

RESEARCH ARTICLE

Functional Stochastic Volatility in Financial Option Surfaces

Phillip A. Jang^a and Michael Jauch^b and David S. Matteson^c

^aCenter for Applied Mathematics, Cornell University, Ithaca, NY, USA; ^bDepartment of Statistics, Florida State University, Tallahassee, FL, USA; ^cDepartment of Statistics and Data Science, Cornell University, Ithaca, NY, USA

ARTICLE HISTORY

Compiled August 11, 2022

ABSTRACT

Stochastic volatility, or variability that is well approximated as a random process, is widespread in modern finance. While understanding that volatility is essential for sound decision making, the structural and data constraints associated with complex financial instruments limit the applicability of classical volatility modeling. This article investigates stochastic volatility in functional time series with the goal of accurately modeling option surfaces. We begin by introducing a functional analogue of the familiar stochastic volatility models employed in univariate and multivariate time series analysis. We then describe how that functional specification can be reduced to a finite dimensional vector time series model and discuss a strategy for Bayesian inference. Finally, we present a detailed application of the functional stochastic volatility model to daily SPX option surfaces. We find that the functional stochastic volatility model, by accounting for the heteroscedasticity endemic to option surface data, leads to improved quantile estimates. More specifically, we demonstrate through backtesting that Value-at-Risk estimates from the proposed functional stochastic volatility model exhibit correct coverage more consistently than those of a constant volatility model.

KEYWORDS

Functional Time Series, Stochastic Volatility, Option Surface, Value-at-Risk

CONTACT Michael Jauch. Email: mjauch@fsu.edu

1. Introduction

A *functional time series* $\{Y_t(\tau)\}$ is a time-ordered sequence ($t = 1, 2, \dots$) of random functions often used to model the time dynamics of random phenomena that live on a continuum ($\tau \in \mathcal{T}$). Functional time series models enable statistical analysis and prediction of curves and surfaces evolving over time. The study of functional time series has become increasingly relevant as the collection of data in high resolution or frequency has become commonplace.

Functional time series can offer key practical advantages over multivariate or vector time series. By modeling random phenomena on a continuum, functional time series meet the needs of real world applications in which measurement locations are often sparse, changing over time, or irregularly spaced. Applying multivariate time series methods in similar settings would require that measurements lie on a regular unchanging grid and demand special handling of missing data. Furthermore, the number of parameters associated with a vector time series model would grow with the number of measurement locations and quickly become unreasonable, whereas the parametrization of a functional time series model is independent of the measurement locations.

More fundamentally, functional time series methods extend the domain of functional data analysis to time dependent functional data. While the celebrated results of Karhunen (1947) and Loève (1945) imply that functional principal component analysis (FPCA) provides an optimal finite dimensional representation for i.i.d. functional data, that optimality no longer holds when observations exhibit time dependence. However, FPCA provides a foundation for the development of methods intended specifically for time dependent functional data. 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).

Functional time series methods 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 modeling (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 modeling (Klepsch et al. 2017), and electricity demand forecasting (Shang 2013; Yasmeeen and Sharif 2015; Chen and Li 2017).

Many important multivariate time series models have been generalized to a functional setting. The functional analogue of the autoregressive integrated moving average (ARIMA) model has been developed by several authors. Bosq (2000) developed the theoretical foundations of the functional autoregression model on Hilbert and Banach spaces, while Aue and Klepsch (2017) studied the functional moving average model. Klepsch et al. (2017) derived sufficient conditions for the wide-sense stationarity of a functional ARMA model on separable Hilbert spaces. Sen and Klüppelberg (2019) explored how a functional ARMA model leads to a vector ARMA structure on the principal component scores. Li et al. (2020) addressed long memory in functional time series by developing fractionally differenced functional ARIMA models. Developments related to forecasting functional time series appear in Hyndman and Shahid Ullah (2007), Bosq (2014), and Aue et al. (2015). Kowal et al. (2017) and Kowal et al. (2019) introduced a Bayesian framework for functional time series through a dynamic linear model formulation.

Compared to the extensive literature on functional generalizations of ARIMA and dynamic linear models, relatively little has been written on the important problem of modeling 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 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) among others. Multivariate GARCH models were examined in Bollerslev et al. (1988) and Bollerslev (1990), while multivariate SV models were considered in Harvey et al.

(1994), Daniélsson (1998), and Asai et al. (2006). Bayesian methods for multivariate GARCH and SV models were proposed in Vrontos et al. (2003) and Yu and Meyer (2006), respectively. To date, there have been several important contributions on the topic of heteroscedastic functional time series models. Hörmann and Kokoszka (2010) formalized the notion of weakly dependent functional time series, while Hörmann et al. (2014) addressed serial correlation through a dynamic Karhunen-Loève expansion, a functional analogue of the dynamic principal component analysis introduced for multivariate time series in Brillinger (1981). Hörmann et al. (2013) and Aue et al. (2017) proposed functional ARCH and GARCH models, respectively.

This article investigates stochastic volatility in the functional time series setting. 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. Although seemingly less popular in applications, stochastic volatility models have been observed to provide a better fit with fewer parameters when compared to GARCH models (Daniélsson 1998; Kim et al. 1998). SV models also provide a natural discrete analogue to the stochastic differential equations used in option pricing, such as the stochastic volatility diffusion of Hull and White (1987).

Stochastic volatility is ubiquitous in financial time series modeling, as it is often observed empirically in the dynamics of asset prices and returns (Hull and White 1987; Heston 1993; Heston and Nandi 2000; Gatheral 2006; Shephard and Andersen 2009). Accordingly, we illustrate the benefits of our functional stochastic volatility (FSV) model in a financial application and model the daily price movements of SPX options contracts¹. For each day from January 1, 2010 to December 31, 2017, we have quotes of hundreds to thousands of option contracts which differ by their strike price and time to maturity. These contracts can be understood as a discrete collection of points observed from an underlying continuous option surface. Exploratory analyses indicate that these SPX option surfaces exhibit stochastic volatility. We demonstrate through a backtest that the FSV model leads to more accurate estimates of the portfolio Value-at-Risk (VaR) compared to a constant volatility model as its quantile estimates consistently

¹SPX options have the Standard and Poor's 500 Index as the underlying asset.

exhibit the correct coverage. Hence the FSV model can benefit portfolio managers, risk managers, and other practitioners who need to monitor risk metrics of portfolios containing futures, options, swaptions, and other derivatives whose price dynamics are determined by an underlying continuum. More broadly, the FSV model can benefit any forecaster who is concerned not only with prediction but also quantile estimation and uncertainty quantification for functional time series data.

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

2. Method

In time series analysis, data often exhibit both time-varying volatility and autocorrelation. Therefore, SV models are commonly used in conjunction with ARMA models by allowing the innovations of the ARMA process to have time-varying volatility. In this section, we formulate the FSV model in the context and notation of a functional ARMA model. Then we show how the FSV model can be reduced to a tractable finite dimensional representation that facilitates inference with familiar techniques from vector time series. Lastly, we describe the basis functions used for dimension reduction.

2.1. Functional Stochastic Volatility

We define an FSV model as a sequence of random functions that can be decomposed as the product of a volatility process with a serially uncorrelated noise process. More precisely, let $\mathcal{T} \subseteq \mathcal{R}^d$ be compact and let $L^2(\mathcal{T})$ be the set of square-integrable functions defined on \mathcal{T} . Suppose η_t is a sequence of random functions in $L^2(\mathcal{T})$ that can be decomposed as the product

$$\eta_t(\tau) = v_t(\tau)z_t(\tau)$$

where

- (1) $\{z_t\}$ is a sequence of i.i.d. mean zero random functions in $L^2(\mathcal{T})$ with a common correlation function $k_z(\tau, u) = \text{Cov}[z_t(\tau), z_t(u)]$ for all t so that $k_z(\tau, \tau) = 1$ for all τ ,
- (2) $\{v_t\}$ is a sequence of non-negative random functions in $L^2(\mathcal{T})$,
- (3) $\{z_t\}$ and $\{v_t\}$ are independent.

Then $\{\eta_t\}$ is a *functional stochastic volatility process* if, for all t , the conditional volatility function v_t is not measurable with respect to the sigma field generated by $\{(v_s, \eta_s)\}_{s=1}^{t-1}$. This is in contrast to a functional ARCH or GARCH process where v_t is measurable with respect to the above sigma field. The FSV process will drive the functional ARMA process introduced in the next section.

2.2. Functional ARMA with Stochastic Volatility

We assume that the data of interest arise from an underlying functional ARMA process $\{Y_t\}$ driven by an FSV process as defined in the previous section. More precisely, 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 $\tau_{t,1}, \dots, \tau_{t,n_t} \in \mathcal{T}$ are the observation locations, $m(\tau)$ is the mean function, and $\varepsilon_{t,i} \stackrel{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$ is the measurement error.

The functional ARMA(p,q) model for $\{Y_t\}$ satisfies

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

where $\{\eta_t\}$ is a sequence of mean zero random functions with $\mathbb{E} \|\eta_t\|^2 \equiv \mathbb{E} \left[\int_{\mathcal{T}} \eta_t(\tau)^2 d\tau \right] < \infty$ and $\mathbb{E}[\eta_t(\tau)\eta_s(\tau)] = 0$ when $t \neq s$ for all τ . In other words, the functions $\{\eta_t\}$ are uncorrelated across time but not necessarily independent. These conditions are satisfied by the FSV model in Section 2.1, and hereafter we will suppose that $\{\eta_t\}$ is an FSV process. The parameter kernel functions $\{\psi_i\}$ and $\{\theta_j\}$ are analogous to the parameter matrices in an multivariate ARMA model and determine the time-varying conditional mean.

Previous work has established sufficient conditions for the existence of a unique wide-sense stationary and causal solution to Equation 1. Suppose the volatility pro-

cess $\{v_t\}$ is wide-sense stationary so that there exists a common function $k_v(\tau, u) = \mathbb{E}[v_t(\tau)v_t(u)]$ for all t , implying that $k_\eta(\tau, u) = \mathbb{E}[\eta_t(\tau)\eta_t(u)] = k_v(\tau, u)k_z(\tau, u)$. It is shown in Theorem 3.8 of Klepsch et al. (2017) and in Li et al. (2020) that a unique wide-sense stationary and causal solution exists provided that k_η and $\{\theta_j\}$ are Hilbert-Schmidt kernels so that $\int_{\mathcal{T}} \int_{\mathcal{T}} k_\eta(\tau, u)^2 d\tau du < \infty$ and $\int_{\mathcal{T}} \int_{\mathcal{T}} \theta_j(\tau, u)^2 d\tau du < \infty$ for each j , and $\{\psi_i\}$ satisfies the summability condition

$$\sum_{i=1}^p \left[\int_{\mathcal{T}} \int_{\mathcal{T}} \psi_i(\tau, u)^2 d\tau du \right]^{1/2} < 1.$$

The functional ARMA model can be defined more generally with bounded linear operators acting on separable Hilbert spaces as in Klepsch et al. (2017).

For the sake of simplicity and computational tractability, we will work with the functional AR(1) process and drop the index on the kernel function ψ_1 so that

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

2.3. Reducing a Functional Time Series to a Vector Time Series

The functional ARMA model becomes more tractable upon making a few assumptions. In particular, we suppose there exist orthonormal deterministic functions $f_1, \dots, f_K \in L^2(\mathcal{T})$ such that the following conditions hold:

- (1) The functional ARMA process $\{Y_t\}$ lies in $\text{Span}(f_1, \dots, f_K)$. In that case, there exist random coefficient vectors $\beta_t = (\beta_{t1}, \dots, \beta_{tK})^\top$ and $\gamma_t = (\gamma_{t1}, \dots, \gamma_{tK})^\top$ with $\beta_{tk} = \langle Y_t, f_k \rangle$ and $\gamma_{tk} = \langle \eta_t, f_k \rangle$ such that $Y_t = \sum_{k=1}^K \beta_{tk} f_k$ and $\eta_t = \sum_{k=1}^K \gamma_{tk} f_k$. Because the functions f_1, \dots, f_K are orthogonal, the entries of each β_t vector are uncorrelated, as are the entries of each γ_t vector.
- (2) The kernel $\psi(\tau, u) \in \text{Span}(f_1, \dots, f_K) \otimes \text{Span}(f_1, \dots, f_K)$, 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 $\psi_{kj} = \langle\langle \psi(\star, \cdot), f_j(\cdot) \rangle, f_k(\star) \rangle = \int_{\mathcal{T}} \int_{\mathcal{T}} \psi(\tau, u) f_j(u) f_k(\tau) du d\tau$.

(3) The coefficient vectors γ_t are a time-dependent sequence of random vectors with

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

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

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

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

Under assumptions 1-3 above, the coefficients $\{\beta_t\}_{t=1}^T$ form a K -dimensional first order vector autoregressive, or VAR(1), process satisfying

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

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

In this article, we use VAR(1) dynamics for \mathbf{h}_t , a choice well-suited for modeling volatility clustering. More precisely, we suppose

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

For the sake of parsimony, we assume the matrices Φ and Σ_{ζ} are diagonal and that the series $\{h_{tj}\}_{t=1}^T$ is independent of the series $\{h_{tk}\}_{t=1}^T$ for all distinct pairs of indices

j and k . Thus, for K -dimensional parameter vectors $\boldsymbol{\phi}$ and $\boldsymbol{\sigma}_\zeta$, we have

$$\begin{aligned}\mathbf{h}_t &= \boldsymbol{\mu} + \boldsymbol{\phi}^\top (\mathbf{h}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\zeta}_t \\ \boldsymbol{\zeta}_t &\stackrel{\text{iid}}{\sim} N_K [\mathbf{0}, \text{diag}(\boldsymbol{\sigma}_\zeta^2)].\end{aligned}$$

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

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

$$\boldsymbol{\beta}_t = \Psi \boldsymbol{\beta}_{t-1} + \boldsymbol{\gamma}_t, \quad \boldsymbol{\gamma}_t | \mathbf{h}_t \sim N_K[\mathbf{0}, \text{diag}(e^{\mathbf{h}_t})], \quad (3)$$

$$\mathbf{h}_t = \boldsymbol{\mu} + \boldsymbol{\phi}^\top (\mathbf{h}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\zeta}_t, \quad \boldsymbol{\zeta}_t \sim N_K[\mathbf{0}, \text{diag}(\boldsymbol{\sigma}_\zeta^2)], \quad (4)$$

$$\mathbf{h}_1 \sim N_K \left[\boldsymbol{\mu}, \text{diag} \left(\frac{\sigma_1^2}{1 - \phi_1^2}, \dots, \frac{\sigma_K^2}{1 - \phi_K^2} \right) \right] \quad (5)$$

where

- (1) $\{\boldsymbol{\varepsilon}_t\}$ and $\{\boldsymbol{\zeta}_t\}$ are independent across time 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 \mathbf{h}_t , and
- (4) \mathbf{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, best approximates the underlying process with minimal information loss. In our application, we use the first K functional principal components, estimated from spline-smoothed functional data, as our basis functions f_1, \dots, f_K . Refer to Appendix Section A for further justification.

3. Inference

This section describes Bayesian inference for the FSV 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, \dots, y_n are sampled from a joint density $p(y_1, \dots, y_n \mid \boldsymbol{\theta})$ 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, \dots, y_n is given by

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

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 $\Theta = \Theta_1 \times \dots \times \Theta_M$ and then iteratively sampling from the full conditional distributions $p(\boldsymbol{\theta}_1 \mid \{y_i\}_{i=1}^n, \{\boldsymbol{\theta}_m\}_{m \neq 1}), \dots, p(\boldsymbol{\theta}_M \mid \{y_i\}_{i=1}^n, \{\boldsymbol{\theta}_m\}_{m \neq M})$. The full conditional distributions are often analytically tractable even when the joint posterior distribution is not. One advantage of the Gibbs sampling approach is that it does not require any tuning parameters. The Gibbs sampling algorithm is also 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 FSV model and the resulting full conditional distributions. In order to write these succinctly, we introduce matrix variate notation to simplify Equations 2-5 in Section 2.3.

From the terms in Equations 3-5, we define the following KT -dimension random vectors $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_T^\top)^\top$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \dots, \boldsymbol{\gamma}_T^\top)^\top$, and $\boldsymbol{h} = (\boldsymbol{h}_1^\top, \dots, \boldsymbol{h}_T^\top)^\top$. Then the random vectors satisfy

$$P\boldsymbol{\beta} = \boldsymbol{\gamma} \tag{6}$$

where

$$P = \begin{bmatrix} \mathbf{I}_K & \mathbf{0}_K & \mathbf{0}_K & \dots & \mathbf{0}_K \\ -\boldsymbol{\Psi} & \mathbf{I}_K & \mathbf{0}_K & \dots & \mathbf{0}_K \\ \mathbf{0}_K & -\boldsymbol{\Psi} & \mathbf{I}_K & \dots & \mathbf{0}_K \\ \dots & \dots & \dots & \ddots & \dots \\ \mathbf{0}_K & \mathbf{0}_K & \mathbf{0}_K & \dots & \mathbf{I}_K \end{bmatrix}.$$

and \mathbf{I}_K and $\mathbf{0}_K$ denote the $K \times K$ identity and zero matrices respectively.

From the terms in Equation 2, we set $n_y \equiv \sum_{t=1}^T n_t$ and define the n_y -dimensional random vectors $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \dots, \boldsymbol{y}_T^\top)^\top$, $\boldsymbol{m} = (\boldsymbol{m}_1^\top, \dots, \boldsymbol{m}_T^\top)^\top$, and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, \dots, \boldsymbol{\varepsilon}_T^\top)^\top$, along with the block diagonal matrix $\boldsymbol{F} = \text{diag}(\boldsymbol{F}_1, \dots, \boldsymbol{F}_T)$. Then we have

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

Writing the SV parameters as $\boldsymbol{\theta}_h = (\boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\sigma}_\zeta)$, it follows from Equation 2 that

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

and

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

For the log-variance process \mathbf{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 \mathbf{h} given $\boldsymbol{\theta}_h$ factors as

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

where

$$\begin{aligned} \mathbf{h}_t \mid \mathbf{h}_{t-1}, \boldsymbol{\theta}_h &\sim N_K \left[\boldsymbol{\mu} + \boldsymbol{\phi}^\top (\mathbf{h}_{t-1} - \boldsymbol{\mu}), \text{diag}(\boldsymbol{\sigma}_\zeta^2) \right] \\ \mathbf{h}_1 \mid \boldsymbol{\theta}_h &\sim N_K \left[\boldsymbol{\mu}, \text{diag} \left(\frac{\sigma_1^2}{1 - \phi_1^2}, \dots, \frac{\sigma_K^2}{1 - \phi_K^2} \right) \right], \end{aligned}$$

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

$$p(\boldsymbol{\theta}_h) = \prod_{k=1}^K p(\mu_k) p(\phi_k) p(\sigma_k^2)$$

with

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

for all k .

To the remaining parameters, we assign the following priors, which yield closed

form full conditional distributions:

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

where $\text{Matrix Normal}_{K,K}(\mathbf{M}, \mathbf{U}, \mathbf{V})$ indicates a $K \times K$ matrix normal distribution with mean matrix \mathbf{M} and scale matrices \mathbf{U} and \mathbf{V} . The density function is given in Appendix Section C. The notation $\text{InvGamma}(a_\varepsilon, b_\varepsilon)$ indicates an inverse gamma distribution with shape parameter a_ε and scale parameter b_ε .

In summary, all full conditional densities will be proportional to the joint density of $(\mathbf{y}, \boldsymbol{\beta}, \mathbf{h}, \boldsymbol{\theta}_h, \boldsymbol{\Psi}, \sigma_\varepsilon^2)$ which factors as

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

The full set of prior hyperparameters is $(b_\mu, B_\mu, a_\phi, b_\phi, \mathbf{M}, \mathbf{U}, \mathbf{V}, a_\varepsilon, b_\varepsilon, B_\sigma)$.

The full conditional distributions of \mathbf{h} and $\boldsymbol{\theta}_h$ along with an associated sampler are presented in Kim et al. (1998) and Kastner and Frühwirth-Schnatter (2014). After Gibbs updates to $\boldsymbol{\beta}$ and $\boldsymbol{\Psi}$, $\boldsymbol{\gamma}$ is recalculated through Equation 6. As each $\{\gamma_{tk}\}_{t=1}^T$ is assumed to be independent of the others with 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). That article carries out Bayesian inference for $\{h_{tk}\}_{t=1}^T$ and its three associated parameters $\{\mu_k, \phi_k, \sigma_k\}$ through a mixture sampler applied to $\{\gamma_{tk}\}_{t=1}^T$. The R package `stochvol` developed by Kastner (2016) employs this mixture sampler as well as the Ancillarity-Sufficiency Interweaving Strategy in Kastner and Frühwirth-Schnatter (2014) to achieve highly efficient Bayesian inference.

The remaining full conditional distributions are as follows:

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

where

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

The full conditional distributions for $\boldsymbol{\beta}$ and σ_ε^2 follow from the well-known Bayesian linear regression with the normal-inverse-gamma conjugate prior. A derivation can be found, for example, in Gelman et al. (2013). Though the dimension of $\boldsymbol{\beta}$ is potentially very large, we can exploit the block structure of its precision matrix. Appendix Section B describes how to sample efficiently from the full conditional for $\boldsymbol{\beta}$. Refer to Appendix Section C for a derivation of the full conditional for $\boldsymbol{\Psi}$.

4. Value-at-Risk Estimation for Option Portfolios

In this section, we apply the FSV model to SPX option surface data in order to estimate Value-at-Risk for option portfolios. The results indicate that modeling stochastic volatility 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 motivates the application of an FSV 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 estimating 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 whether the same pattern holds for a functional time series.

In the present application, the functional time series $\{Y_t(\tau)\}_{t=1}^T$ represents a surface of option price quotes and is modeled with the proposed FSV 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 κ 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 κ 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 $\Pi_t : L^2(\mathcal{T}) \rightarrow \mathbb{R}$. 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 . Both 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

$$\text{VaR}(L_{t+1}, 1 - \alpha) = \sup \{x \in \mathbb{R} : \mathbb{P}(L_{t+1} \leq x) < 1 - \alpha\} \quad (7)$$

$$= F_{L_{t+1}}^{-1}(1 - \alpha) \quad \text{if } L_{t+1} \text{ is a continuous R.V.} \quad (8)$$

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 $\widehat{\text{VaR}}(L_{t+1}, 1 - \alpha)$ is to compare the proportion of observed exceedences of the quantile estimate to the true level of the quantile. Define the exceedence indicator variables $X_t = \mathbb{1}_{\{L_{t+1} > \widehat{\text{VaR}}(L_{t+1}, 1 - \alpha)\}}$ for $t = 1, \dots, T - 1$ 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 $N_{T-1} \equiv \sum_{t=1}^{T-1} X_t \sim \text{Binomial}(T-1, \alpha_N)$. Kupiec (1995) proposes 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 $\widehat{\text{VaR}}(L_{t+1}, 1 - \alpha)$ 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 method 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 11). 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:

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

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 analysis.

On any given day t , the observation points $\tau_{t,1}, \dots, \tau_{t,n_t} \in \mathcal{T}$ 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 (that is, when $\tau^1 = 0$). 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 an FSV 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 suggest that the functional time series exhibits stochastic volatility.

We used penalized spline smoothing, as implemented in the R package `mgcv` by Wood (2011), to estimate the mean function $m(\tau)$ and functional principal components f_1, \dots, f_K . These are shown in Figure 1. The following knot sequences, chosen based on empirical quantiles of the observation points, were used in each dimension of the domain to define the cubic tensor splines:

$$\mathcal{K}^1 = (1.4, 1.4, 1.4, 1.4, 8.1, 12.7, 19.3, 33.1, 33.1, 33.1, 33.1),$$

$$\mathcal{K}^2 = [0.10, 0.10, 0.10, 0.10, 0.37, 0.57, 0.71, 0.79, 0.86, 0.90, 0.90, 0.90, 0.90].$$

The 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.

The first principal component $f_1(\tau)$ is negative when time to maturity is small while positive when time to maturity is large, resulting in a tilt of the surface with respect to maturity. 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 maturity end of the surface with respect to delta. 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.

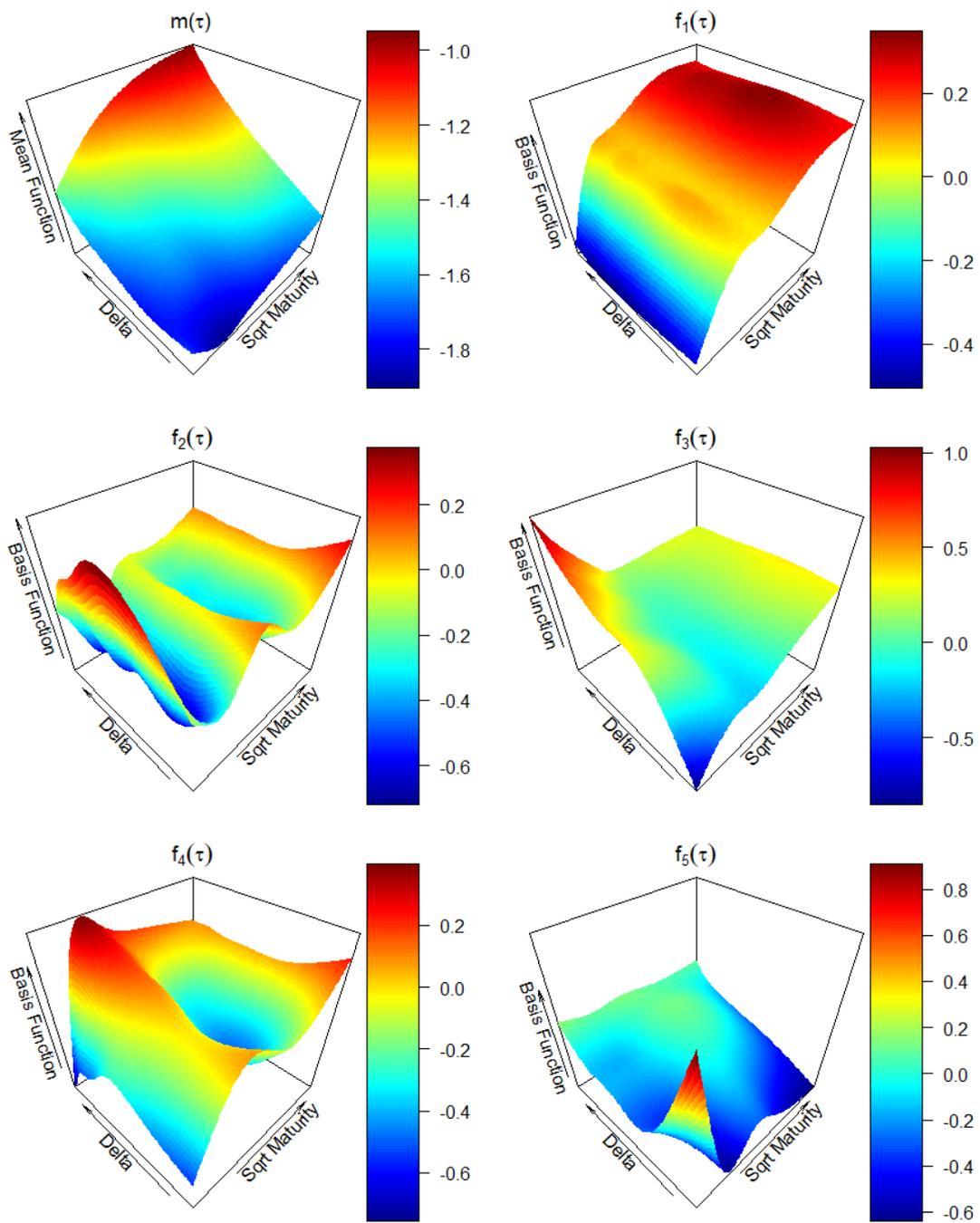


Figure 1.: Mean Function $m(\tau)$ and Functional Principal Components $f_1(\tau), \dots, f_5(\tau)$.

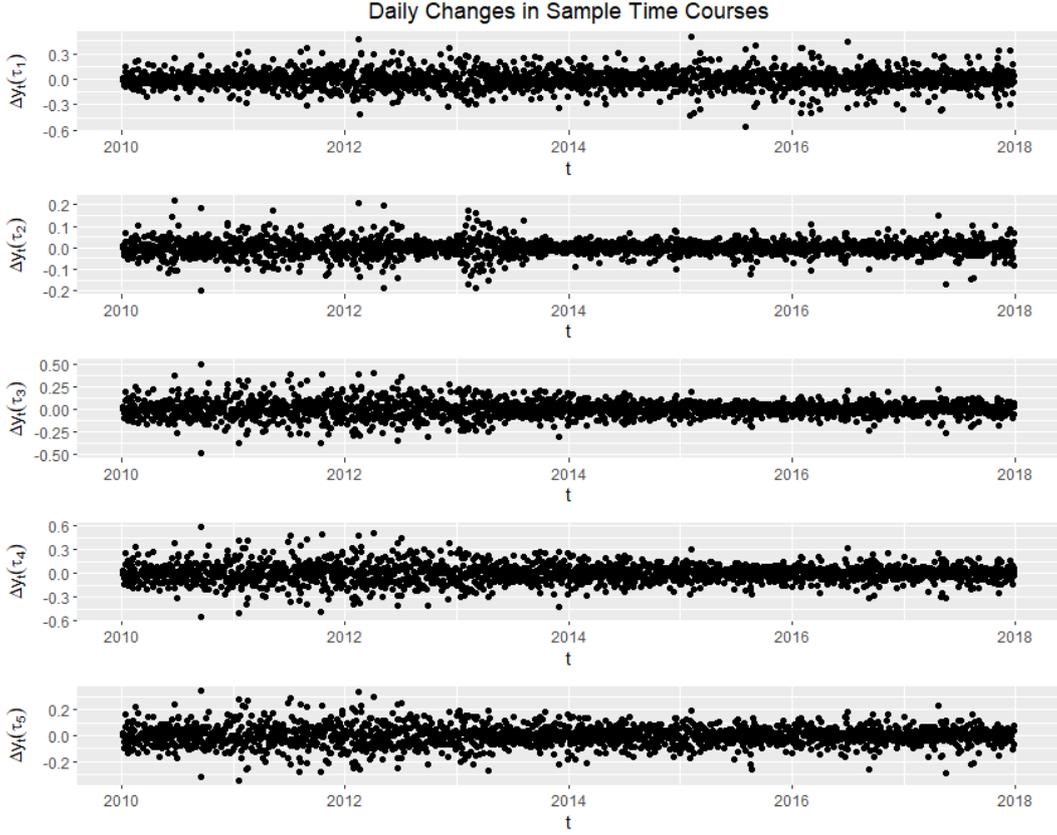


Figure 2.: Differenced time courses $\{Y_{t+1}(\tau_j) - Y_t(\tau_j)\}_{t=1}^{T-1}$ for five sample locations τ_1, \dots, τ_5 , which exhibit stochastic volatility.

In an exploratory analysis, we see evidence for stochastic volatility. For five fixed sample domain points τ_1, \dots, τ_5 , the 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 are shown in Figure 3. Both sets of series exhibit stochastic volatility and motivate the application of the FSV model. Because the amount of stochastic volatility observed in the previous section varies year-by-year, the model is fit separately to each of the 8 years of data.

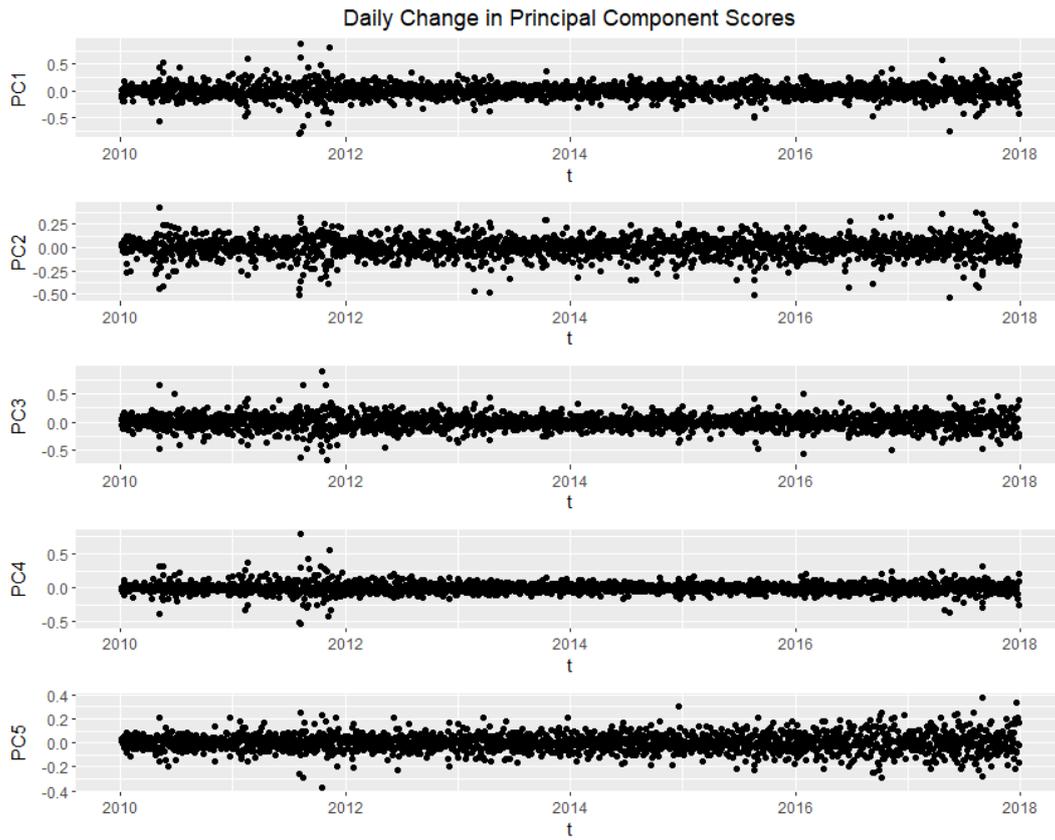


Figure 3.: Differenced Principal Component Scores $\beta_{t+1,k} - \beta_{tk}$, which exhibit stochastic volatility.

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 $\Pi_t : L^2(\mathcal{T}) \rightarrow \mathbb{R}$. The option portfolio pricing functional Π_t is explicitly described by Equations 11–14. The forecast distribution $\{\boldsymbol{\theta}^j\}_{j=1}^J$ is constructed from Equations 9–13. 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 7.

For this application, we complete the prior specification as follows. For each k ,

$$\begin{aligned}\mu_k &\sim N(0, 1), \\ (\phi_k + 1)/2 &\sim \text{Beta}(20, 1.5), \\ \sigma_k^2 &\sim \chi_1^2,\end{aligned}$$

while

$$\begin{aligned}\boldsymbol{\Psi} &\sim \text{Matrix Normal}_{K,K}(\mathbf{0}_K, 10^6 \cdot \mathbf{I}_K, \mathbf{I}_K), \\ \sigma_\varepsilon^2 &\sim \text{InvGamma}(0.001, 0.001).\end{aligned}$$

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. We have found that the results are not sensitive to the choice of priors for μ_k and σ_k provided the series has a reasonable length. The hyperparameters of the matrix normal prior for $\boldsymbol{\Psi}$ and the inverse gamma prior for σ_ε^2 were chosen so that these prior distributions do not strongly influence the results.

As we saw in our exploratory analysis (see Figures 2 and 3), the amount of stochastic volatility varies for each year. Thus, we fit the FSV model separately to each year of data. Figure 4 presents the 95% pointwise credible intervals for each of the

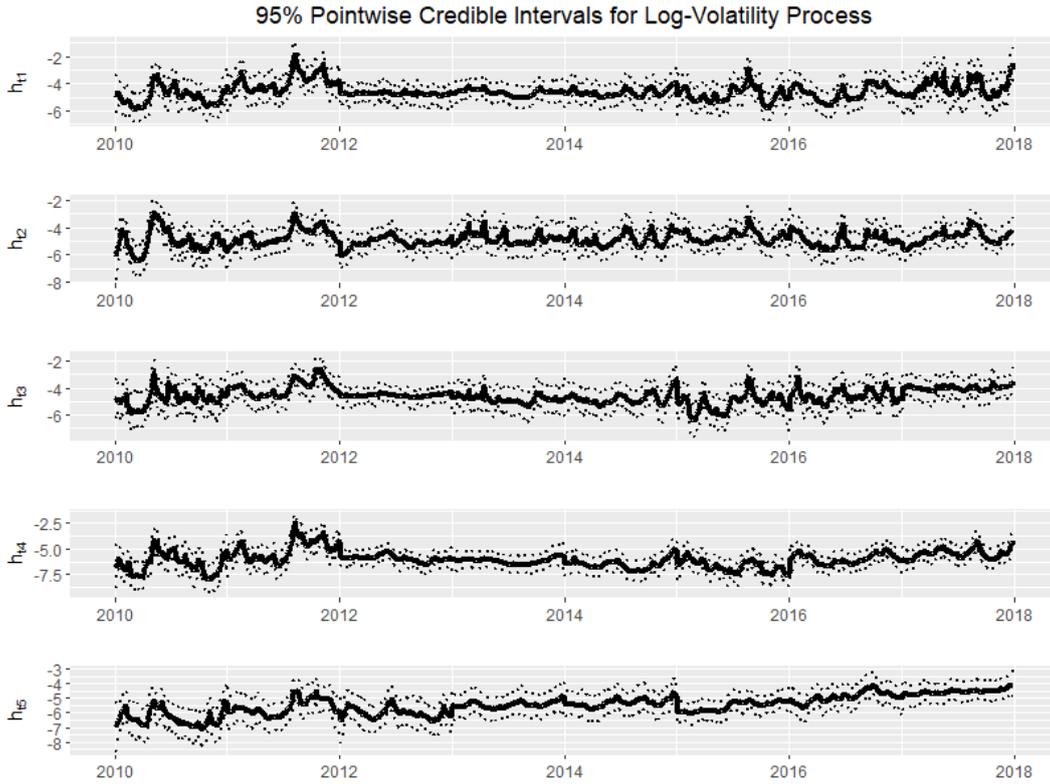


Figure 4.: 95% pointwise credible bands for the latent log-volatility process with posterior median in bold. Dates range from start of 2010 to end of 2017. The roughness of random variation in h_{tk} indicates a notable presence of stochastic volatility in years 2010-2011 and 2015-2017.

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 by the posterior histograms of the stochastic volatility parameter σ_1 appearing in Figure 5. The histograms are bounded away from zero, especially in the years 2010-2011 and 2015-2017. The results are similar for $\sigma_2, \dots, \sigma_K$.

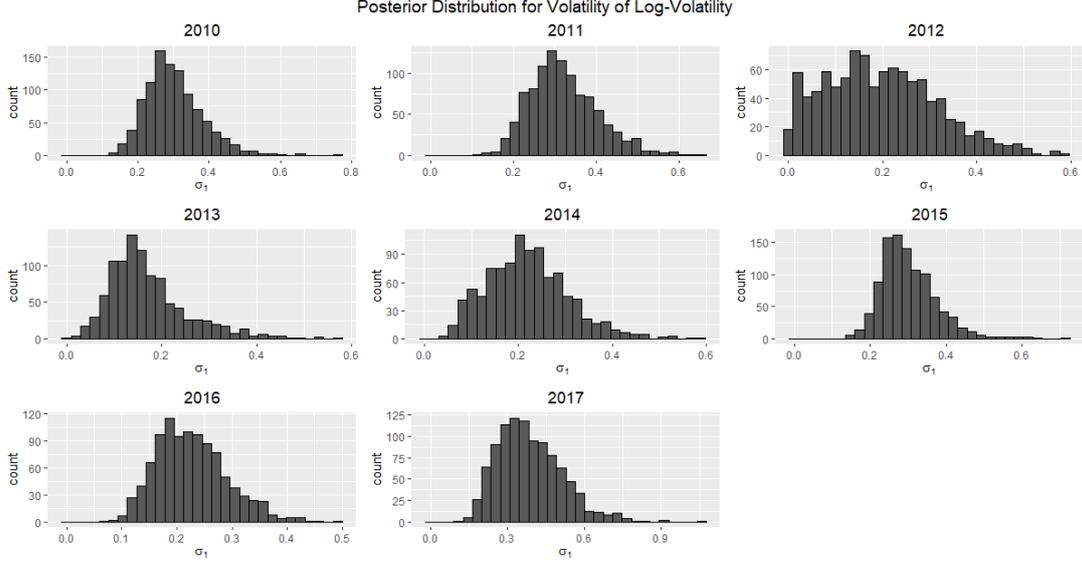


Figure 5.: Posterior distributions for σ_1 , the volatility of h_{t1} . The histograms exhibit most of their probability mass away from 0 particularly in the years 2010-2011 and 2015-2017, indicating the presence of stochastic volatility.

To simplify the presentation of what follows, it should be understood that a posterior draw corresponds to the fit from the year containing t . For each posterior sample from the associated year's model parameters $\theta^j = (\{\beta_t\}^j, \Psi^j, \sigma_\varepsilon^{2j}, \{h_t\}^j, \mu^j, \phi^j, \sigma^j)$, we forecast the option surface at time $t + 1$ from time t as follows:

$$\begin{aligned}
\zeta_{t+1}^{*j} &\sim N_K [0, \text{diag}(\sigma^{2j})] \\
h_{t+1}^{*j} &= \mu^j + \phi^{j\top}(h_t^j - \mu^j) + \zeta_{t+1}^{*j} \\
\mathbf{u}_{t+1}^{*j} \mid h_{t+1}^{*j} &\sim N_K [\mathbf{0}, \text{diag}(e^{h_{t+1}^{*j}})] \\
\beta_{t+1}^{*j} &= \Psi^j \beta_t^j + \mathbf{u}_{t+1}^{*j} \\
Y_{t+1}^{*j}(\tau) &= F(\tau)^T \beta_{t+1}^{*j}
\end{aligned}$$

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

Given this forecast of the 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 $\{y_{t+1,i}^{*j}\}_{i=1}^{n_t}$ as

$$y_{t+1,i}^{*j} = m(\tau_{t+1|t,i}) + Y_{t+1}^{*j}(\tau_{t+1|t,i}) + \varepsilon_{t+1,i}^{*j}, \quad \varepsilon_{t+1,i}^{*j} \stackrel{\text{iid}}{\sim} N(0, \sigma_\varepsilon^{j2}). \quad (9)$$

With their implied volatilities known, the options in the portfolio can 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

$$\text{Prem} [M_t, S_t, \kappa, CP, r_t(M_t), q_t, y_t] = \quad (10)$$

$$CP \left[S_t e^{-q_t M_t} \Phi(CP \times d_+) - \kappa e^{-r_t(M_t) M_t} \Phi(CP \times d_-) \right] \quad (11)$$

where

$$d_{\pm} = \frac{\log(S_t/\kappa) + [r_t(M_t) - q_t \pm \frac{1}{2}e^{2y_t}] M_t}{e^{y_t} \sqrt{M_t}} \quad (12)$$

and

- $M_t \equiv M_t(\tau^1) = (\tau^1)^2 / 365$ is the time to maturity in years at time t ,
- S_t is the spot price of S&P500 at time t ,
- κ 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:

- (1) Days to maturity $\tau_{t+1|t,i}^1$ is reduced by the number of trading days between time t and $t+1$. This is usually one day but can be more because of weekends and holidays.
- (2) Option delta $\tau_{t+1|t,i}^2$ is assumed equal to the previous day's value $\tau_{t,i}^2$.
- (3) The other option parameters $S_{t+1|t}, \kappa, r_{t+1|t}(M_{t+1}), q_{t+1|t}$ are held equal to day t 's values $S_t, \kappa, r_t(M_t), q_t$ to prevent looking ahead to future data.

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

option prices $\{FP_{t+1,i}^{*j}\}_{i=1}^{n_t}$ where

$$FP_{t+1,i}^{*j} = \text{Prem} \left[M_{t+1}, S_t, \kappa, CP, r_t(M_t), q_t, y_{t+1,i}^{*j} \right].$$

These forecasted option prices can 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 $i = 1, \dots, n_t$, we can compute the forecasted loss distribution $\{L_{t+1}^{*j}\}_{j=1}^J$ and actual loss L_{t+1} of the portfolio as

$$L_{t+1}^{*j} = \sum_{i=1}^{n_t} c_{ti} \left(AP_{t,i} - FP_{t+1,i}^{*j} \right) \quad (13)$$

$$L_{t+1} = \sum_{i=1}^{n_t} c_{ti} (AP_{t,i} - AP_{t+1,i}). \quad (14)$$

We can take sample quantile of the forecasted loss distribution as our estimate of the $(1 - \alpha)100\%$ one-day Value-at-Risk, setting

$$\widehat{\text{VaR}}(L_{t+1}, 1 - \alpha) = \text{quantile} \left(\{L_{t+1}^{*j}\}_{j=1}^J, 1 - \alpha \right). \quad (15)$$

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 exceedence rates of our quantile estimates are consistent with the true quantile levels. This process of evaluation based on historical data is known as backtesting. A 95% Value-at-Risk estimate should be breached by approximately 5% of observed losses if the observed trials are independent and the estimate is accurate. We can test whether the estimator proposed in Subsection 4.4 is accurate in this sense using the binomial hypothesis test discussed in Subsection 4.1.

We focus on option portfolios whose values are most sensitive to movements in

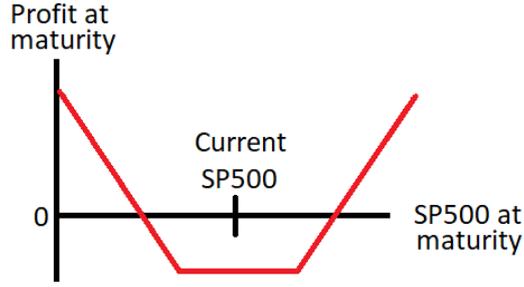


Figure 6.: Profit diagram for an option strangle portfolio.

volatility. An option strangle is a portfolio consisting of an out-of-the-money call option (whose strike is below the current S&P 500 level) and an out-of-the-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 $e^{Y_i(\tau)}$, 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 binomial test of Kupiec (1995) described 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 $t = 1, \dots, T - 1$, a set of random out-of-the-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-the-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 $\kappa_1 > S_t$ is selected, then the put option whose strike is closest to $\kappa_2 = S_t - (\kappa_1 - S_t)$ 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-the-money.

For each $t = 1, \dots, 2012$, 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 13 and 14. Since the actual price movements of these portfolios are known, we can apply the binomial hypothesis test described in Subsection 4.1 to the one-day Value-at-Risk estimator in Equation 15. A separate binomial test is done for each year. 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 . Setting $\mathbf{v} = (v_1, \dots, v_K)^\top$, the benchmark model satisfies

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

for all t . Each v_k is assigned an independent $\text{InvGamma}(a_k, b_k)$ prior. The full conditional distribution for \mathbf{v} is standard, and all other full conditional distributions remain the same with the substitution $\mathbf{h}_t = \log(\mathbf{v})$ for each t .

The results in Figure 7 indicate that the exceedence rates of the stochastic volatility-based VaR estimates are generally closer to the true quantile levels than those of the constant volatility-based VaR estimates. The binomial confidence intervals associated with the stochastic volatility model's VaR estimates include the true level in almost all years, while the intervals associated with the constant volatility model's VaR estimates miss far more frequently. These findings indicate that incorporating stochastic volatility leads to better quantile estimates.

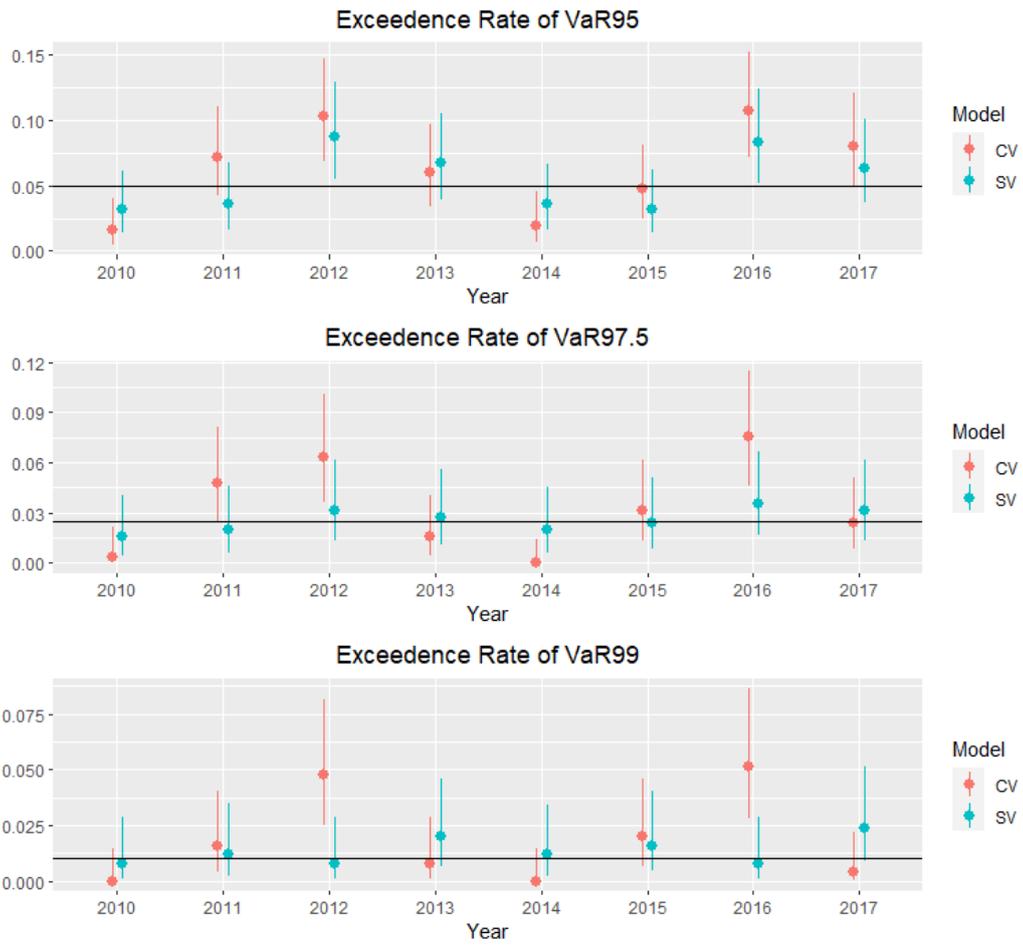


Figure 7.: 95% confidence intervals for the exceedence rates of Value-at-Risk.

5. Conclusion

To review, we introduced an analogue of the familiar stochastic volatility model for functional time series, with the goal of accurately modeling option surfaces. We reduced that functional specification to a finite dimensional vector time series model and discussed a strategy for Bayesian inference that enables forecasting, quantile estimation, and other uncertainty quantification. Motivated by an exploratory analysis that revealed evidence of stochastic volatility, we applied the FSV model to SPX option surfaces and demonstrated through backtesting that the FSV model leads to improved quantile estimates and thus improved Value-at-Risk estimates for option portfolios.

References

- Asai M, McAleer M, Yu J. 2006. Multivariate Stochastic Volatility: A Review. *Econometric Reviews*. 25:145–175.
- Aue A, Horváth L, F Pellatt D. 2017. Functional Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Time Series Analysis*. 38(1):3–21.
- Aue A, Klepsch J. 2017. Estimating Functional Time Series by Moving Average Model Fitting.
- Aue A, Norinho DD, Hörmann S. 2015. On the Prediction of Stationary Functional Time Series. *Journal of the American Statistical Association*. 110(509):378–392.
- Besse PC, Cardot H, Stephenson DB. 2000. Autoregressive Forecasting of Some Functional Climatic Variations. *Scandinavian Journal of Statistics*. 27(4):673–687.
- Black F, Scholes M. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*. 81(3):637–654.
- 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, Engle R, Wooldridge J. 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*. Springer New York. *Lecture Notes in Statistics*.
- 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 DR. 1981. *Time Series: Data Analysis and Theory*. vol. 36. SIAM.
- Bui Quang P, Klein T, Nguyen NH, Walther T. 2018. Value-at-Risk for South-East Asian Stock Markets: Stochastic Volatility vs. GARCH. *Journal of Risk and Financial Management*. 11(2).
- Chen Y, Li B. 2017. An Adaptive Functional Autoregressive Forecast Model to Predict Electricity Price Curves. *Journal of Business & Economic Statistics*. 35(3):371–388.
- Cressie N, Wikle C. 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 RF. 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*. John Wiley & Sons. Wiley finance series.
- Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. 2013. *Bayesian Data Analysis, Third Edition*. Taylor & Francis. Chapman & Hall/CRC Texts in Statistical Science.
- Ghysels E, Harvey A, Renault E. 1996. *Stochastic Volatility*. Centre Interuniversitaire de Recherche en Économie Quantitative, CIREQ. Cahiers de recherche.
- Han CH, Liu WH, Chen TY. 2014. VaR/CVaR Estimation Under Stochastic Volatility Models. *International Journal of Theoretical and Applied Finance*. 17(02):1450009.
- Harvey A, Ruiz E, Shephard N. 1994. Multivariate Stochastic Variance Models. *The Review of Economic Studies*. 61(2):247–264.
- Heston SL. 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 SL, Nandi S. 2000. A Closed-Form GARCH Option Valuation Model. *The Review of Financial Studies*. 13(3):585–625.
- Hörmann S, Horváth L, Reeder R. 2013. A Functional Version of the ARCH Model. *Econometric Theory*. 29(2):267–288.
- Hörmann S, Kidziński Lu, Hallin M. 2014. Dynamic Functional Principal Components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 77(2):319–348.
- Hörmann S, Kokoszka P. 2010. Weakly Dependent Functional Data. *The Annals of Statistics*. 38(3):1845 – 1884.
- Horváth L, Kokoszka P. 2012. *Inference for Functional Data with Applications*. Springer New

- York. Springer Series in Statistics.
- Hsing T, Eubank R. 2015. Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. Wiley. Wiley Series in Probability and Statistics.
- Huang AY. 2015. Value at Risk Estimation by Threshold Stochastic Volatility Model. *Applied Economics*. 47(45):4884–4900.
- Huang SF, Guo M, Chen MR. 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, White A. 1987. The Pricing of Options on Assets with Stochastic Volatilities. *The Journal of Finance*. 42(2):281–300.
- Hyndman R, Booth H. 2008. Stochastic Population Forecasts Using Functional Data Models for Mortality, Fertility and Migration. *International Journal of Forecasting*. 24(3):323–342.
- Hyndman R, Shahid Ullah M. 2007. Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach. *Computational Statistics & Data Analysis*. 51(10):4942–4956.
- Hyndman RJ, Booth H, Yasmeeen F. 2013. Coherent Mortality Forecasting: The Product-Ratio Method With Functional Time Series Models. *Demography*. 50(1):261–283.
- Jang PA, Matteson DS. 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, Frühwirth-Schnatter S. 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, Shephard N, Chib S. 1998. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *The Review of Economic Studies*. 65(3):361–393.
- King MC, Staicu AM, Davis JM, Reich BJ, Eder B. 2018. A Functional Data Analysis of Spatiotemporal Trends and Variation in Fine Particulate Matter. *Atmospheric Environment*. 184:233–243.
- Klepsch J, Klüppelberg C, Wei T. 2017. Prediction of Functional ARMA Processes with an Application to Traffic Data. *Econometrics and Statistics*. 1(C):128–149.

- Kowal DR, Matteson DS, Ruppert D. 2017. A Bayesian Multivariate Functional Dynamic Linear Model. *Journal of the American Statistical Association*. 112(518):733–744.
- Kowal DR, Matteson DS, Ruppert D. 2019. Functional Autoregression for Sparsely Sampled Data. *Journal of Business & Economic Statistics*. 37(1):97–109.
- Kupiec PH. 1995. Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives*. 3(2):73–84.
- Li D, Robinson PM, Shang HL. 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, Ramsay J, Silverman B, Media SS, Silverman H. 2005. *Functional Data Analysis*. Springer. Springer Series in Statistics.
- Ruiz-Medina M, Espejo R, Ugarte M, Militino A. 2013. 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, Klüppelberg C. 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 HL, Booth H, Hyndman R. 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*. London: Chapman & Hall. p. 1–67. 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, Andersen TG. 2009. *Stochastic Volatility: Origins and Overview*. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 233–254.
- 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. Modelling Stochastic Volatility: A Review and Comparative Study. *Mathematical Finance*. 4(2):183–204.
- Vrontos ID, Dellaportas P, Politis DN. 2003. A Full-Factor Multivariate GARCH Model. *The Econometrics Journal*. 6(2):312–334.
- Wood SN. 2011. Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. *Journal of the Royal Statistical Society (B)*. 73(1):3–36.
- Yasmeen F, Sharif M. 2015. Functional Time series (FTS) Forecasting of Electricity Consumption in Pakistan. *International Journal of Computer Applications*. 124:975–8887.
- Yu J, Meyer R. 2006. Multivariate Stochastic Volatility Models: Bayesian Estimation and Model Comparison. *Econometric Reviews*. 25(2-3):361–384.

Appendix A. 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

$$\text{Cov}(\eta_t(\tau), \eta_t(u)) = k_\eta(\tau, u) \quad \forall t.$$

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

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

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

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

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

Appendix B. Sampling the Full Conditional for β

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

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:

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

where

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

and

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

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:

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

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

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

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

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

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

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

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

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

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

we have

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

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

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

we have

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

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

$$\mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_T \end{bmatrix} \sim N_{TK}(\mathbf{0}, I)$$

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

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

and compute

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

Appendix C. 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 $\Sigma_t = \text{diag}(\exp(\mathbf{h}_t))$. Then

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

It is sufficient to look at the quadratic form Ω 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

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

we have

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

After simplifying the terms below,

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

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

we get

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

which is the quadratic form for the multivariate normal below:

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