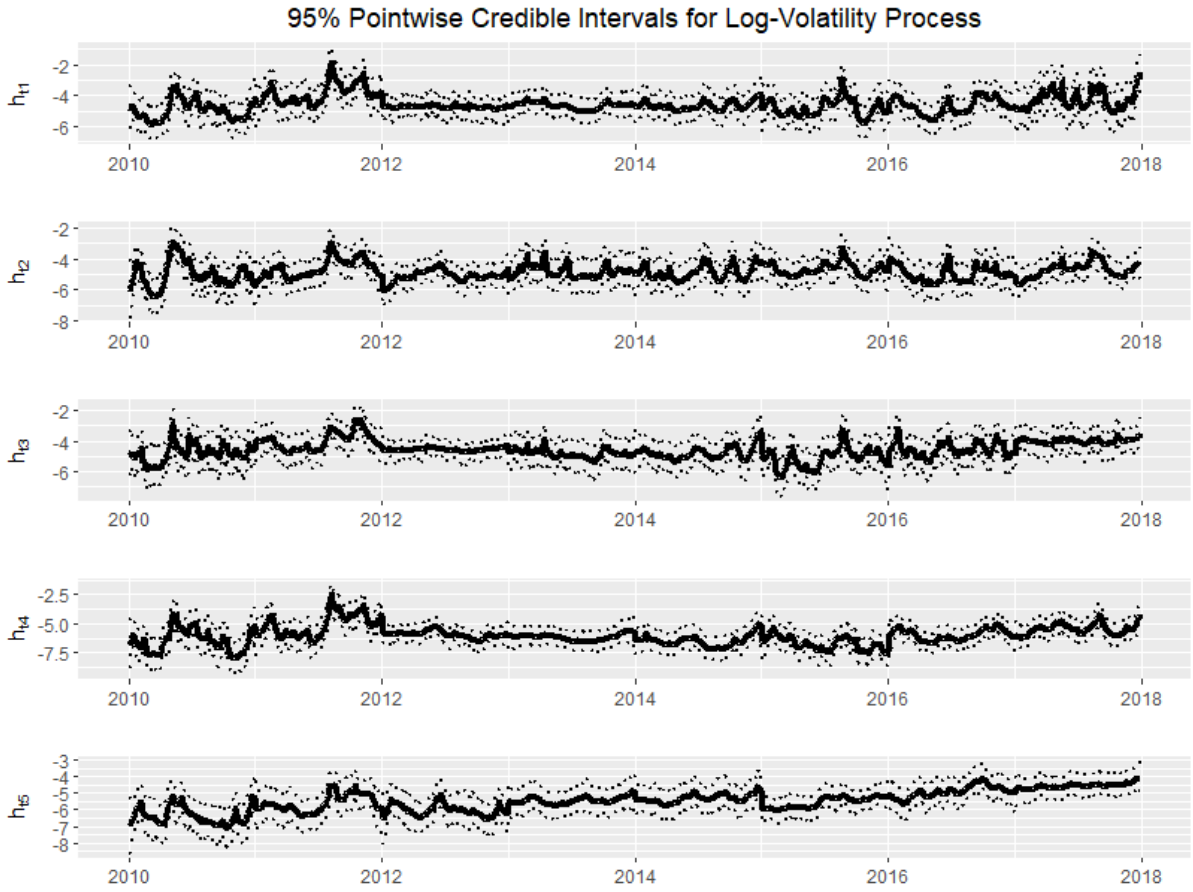

Figure 4: 95% pointwise credible bands for the latent log-volatility process with posterior median in bold. Dates range from start of 2010 to end of 2017. The roughness of random variation in h_{tk} indicates a notable presence of stochastic volatility in years 2010-2011 and 2015-2017.

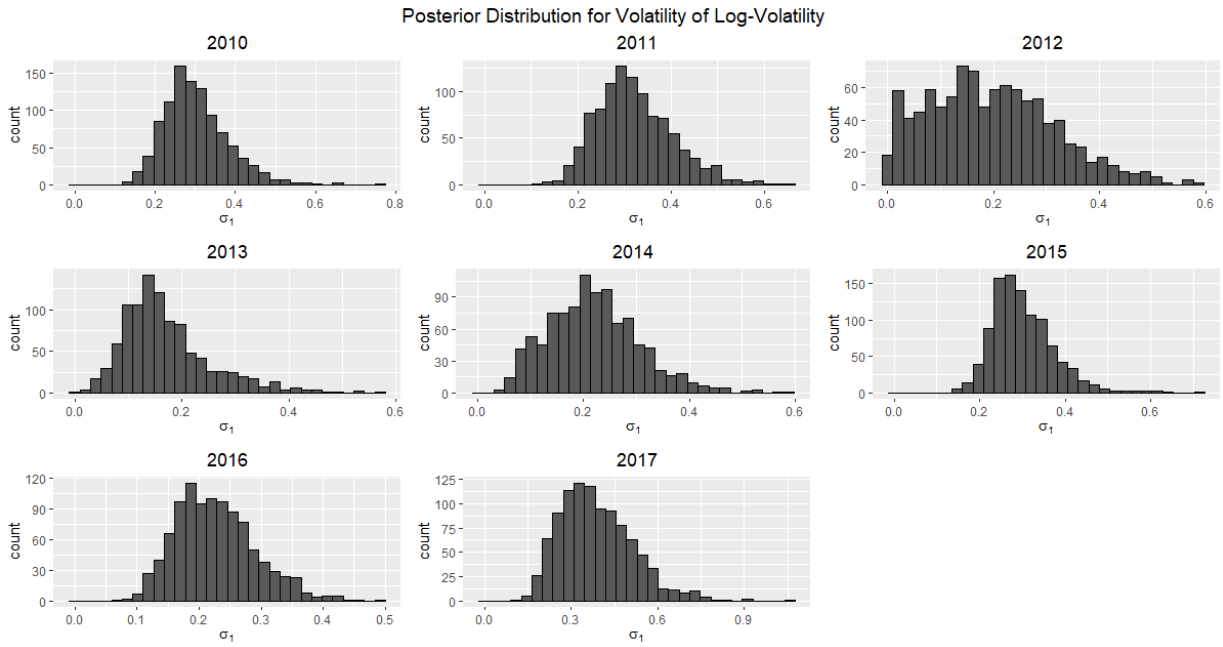

Figure 5: Posterior distributions for σ_1, the volatility of h_{t1}. The histograms exhibit most of their probability mass away from 0 particularly in the years 2010-2011 and 2015-2017, indicating the presence of stochastic volatility.

For the sake of brevity, it will be understood that a posterior draw will be for the fit on the year containing t. For each posterior sample of the associated year's model parameters $\boldsymbol{\theta}^j = (\{\boldsymbol{\beta}_t\}^j, \Psi^j, \sigma_\varepsilon^{2j}, \{\boldsymbol{h}_t\}^j, \boldsymbol{\mu}^j, \boldsymbol{\phi}^j, \boldsymbol{\sigma}^j)$, we forecast the option surface at time $t+1$ from time t as follows:

$$\begin{aligned}
\boldsymbol{\eta}_{t+1}^{*j} &\sim N_K(0, \text{diag}\{\boldsymbol{\sigma}^{2j}\}) \\
\boldsymbol{h}_{t+1}^{*j} &= \boldsymbol{\mu}^j + \boldsymbol{\phi}^j \cdot (\boldsymbol{h}_t^j - \boldsymbol{\mu}^j) + \boldsymbol{\eta}_{t+1}^{*j} \\
\boldsymbol{u}_{t+1}^{*j} | \boldsymbol{h}_{t+1}^{*j} &\sim N_K(\mathbf{0}, \text{diag}\{e^{\boldsymbol{h}_{t+1}^{*j}}\}) \\
\boldsymbol{\beta}_{t+1}^{*j} &= \Psi^j \boldsymbol{\beta}_t^j + \boldsymbol{u}_{t+1}^{*j} \\
Y_{t+1}^{*j}(\tau) &= F(\tau)^T \boldsymbol{\beta}_{t+1}^{*j}
\end{aligned}$$

where $F(\tau) = [f_1(\tau), \ldots, f_K(\tau)]^\top$.

Given this forecasted implied volatility surface $Y_{t+1}^{*j}(\tau)$ and observation points $\{\tau_{t+1|t,i}\}_{i=1}^{n_t}$,

we can compute the Black-Scholes log-implied volatilities $\{y^{*j}_{t+1,i}\}_{i=1}^{n_t}$ as

$$y^{*j}_{t+1,i} = m(\tau_{t+1|t,i}) + Y^{*j}_{t+1}(\tau_{t+1|t,i}) + \varepsilon^{*j}_{t+1,i}, \qquad \varepsilon^{*j}_{t+1,i} \overset{iid}{\sim} N(0, \sigma^{j2}_{\varepsilon}). \tag{8}$$

With their implied volatilities known, the options in the portfolio can then be priced with the Black-Scholes formula (Black and Scholes, 1973): If an option's log-implied volatility at time t is y_t, then its price is given by

$$\text{Prem}(M_t, S_t, K, CP, r_t(M_t), q_t, y_t) = CP[S_t e^{-q_t M_t}\Phi(CP \times d_+) - K e^{-r_t(M_t)M_t}\Phi(CP \times d_-)] \tag{9}$$

$$d_\pm = \frac{\log(S_t/K) + (r_t(M_t) - q_t \pm \frac{1}{2}e^{2y_t})M_t}{e^{y_t}\sqrt{M_t}} \tag{10}$$

where

- $M_t \equiv M_t(\tau^1) = (\tau^1)^2/365$ is the time to maturity in years at time t,
- S_t is the spot price of S&P500 at time t,
- K is the option strike price,
- $CP = \begin{cases} 1 \text{ for a call option,} \\ -1 \text{ for a put option,} \end{cases}$
- $r_t(M)$ is the risk-free rate at time t associated with time to maturity M,
- q_t is the dividend yield of S&P500 at time t,
- y_t is the log of implied volatility at time t,
- Φ is the standard normal cumulative distribution function.

Suppose on day t the set of traded options are at locations $\{\tau_{t,i}\}_{i=1}^{n_t}$. Then the updated observation points $\{\tau_{t+1|t,i}\}_{i=1}^{n_t}$ on day $t+1$ are calculated as follows: Days to maturity $\tau^1_{t+1|t,i}$ is reduced by the number of trading days to between time t and $t+1$ (usually 1 day,

though possibly more for weekends and holidays), and option delta $\tau^2_{t+1|t,i}$ is assumed equal to the previous day's value $\tau^2_{t,i}$. The other option parameters $S_{t+1|t}, K, r_{t+1|t}(M_{t+1}), q_{t+1|t}$ are held equal to day t's values $S_t, K, r_t(M_t), q_t$ to prevent looking ahead to future data.

Thus for each posterior sample $\{y^{*j}_{t+1,i}\}_{i=1}^{n_t}$ we can produce joint forecasts for the option prices $\{FP^{*j}_{t+1,i}\}_{i=1}^{n_t}$ where

$$FP^{*j}_{t+1,i} = \text{Prem}(M_{t+1}, S_t, K, CP, r_t(M_t), q_t, y^{*j}_{t+1,i})$$

These forecasted option prices can then be compared to the actual option prices $\{AP_{t,i}\}_{i=1}^{n_t}$ and $\{AP_{t+1,i}\}_{i=1}^{n_t}$ on days t and $t+1$, respectively, to assess the performance of the forecast. Given a portfolio with c_{ti} units of option i for $i = 1, \ldots, n_t$, we can compute the forecasted loss distribution $\{L^{*j}_{t+1}\}_{j=1}^{J}$ and actual loss L_{t+1} of the portfolio as

$$L^{*j}_{t+1} = \sum_{i=1}^{n_t} c_{ti}[AP_{t,i} - FP^{*j}_{t+1,i}] \tag{11}$$

$$L_{t+1} = \sum_{i=1}^{n_t} c_{ti}[AP_{t,i} - AP_{t+1,i}] \tag{12}$$

Then an estimate for the $(1-\alpha)100\%$ one-day Value-at-Risk for L_{t+1} is the sample quantile of our forecasted loss distribution.

$$\widehat{\text{VaR}}(L_{t+1}, 1-\alpha) = \text{quantile}(\{L^{*j}_{t+1}\}_{j=1}^{J}, 1-\alpha) \tag{13}$$

The next step is to assess whether the quantiles of the forecast distribution match the quantiles of the true loss distribution.

4.5 Backtesting Value-at-Risk

In order to evaluate the accuracy of our Value-at-Risk estimates, we can test if the historically observed frequencies of exceedances of the quantile estimates are consistent with

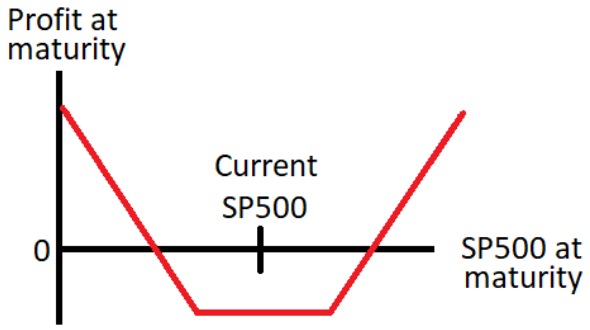

Figure 6: Profit diagram for an option strangle portfolio.

the true quantile levels. This process of evaluation based on historical data is known as backtesting. A 95% Value-at-Risk should be breached by approximately 5% of our observed losses if the observed trials are independent and our estimate is accurate. We can test for the latter using the Binomial hypothesis test proposed in Subsection 4.1 and the estimator proposed in Subsection 4.4.

We focus on option portfolios whose values are most sensitive to movements in volatility. An option strangle is a portfolio consisting of an out-of-money call option (whose strike is below the current S&P 500 level) and an out-of-money put option (whose strike is above the current S&P 500 level) where the call and put strikes are roughly the same distance apart from the current S&P 500 level. Such portfolios are particularly sensitive to movements in implied volatility $e^{Y_t(\tau)}$, since a strangle profits from large moves in the S&P but loses money on small moves. The profit diagram of a strangle portfolio is illustrated in Figure 6.

The proposed binomial test in Subsection 4.1 requires independent trials, so we use a different randomized option portfolio on each trading day in order to decorrelate the trials. To construct these randomized portfolios, on each day $t = 1, \ldots, T-1$, a set of random out-of-money options are chosen from the subset of options common to both days t and $t+1$ so that the actual price movement can be computed. These out-of-money options are chosen in 25 unique pairs of calls and puts so that we can form strangle positions. Hence, if a call option with strike price $K_1 > S_t$ is selected, then the put option whose strike is closest to $K_2 = S_t - (K_1 - S_t)$ is also chosen so that the put and call options are roughly

the same distance apart from the current spot price with both options out-of-money.

For each $t = 1, \ldots, 2012$, consider a simple randomized portfolio $\{c_{ti}\}_{i=1}^{50}$ on these 50 options where each c_{ti} is either 1 or -1 with equal probability. These coefficients determine the portfolio pricing functionals $\{\Pi_t\}$ through Equations 11 and 12. Since the actual price movements of these portfolios are known, we can perform the proposed binomial hypothesis test in Subsection 4.1 for the one-day Value-at-Risk estimator in Equation 13. A separate Binomial test is done for each year, and the 95% confidence intervals of the exceedence rates of VaR95, VaR97.5, and VaR99 are shown in blue in Figure 7.

For comparison, we also made forecasts based on a benchmark model with constant volatility. The benchmark model assumes each coefficient series $\{\beta_{tk}\}_{t=1}^{T}$ has constant volatility v_k, and denoting $\boldsymbol{v} = (v_1, \ldots, v_K)^\top$,

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{m}_t + F_t\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t &\sim N_{n_t}(\mathbf{0}, \sigma_\varepsilon^2 I) & \forall t \geq 1, \\ \boldsymbol{\beta}_t &= \Psi\boldsymbol{\beta}_{t-1} + \boldsymbol{\gamma}_t, & \boldsymbol{\gamma}_t &\sim N_K(\mathbf{0}, \mathrm{diag}\{\boldsymbol{v}\}) & \forall t \geq 1. \end{aligned}$$

Each v_k is assigned an independent InvGamma(a_k, b_k) prior . The full conditional for $\boldsymbol{v}$ is standard, and all other full conditionals remain the same with the substitution $\boldsymbol{h}_t = \log(\boldsymbol{v})$ for each t. The results in Figure 7 indicate that the exceedence rates of the stochastic volatility VaR's are generally closer to the true quantile levels than those of the constant volatility VaR's, as the stochastic volatility model yields confidence intervals that capture the true level in almost all years, while the constant volatility model's confidence intervals miss far more frequently. These findings indicate that incorporating stochastic volatility leads to better quantile estimates.

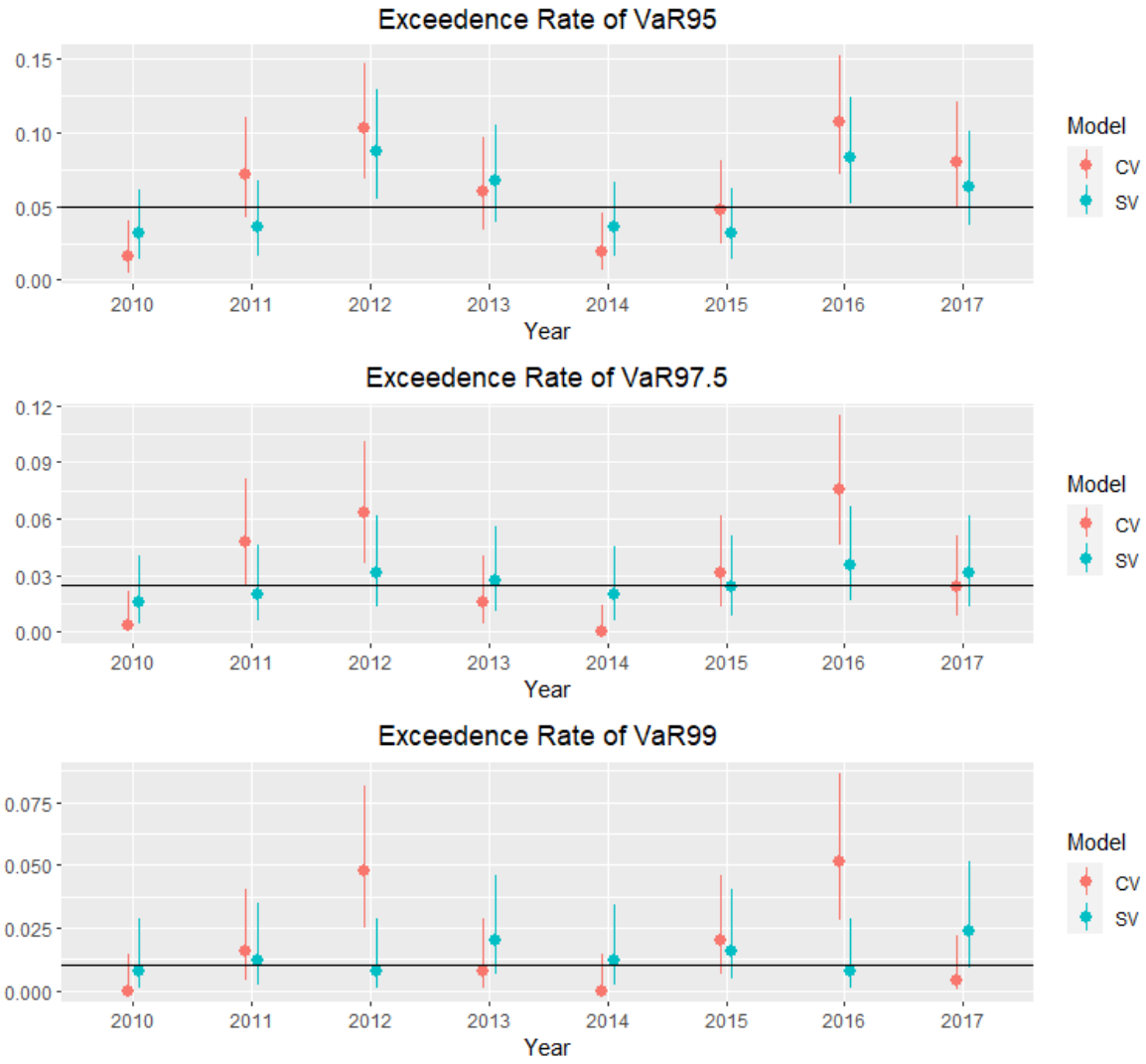

Figure 7: 95% confidence intervals for the exceedence rates of Value-at-Risk.

5 Conclusion

In this article, we have formulated an analogue of the familiar stochastic volatility model for functional time series, and constructed an example motivated through an application. We have derived a Bayesian hierarchical specification which can be used for uncertainty quantification, forecasting, and quantile estimation. We have demonstrated through the application an empirical example of functional stochastic volatility in SPX option surfaces and have shown how our methodology is well-suited for the daily shifts in the observation set. Through our exploratory analysis, and later in Figures 4 and 5, we have shown evidence of stochastic volatility in the functional time series. We have also shown through a backtesting exercise that the functional stochastic volatility model can improve quantile estimates for random functionals of the surfaces, which is of benefit in Value-at-Risk estimation for option portfolios.

6 Appendix

6.1 Hilbert-Schmidt Theorem and Dimension Reduction

Assuming $\mathcal{T}$ is uncountable, the functional time series $\{Y_t\}$ represents a random process with uncountably infinite dimension, but its reduction to a countably infinite L^2 approximation can be justified using the Hilbert-Schmidt theorem.

Let $v_t(\tau) = e^{h_t(\tau)}$, $\tilde{h}_t(\tau) = h_t(\tau) - \mu(\tau)$, and

$$\tilde{h}_t(\tau) = c(\tau) + \int_{\mathcal{T}} \phi(\tau, u)\tilde{h}_{t-1}(u)du + \zeta_t(\tau)$$

Suppose the parameter functions μ, ψ, k_ζ of the volatility process are such that $\{Y_t\}$ is a mean-zero covariance stationary process with a defined unconditional covariance function

$$Cov(\eta_t(\tau), \eta_t(u)) = k_\eta(\tau, u) \ \forall t.$$

If $k_\eta \in L^2(\mathcal{T} \times \mathcal{T})$, then the associated covariance operator $f \mapsto \int_{\mathcal{T}} k_\eta(\cdot, u) f(u) du$ is a Hilbert-Schmidt operator on $L^2(\mathcal{T})$, so by the Hilbert-Schmidt theorem we can decompose the covariance function k_η by its spectral expansion

$$k_\eta(\tau, u) \overset{L^2}{=} \sum_{k=1}^{\infty} \lambda_k f_k(\tau) f_k(u)$$

where the eigenfunctions $\{f_k\}_{k=1}^{\infty}$ form an orthonormal basis of $L^2(\mathcal{T})$ and the eigenvalues $\{\lambda_k\}_{k=1}^{\infty}$ are a non-negative decreasing square-summable sequence. For $\mathcal{T}$ compact we can strengthen the result to uniform convergence with Mercer's theorem if $k_\eta(\tau, \tau) \in L^1(\mathcal{T})$ or $k_\eta(\tau, u)$ is continuous.

To reduce from countably-infinite dimension to finite dimension, we appeal to the Karhunen-Loève theorem and truncate to the first K principal components $\{f_k\}_{k=1}^{K}$. We can get an empirical estimate of these eigenfunctions using FPCA on the sample covariance function of the spline-smoothed surfaces $\{\hat{Y}_t\}$:

$$\hat{k}_\eta(\tau, u) = \frac{1}{T} \sum_{t=1}^{T} \hat{Y}_t(\tau) \hat{Y}_t(u)$$

If we assume $\eta_t \in \text{Span}\{f_1, \ldots, f_K\}$ as in Section 2.3, then we can then write the conditional covariance functions as

$$Cov(\eta_t(\tau), \eta_t(u) | F_{t-1}) = \sum_{k=1}^{K} e^{h_{tk}} f_k(\tau) f_k(u)$$

and our model for functional stochastic volatility reduces to a vector stochastic volatility model on the log-eigenvalues $\{\boldsymbol{h}_t\}$. We know that $\{\boldsymbol{h}_t\}$ is a stochastic volatility model since $\boldsymbol{h}_t$ depends on ζ_t which is not $\mathcal{F}_{t-1}$ measurable.

6.2 Sampling the Full Conditional for $\boldsymbol{\beta}$

For the sake of brevity, in this section $e^{\boldsymbol{x}}$ for a vector $\boldsymbol{x} \in \mathbb{R}^d$ will be shorthand for the $d \times d$ diagonal matrix $\text{diag}\{e^{\boldsymbol{x}}\}$.

The precision matrix Λ_β takes on the following block matrix form, where each block is $K \times K$, and there are T rows and T columns of blocks:

$$
\begin{aligned}
\Lambda_\beta &= F^\top F + \sigma_\varepsilon^2 P^\top e^{-\boldsymbol{h}} P \\
&= \begin{bmatrix} F_1^\top F_1 & 0 & \cdots & 0 \\ 0 & F_2^\top F_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & F_T^\top F_T \end{bmatrix} \\
&\quad + \sigma_\varepsilon^2 \begin{bmatrix} I & -\Psi^\top & \cdots & 0 \\ 0 & I & \cdots & 0 \\ \vdots & \vdots & \ddots & -\Psi^\top \\ 0 & 0 & \cdots & I \end{bmatrix} \begin{bmatrix} e^{-\boldsymbol{h}_1} & 0 & \cdots & 0 \\ 0 & e^{-\boldsymbol{h}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & e^{-\boldsymbol{h}_T} \end{bmatrix} \begin{bmatrix} I & 0 & \cdots & 0 \\ -\Psi & I & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & -\Psi & I \end{bmatrix} \\
&= \begin{bmatrix} G_1 & H_1^\top & \cdots & 0 \\ H_1 & G_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & H_{T-1}^\top \\ 0 & \cdots & H_{T-1} & G_T \end{bmatrix}
\end{aligned}
$$

where

$$
G_t = \begin{cases} F_t^\top F_t + \sigma_\varepsilon^2 e^{-\boldsymbol{h}_t} + \sigma_\varepsilon^2 \Psi^\top e^{-\boldsymbol{h}_{t+1}} \Psi & \text{for } t = 1, \ldots, T-1 \\ F_T^\top F_T + \sigma_\varepsilon^2 e^{-\boldsymbol{h}_T} & \text{for } t = T \end{cases}
$$

and

$$
H_t = -\sigma_\varepsilon^2 e^{-\boldsymbol{h}_{t+1}} \Psi \text{ for } t = 1, \ldots, T-1
$$

As Λ_β is a block tridiagonal matrix, its Cholesky factor L is bidiagonal where $\Lambda_\beta = LL^\top$ as follows:

$$
\Lambda_\beta = \begin{bmatrix} G_1 & H_1^\top & \cdots & 0 \\ H_1 & G_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & H_{T-1}^\top \\ 0 & \cdots & H_{T-1} & G_T \end{bmatrix} = \underbrace{\begin{bmatrix} Q_1 & 0 & \cdots & 0 \\ R_1 & Q_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & R_{T-1} & Q_T \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} Q_1^\top & R_1^\top & \cdots & 0 \\ 0 & Q_2^\top & \ddots & 0 \\ \vdots & \ddots & \ddots & R_{T-1}^\top \\ 0 & \cdots & 0 & Q_T^\top \end{bmatrix}}_{L^\top}
$$

Blockwise multiplication yields the following system of equations:

$$\begin{aligned} G_1 &= Q_1 Q_1^\top \\ H_t &= R_t Q_t^\top && \text{for } t = 1, \ldots, T-1 \\ G_t &= R_{t-1} R_{t-1}^\top + Q_t Q_t^\top && \text{for } t = 2, \ldots, T \end{aligned}$$

and therefore we can recursively construct the blocks of L as follows:

$$\begin{aligned} Q_1 &= \text{Chol}(G_1) \\ R_t &= H_t Q_t^{-\top} && \text{for } t = 1, \ldots, T-1 \\ Q_t &= \text{Chol}(G_t - R_{t-1} R_{t-1}^\top) && \text{for } t = 2, \ldots, T \end{aligned}$$

Hence to compute $\boldsymbol{\mu}_\beta = \Lambda_\beta^{-1} F^\top (\boldsymbol{y} - \boldsymbol{m}) = L^{-\top} L^{-1} F^\top (\boldsymbol{y} - \boldsymbol{m})$, we can successively solve the systems below

$$\begin{aligned} L \tilde{\boldsymbol{\mu}}_\beta &= F^\top (\boldsymbol{y} - \boldsymbol{m}) \\ L^\top \boldsymbol{\mu}_\beta &= \tilde{\boldsymbol{\mu}}_\beta \end{aligned}$$

To obtain the explicit solution, first write the vectors as follows:

$$\boldsymbol{\mu}_\beta = \begin{bmatrix} \hat{\boldsymbol{m}}_1 \\ \vdots \\ \hat{\boldsymbol{m}}_T \end{bmatrix}, \qquad \tilde{\boldsymbol{\mu}}_\beta = \begin{bmatrix} \tilde{\boldsymbol{m}}_1 \\ \vdots \\ \tilde{\boldsymbol{m}}_T \end{bmatrix}, \qquad F^\top (\boldsymbol{y} - \boldsymbol{m}) = \begin{bmatrix} \boldsymbol{\xi}_1 \\ \vdots \\ \boldsymbol{\xi}_T \end{bmatrix}$$

where each entry above is a $K \times 1$ sub-vector.

To solve for $\tilde{\boldsymbol{\mu}}_\beta$ from $L \tilde{\boldsymbol{\mu}}_\beta = F^\top (\boldsymbol{y} - \boldsymbol{m})$,

$$\begin{bmatrix} Q_1 & 0 & \cdots & 0 \\ R_1 & Q_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & R_{T-1} & Q_T \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{m}}_1 \\ \vdots \\ \tilde{\boldsymbol{m}}_T \end{bmatrix} = \begin{bmatrix} \boldsymbol{\xi}_1 \\ \vdots \\ \boldsymbol{\xi}_T \end{bmatrix}$$

we have

$$\tilde{\boldsymbol{m}}_t = \begin{cases} Q_1^{-1}\boldsymbol{\xi}_1 & \text{for } t = 1, \\ Q_t^{-1}(\boldsymbol{\xi}_t - R_{t-1}\tilde{\boldsymbol{m}}_{t-1}) & \text{for } t = 2, \ldots, T \end{cases}$$

.

Then to solve for $\boldsymbol{\mu}_\beta$ from $L^\top \boldsymbol{\mu}_\beta = \tilde{\boldsymbol{\mu}}_\beta$,

$$\begin{bmatrix} Q_1^\top & R_1^\top & \cdots & 0 \\ 0 & Q_2^\top & \ddots & 0 \\ \vdots & \ddots & \ddots & R_{T-1}^\top \\ 0 & \cdots & 0 & Q_T^\top \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{m}}_1 \\ \vdots \\ \hat{\boldsymbol{m}}_T \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{m}}_1 \\ \vdots \\ \tilde{\boldsymbol{m}}_T \end{bmatrix}$$

we have

$$\hat{\boldsymbol{m}}_t = \begin{cases} Q_T^{-\top}\tilde{\boldsymbol{m}}_T & \text{for } t = T, \\ Q_t^{-\top}(\tilde{\boldsymbol{m}}_t - R_t^\top \hat{\boldsymbol{m}}_{t+1}) & \text{for } t = T-1, \ldots, 1 \end{cases}$$

.

Lastly, to sample the full conditional of $\boldsymbol{\beta}$, we draw a $TK \times 1$ vector

$$\boldsymbol{z} = \begin{bmatrix} \boldsymbol{z}_1 \\ \vdots \\ \boldsymbol{z}_T \end{bmatrix} \sim N_{TK}(\boldsymbol{0}, I)$$

where each $\boldsymbol{z}_i$ is a $K \times 1$ sub-vector. To compute $\boldsymbol{\beta}^{\text{post}} = \boldsymbol{\mu}_\beta + \sigma_\varepsilon L^{-\top}\boldsymbol{z}$, segment $\boldsymbol{\beta}^{\text{post}}$ into $K \times 1$ subvectors

$$\boldsymbol{\beta}^{\text{post}} = \begin{bmatrix} \boldsymbol{\beta}_1^{\text{post}} \\ \vdots \\ \boldsymbol{\beta}_T^{\text{post}} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{m}}_1 + \sigma_\varepsilon \hat{\boldsymbol{z}}_1 \\ \vdots \\ \hat{\boldsymbol{m}}_T + \sigma_\varepsilon \hat{\boldsymbol{z}}_T \end{bmatrix}$$

and compute

$$\hat{\boldsymbol{z}}_t = \begin{cases} Q_T^{-\top}\boldsymbol{z}_T & \text{for } t = T, \\ Q_t^{-\top}(\boldsymbol{z}_t - R_t^\top \hat{\boldsymbol{z}}_{t+1}) & \text{for } t = T-1, \ldots, 1 \end{cases}$$

6.3 Derivation of Full Conditional for Ψ

Consider a matrix normal prior for $\Psi \sim \text{Matrix Normal}_{K,K}(M, U, V)$:

$$p(\Psi) = (2\pi)^{-K^2/2}|U|^{-K/2}|V|^{-K/2}\exp\left(-\frac{1}{2}\text{tr}[V^{-1}(\Psi - M)^\top U^{-1}(\Psi - M)]\right)$$

Let $\Sigma_t = \text{diag}(\exp(\boldsymbol{h}_t))$. Then

$$\begin{aligned}
p(\Psi|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{h}, \sigma_\varepsilon^2) &\propto p(\boldsymbol{\beta}|\boldsymbol{h}, \Psi)p(\Psi) \\
&\propto \exp\left(-\frac{1}{2}\sum_{t=2}^{T}(\boldsymbol{\beta}_t - \Psi\boldsymbol{\beta}_{t-1})^\top\Sigma_t^{-1}(\boldsymbol{\beta}_t - \Psi\boldsymbol{\beta}_{t-1}) - \frac{1}{2}\text{tr}[V^{-1}(\Psi - M)^\top U^{-1}(\Psi - M)]\right) \\
&= \exp\left(-\Omega\right)
\end{aligned}$$

It is sufficient to look at the quadratic form $\boldsymbol{\Omega}$ to determine the full conditional distribution.

$$\begin{aligned}
\Omega &= \frac{1}{2}\sum_{t=2}^{T}(\boldsymbol{\beta}_t - \Psi\boldsymbol{\beta}_{t-1})^\top\Sigma_t^{-1}(\boldsymbol{\beta}_t - \Psi\boldsymbol{\beta}_{t-1}) + \frac{1}{2}\text{tr}\left[V^{-1}(\Psi - M)^\top U^{-1}(\Psi - M)\right] \\
&= \frac{1}{2}\sum_{t=2}^{T}\left(\frac{\boldsymbol{\beta}_t\boldsymbol{\beta}_{t-1}^\top}{\|\boldsymbol{\beta}_{t-1}\|^2}\boldsymbol{\beta}_{t-1} - \Psi\boldsymbol{\beta}_{t-1}\right)^\top\Sigma_t^{-1}\left(\frac{\boldsymbol{\beta}_t\boldsymbol{\beta}_{t-1}^\top}{\|\boldsymbol{\beta}_{t-1}\|^2}\boldsymbol{\beta}_{t-1} - \Psi\boldsymbol{\beta}_{t-1}\right) \\
&\quad + \frac{1}{2}\text{tr}\left[V^{-1}(\Psi - M)^\top U^{-1}(\Psi - M)\right] \\
&= \frac{1}{2}\sum_{t=2}^{T}\boldsymbol{\beta}_{t-1}^\top\left(\Psi - \frac{\boldsymbol{\beta}_t\boldsymbol{\beta}_{t-1}^\top}{\|\boldsymbol{\beta}_{t-1}\|^2}\right)^\top\Sigma_t^{-1}\left(\Psi - \frac{\boldsymbol{\beta}_t\boldsymbol{\beta}_{t-1}^\top}{\|\boldsymbol{\beta}_{t-1}\|^2}\right)\boldsymbol{\beta}_{t-1} \\
&\quad + \frac{1}{2}\text{tr}\left[V^{-1}(\Psi - M)^\top U^{-1}(\Psi - M)\right] \\
&= \frac{1}{2}\sum_{t=2}^{T}\text{tr}\left[\boldsymbol{\beta}_{t-1}^\top\left(\Psi - \frac{\boldsymbol{\beta}_t\boldsymbol{\beta}_{t-1}^\top}{\|\boldsymbol{\beta}_{t-1}\|^2}\right)^\top\Sigma_t^{-1}\left(\Psi - \frac{\boldsymbol{\beta}_t\boldsymbol{\beta}_{t-1}^\top}{\|\boldsymbol{\beta}_{t-1}\|^2}\right)\boldsymbol{\beta}_{t-1}\right] \\
&\quad + \frac{1}{2}\text{tr}\left[V^{-1}(\Psi - M)^\top U^{-1}(\Psi - M)\right] \\
&= \frac{1}{2}\sum_{t=2}^{T}\text{tr}\left[(\Psi - \xi_t)^\top\Sigma_t^{-1}(\Psi - \xi_t)\,\boldsymbol{\beta}_{t-1}\boldsymbol{\beta}_{t-1}^\top\right] \\
&\quad + \frac{1}{2}\text{tr}\left[(\Psi - M)^\top U^{-1}(\Psi - M)V^{-1}\right]
\end{aligned}$$

where $\xi_t = \frac{\boldsymbol{\beta}_t\boldsymbol{\beta}_{t-1}^\top}{\|\boldsymbol{\beta}_{t-1}\|^2}$. Using the identity

$$\text{tr}(A^\top BCD^\top) = \text{vec}(A)^\top(D \otimes B)\text{vec}(C)$$

we have

$$\begin{aligned}\Omega = &\frac{1}{2}\sum_{t=2}^{T}\left[\text{vec}(\Psi)-\text{vec}(\xi_t)\right]^\top\left(\boldsymbol{\beta}_{t-1}\boldsymbol{\beta}_{t-1}^\top\otimes\Sigma_t^{-1}\right)\left[\text{vec}(\Psi)-\text{vec}(\xi_t)\right]\\ &+\frac{1}{2}\left[\text{vec}(\Psi)-\text{vec}(M)\right]^\top\left(V^{-1}\otimes U^{-1}\right)\left[\text{vec}(\Psi)-\text{vec}(M)\right]\\ =&\frac{1}{2}\text{vec}(\Psi)^\top\left[\sum_{t=2}^{T}(\boldsymbol{\beta}_{t-1}\boldsymbol{\beta}_{t-1}^\top\otimes\Sigma_t^{-1})+(V^{-1}\otimes U^{-1})\right]\text{vec}(\Psi)\\ &-\text{vec}(\Psi)^\top\left[\sum_{t=2}^{T}(\boldsymbol{\beta}_{t-1}\boldsymbol{\beta}_{t-1}^\top\otimes\Sigma_t^{-1})\text{vec}(\xi_t)+(V^{-1}\otimes U^{-1})\text{vec}(M)\right]+\text{const}\end{aligned}$$

After simplifying the terms below,

$$\begin{aligned}(\boldsymbol{\beta}_{t-1}\boldsymbol{\beta}_{t-1}^\top\otimes\Sigma_t^{-1})\text{vec}(\xi_t) &= (\boldsymbol{\beta}_{t-1}\boldsymbol{\beta}_{t-1}^\top\otimes\Sigma_t^{-1})\frac{\text{vec}(\boldsymbol{\beta}_t\boldsymbol{\beta}_{t-1}^\top)}{\|\boldsymbol{\beta}_{t-1}\|^2}\\ &= \text{vec}(\Sigma_t^{-1}\boldsymbol{\beta}_t\frac{\boldsymbol{\beta}_{t-1}^\top}{\|\boldsymbol{\beta}_{t-1}\|^2}\boldsymbol{\beta}_{t-1}\boldsymbol{\beta}_{t-1}^\top)\\ &= \text{vec}(\Sigma_t^{-1}\boldsymbol{\beta}_t\boldsymbol{\beta}_{t-1}^\top)\end{aligned}$$

$$(V^{-1}\otimes U^{-1})\text{vec}(M) = \text{vec}(U^{-1}MV^{-1})$$

we get

$$\begin{aligned}\boldsymbol{\Omega} = &\frac{1}{2}\text{vec}(\Psi)^\top\left[\sum_{t=2}^{T}(\boldsymbol{\beta}_{t-1}\boldsymbol{\beta}_{t-1}^\top\otimes\Sigma_t^{-1})+(V^{-1}\otimes U^{-1})\right]\text{vec}(\Psi)\\ &-\text{vec}(\Psi)^\top\text{vec}\left(\sum_{t=2}^{T}\Sigma_t^{-1}\boldsymbol{\beta}_t\boldsymbol{\beta}_{t-1}^\top+U^{-1}MV^{-1}\right)+\text{const}\end{aligned}$$

which is the quadratic form for the multivariate normal below:

$$\begin{aligned}&\text{vec}(\Psi)|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{h},\sigma_\varepsilon^2\sim N_{K^2}(\boldsymbol{\mu}_\Psi,\Sigma_\Psi)\\ &\boldsymbol{\mu}_\Psi = \Sigma_\Psi\text{vec}\left(\sum_{t=2}^{T}\Sigma_t^{-1}\boldsymbol{\beta}_t\boldsymbol{\beta}_{t-1}^\top+U^{-1}MV^{-1}\right)\\ &\Sigma_\Psi = \left[\sum_{t=2}^{T}(\boldsymbol{\beta}_{t-1}\boldsymbol{\beta}_{t-1}^\top\otimes\Sigma_t^{-1})+(V^{-1}\otimes U^{-1})\right]^{-1}\end{aligned}$$

References

Asai, M., M. McAleer, and J. Yu (2006, 09). Multivariate stochastic volatility: A review. *Econometric Reviews 25*, 145–175.

Aue, A., L. Horváth, and D. F. Pellatt (2017). Functional generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis 38*(1), 3–21.

Aue, A. and J. Klepsch (2017, 01). Estimating functional time series by moving average model fitting.

Aue, A., D. D. Norinho, and S. Hörmann (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association 110*(509), 378–392.

Besse, P. C., H. Cardot, and D. B. Stephenson (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics 27*(4), 673–687.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of political economy 81*(3), 637.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics 72*(3), 498–505.

Bollerslev, T., R. Engle, and J. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy 96*(1), 116–31.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. Lecture Notes in Statistics. Springer New York.

Bosq, D. (2014). Computing the best linear predictor in a hilbert space. applications to general armah processes. *Journal of Multivariate Analysis 124*, 436–450.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*, Volume 36. SIAM.

Bui Quang, P., T. Klein, N. H. Nguyen, and T. Walther (2018). Value-at-risk for south-east asian stock markets: Stochastic volatility vs. garch. *Journal of Risk and Financial Management 11*(2).

Chen, Y. and B. Li (2017). An adaptive functional autoregressive forecast model to predict electricity price curves. *Journal of Business & Economic Statistics 35*(3), 371–388.

Cressie, N. and C. Wikle (2015). *Statistics for Spatio-Temporal Data*. Wiley.

Daniélsson, J. (1998). Multivariate stochastic volatility models: Estimation and a comparison with vgarch models. *Journal of Empirical Finance 5*(2), 155–173.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*(4), 987–1007.

Gatheral, J. (2006). *The Volatility Surface: A Practitioner's Guide*. Wiley finance series. John Wiley & Sons.

Gelman, A., J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin (2013). *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.

Ghysels, E., A. Harvey, and E. Renault (1996). Stochastic volatility. Cahiers de recherche, Centre interuniversitaire de recherche en économie quantitative, CIREQ.

Han, C.-H., W.-H. Liu, and T.-Y. Chen (2014). VaR/CVaR estimation under stochastic volatility models. *International Journal of Theoretical and Applied Finance 17*(02), 1450009.

Harvey, A., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *The Review of Economic Studies 61*(2), 247–264.

Heston, S. L. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies 6*(2), 327–343.

Heston, S. L. and S. Nandi (2000, 07). A Closed-Form GARCH Option Valuation Model. *The Review of Financial Studies 13*(3), 585–625.

Hörmann, S., L. Horváth, and R. Reeder (2013). A functional version of the ARCH model. *Econometric Theory 29*(2), 267–288.

Hörmann, S., L. u. Kidziński, and M. Hallin (2014). Dynamic functional principal components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77*(2), 319–348.

Hörmann, S. and P. Kokoszka (2010). Weakly dependent functional data. *The Annals of Statistics 38*(3), 1845 – 1884.

Horváth, L. and P. Kokoszka (2012). *Inference for Functional Data with Applications*. Springer Series in Statistics. Springer New York.

Hsing, T. and R. Eubank (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley Series in Probability and Statistics. Wiley.

Huang, A. Y. (2015). Value at risk estimation by threshold stochastic volatility model. *Applied Economics 47*(45), 4884–4900.

Huang, S.-F., M. Guo, and M.-R. Chen (2020). Stock market trend prediction using a functional time series approach. *Quantitative Finance 20*(1), 69–79.

Hull, J. (2018). *Options, Futures, and Other Derivatives*. Pearson.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance 42*(2), 281–300.

Hyndman, R. and H. Booth (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting 24*(3), 323–342.

Hyndman, R. and M. Shahid Ullah (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis 51*(10), 4942–4956.

Hyndman, R. J., H. Booth, and F. Yasmeen (2013). Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography 50*(1), 261–283.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics 12*(4), 371–389.

Jang, P. A., A. E. Loeb, M. B. Davidow, and A. G. Wilson (2017). Scalable Lévy process priors for spectral kernel learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 3943–3952.

Jang, P. A. and D. S. Matteson (2018). Spatial correlation in weather forecast accuracy: A functional time series approach. *JSM 2018 Proceedings, Statistical Computing Section*, 2776–2784.

Karhunen, K. (1947). über lineare methoden in der wahrscheinlichkeitsrechnung. *Annales Academiae Scientiarum Fennicae 37*, 1–79.

Kastner, G. (2016). Dealing with stochastic volatility in time series using the r package stochvol. *Journal of Statistical Software, Articles 69*(5), 1–30.

Kastner, G. and S. Frühwirth-Schnatter (2014). Ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc estimation of stochastic volatility models. *Computational Statistics & Data Analysis 76*, 408–423. CFEnetwork: The Annals of Computational and Financial Econometrics.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies 65*(3), 361–393.

King, M. C., A.-M. Staicu, J. M. Davis, B. J. Reich, and B. Eder (2018). A functional data analysis of spatiotemporal trends and variation in fine particulate matter. *Atmospheric Environment 184*, 233–243.

Klepsch, J., C. Klüppelberg, and T. Wei (2017). Prediction of functional ARMA processes with an application to traffic data. *Econometrics and Statistics 1*(C), 128–149.

Kowal, D. R., D. S. Matteson, and D. Ruppert (2017). A bayesian multivariate functional dynamic linear model. *Journal of the American Statistical Association 112*(518), 733–744.

Kowal, D. R., D. S. Matteson, and D. Ruppert (2019). Functional autoregression for sparsely sampled data. *Journal of Business & Economic Statistics 37*(1), 97–109.

Li, D., P. M. Robinson, and H. L. Shang (2020). Long-range dependent curve time series. *Journal of the American Statistical Association 115*(530), 957–971.

Loève, M. (1945). Fonctions aléatoires de second ordre. *Comptes Rendus de l'Académie des Sciences, Série I: Mathématique 220*, 469.

OptionMetrics (2017). IvyDB US File and Data Reference Manual, Version 3.1, Rev. 1/17/2017. 1776 Broadway, Suite 1800, New York, NY 10019.

Ramsay, J., J. Ramsay, B. Silverman, S. S. Media, and H. Silverman (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer.

Ruiz-Medina, M., R. Espejo, M. Ugarte, and A. Militino (2013, 05). Functional time series analysis of spatio–temporal epidemiological data. *Stochastic Environmental Research and Risk Assessment 28*, 943–954.

Sadorsky, P. (2005). Stochastic volatility forecasting and risk management. *Applied Financial Economics 15*(2), 121–135.

Sen, R. and C. Klüppelberg (2019). Time series of functional data with application to yield curves. *Applied Stochastic Models in Business and Industry 35*(4), 1028–1043.

Shang, H. (2013). Functional time series approach for forecasting very short-term electricity demand. *Journal of Applied Statistics 40*(1), 152–168.

Shang, H. L., H. Booth, and R. Hyndman (2011). Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods. *Demographic Research 25*, 173–214.

Shephard, N. (1996). *Statistical aspects of ARCH and stochastic volatility* ((edited by D.R. Cox, David V. Hinkley and Ole E. Barndorff-Neilsen) ed.)., pp. 1–67. London: Chapman & Hall. Reprinted in the Survey of Applied and Industrial Mathematics, issue on Financial and insurance mathematics, 3, 764-826, Scientific Publisher TVP, Moscow, 1996 (in Russian).

Shephard, N. and T. G. Andersen (2009). *Stochastic Volatility: Origins and Overview*, pp. 233–254. Berlin, Heidelberg: Springer Berlin Heidelberg.

Taylor, S. (1982). Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. *Time Series Analysis : Theory and Practice 1*, 203–226.

Taylor, S. (1986). *Modelling Financial Time Series*. Wiley.

Taylor, S. (1994, April). Modelling stochastic volatility: a review and comparative study. *Mathematical Finance 4*(2), 183–204.

Vrontos, I. D., P. Dellaportas, and D. N. Politis (2003). A full-factor multivariate garch model. *The Econometrics Journal 6*(2), 312–334.

Yasmeen, F. and M. Sharif (2015, 09). Functional time series (fts) forecasting of electricity consumption in pakistan. *International Journal of Computer Applications 124*, 975–8887.

Yu, J. and R. Meyer (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews 25*(2-3), 361–384.