Sensitivity Diagnostics and Adaptive Tuning of the Multivariate Stochastic Volatility Model

by

Sebastian Baldivieso

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Mathematical Sciences

Dissertation Committee:
Dacian N. Daescu, Chair
Steven Bleiler
Daniel Taylor-Rodriguez
Bin Jiang
Rossitza Wooster
John Gallup

Portland State University
2020

Abstract

New methodologies for diagnostic analysis and adaptive tuning based on sensitivity information of the Multivariate Stochastic Volatility (MSV) model are established in this dissertation. The main focus is on obtaining optimal conditional volatilities from a time series set of financial data observed in the market by specifying a State-Space model with error covariance adaptive tuning of the MSV model. Variational Data Assimilation methods are used in this research as tools for obtaining the optimal *a posteriori* estimates of the multivariate series of volatilities. Calculus of Variations techniques are then applied to a forecast score function to derive the sensitivities of the forecasted volatilities in terms of the input parameters. In summary, this dissertation achieves the development of these new methodologies by

1. Developing the sensitivity information of the multivariate conditional volatilities to observations, covariance specifications and prior estimates,

2. Developing tools for assessing multivariate volatility forecasts. For each time period, sensitivity information provides forecasted volatility diagnostics of the MSV model to give guidance on model performance, and

3. Developing an adaptive tuning procedure based on the multivariate volatility sensitivity information to update the observation error covariance matrix during each assimilation with the main objective of providing improved results in an online manner.

Applications of the new sensitivity diagnostics and adaptive tuning procedures of the MSV model are explored in two experiments. The first experiment is a proof-of-concept experiment where a multivariate series of volatilities is simulated through the specification of a MSV model and serves as a placeholder for true volatilities. The MSV model is then estimated on the resulting time series dataset and the adaptive tuning procedure is performed to demonstrate superior estimation results over the current literature methodologies. In the second experiment, a time series set of Foreign Exchange (FX) rate data is used to estimate the MSV model to provide a time series of conditional volatility estimates of each FX rate. The sensitivity information of each FX rate's conditional volatility forecasts is implemented to derive model performance diagnostics, while the adaptive tuning procedure is implemented to provide improved conditional volatility estimates. Furthermore, an objective assessment and validation of the newly developed methodology is achieved by using an extended data set that is independent on the training set used to calibrate the model.

Dedication

This dissertation is dedicated to my parents, Jorge and Rose, who have sacrificed so much to make this degree a reality. Thank you for always showing me to walk in the ways of the Lord, for encouraging me to pursue my dreams and to be the best person I can be every day.

Acknowledgments

I would like to express my sincere gratitude to my advisor Prof. Dacian Daescu for his continuous support throughout my Ph.D studies and for sharing with me his immense knowledge on data assimilation. His patience, motivation, inspiration and guidance were vital in my development as a mathematician and in the completion of this dissertation. I would give Prof. Dacian Daescu most of the credit for helping me become the mathematician that I am today.

Table of Contents

List of Tables

List of Figures

List of Algorithms

Chapter 1

Introduction

This dissertation establishes new methodologies for diagnostic analysis and adaptive tuning based on sensitivity information of the Multivariate Stochastic Volatility (MSV) model. The main focus is on obtaining optimal conditional volatilities from a time series set of financial data observed in the market by specifying a State-Space model with error covariance adaptive tuning of the MSV model. The following outlines how this dissertation accomplishes its main focus.

Chapter 2 presents an overview of a number of variational data assimilation methods that are used in this research as tools for obtaining the optimal *a posteriori* estimates of the multivariate series of volatilities. The Chapter begins with the fundamental formulation of discrete stochastic differential equations as State-Space models and presents their corresponding probabilistic properties to understand the estimation process that leads to the main framework that is utilized by the data assimilation methods. The Kalman filter method, based on the works of Rudolf Kalman [36], is the first data assimilation method that is reviewed in the Chapter along with a discussion of their fundamental properties. The Kalman smoother algorithm is also presented based on a fixed interval smoothing algorithm and it provides a variant of the Kalman filter as it incorporates all data during a data assimilation window, as opposed to a single iteration as presented in the Kalman filter. These algorithms are well suited when the observation and state equations in the State-Space models are

linear; that is, these algorithms provide optimal results. For the nonlinear case, an extended Kalman filter algorithm is presented as means of overcoming the nonlinearity, however, the estimates are now sub-optimal estimates. The 4D-Var and the weak 4D-Var algorithms are presented as an extension to the extended Kalman filter for the nonlinear case and their probabilistic properties, along with their relationship to the extended Kalman filter and smoother, are also presented in the Chapter. An overview of the Particle filter algorithm that is based on a Markov chain Monte Carlo (MCMC) alternative to the direct optimization scheme is presented at the end of the Chapter.

Chapter 3 introduces the concept of stochastic volatility as it is understood by the financial mathematics and financial econometrics literature. The Chapter begins with an overview of the fundamental option pricing theory based on the work of Black and Scholes [10] and provides an alternative pricing formulae when the simplistic assumptions of having constant variance (volatility) of the underlying asset is extended from the Black-Scholes assumptions. The univariate Stochastic Volatility (SV) model considered for estimation in this dissertation is presented from a continuous-time process and their direct discrete multivariate extensions is formulated. From the multivariate extensions of the SV model, two MSV models are considered: an MSV model with no-leverage effects (the time series and volatilities are not correlated) and an MSV model with leverage effects. For this dissertation, the MSV model with no-leverage effects is selected for modeling multivariate stochastic volatilities.

Chapter 4 presents the Calculus of Variations techniques that are applied to derive the sensitivities of the forecasted volatilities in terms of the input parameters. The Chapter begins by deriving the sensitivity information of a variational data assimilation system to the various input parameters. Then, the sensitivity information is extended to derive the sensitivities of a forecast score function of the multivari-

ate conditional volatilities with respect to observations, covariance specifications and prior estimates. The sensitivity information provides the necessary tools for assessing the multivariate volatility forecasts from the model. For each time period, sensitivity information provides forecasted volatility diagnostics of the MSV model to give guidance on model performance and guidance on adaptive tuning. A new adaptive tuning procedure is derived on the multivariate volatility sensitivity information to update the observation error covariance matrix during each assimilation with the main objective of providing improved results in an online manner.

Chapter 5 presents the applications of the newly developed sensitivity diagnostics and adaptive tuning procedures of the MSV model in two experiments. The first experiment is a proof-of-concept experiment where a multivariate series of volatilities is simulated through the specification of a MSV model and serves as a placeholder for true volatilities. The MSV model is then estimated on the resulting time series dataset and the adaptive tuning procedure is performed to demonstrate superior estimation results over the current literature methodologies. In the second experiment, a time series set of Foreign Exchange (FX) rate data is used to estimate the MSV model to provide a time series of conditional volatilities of each FX rate. The sensitivity information of each FX rate's conditional volatility forecasts is employed to derive model performance diagnostics, while the adaptive tuning procedure is implemented to provide improved conditional volatility estimates. Furthermore, an objective assessment and validation of the newly developed methodology is achieved by using an extended data set that is independent on the training set used to calibrate the model.

Chapter 6 presents the conclusions observed during the development and application of the sensitivity analysis and adaptive tuning of the MSV model. Future research directions are also discussed in this Chapter to extend these methodologies to more advance data assimilation methods and complex industry problems.

Chapter 2

Overview of Data Assimilation Methods

Data assimilation methods are concerned with the estimation of the state of a physical process described by stochastic dynamical systems. The general modeling approach involves the assumption of two equations, where one equation describes the dynamics of the true state and the other equation describes the dynamics of the observations given the states. The estimation approach attempts to recover the true state from a set of noisy observations taken on the state variable. Data assimilation methods attempt to provide an optimal estimate (analysis) of the evolving state of the system by incorporating all available sources of information: observational data, an a priori estimate - typically produced by a model forecast, and the associated error statistics. The following sections of this chapter provide an overview of well-established data assimilation methods and their fundamental properties.

2.1 General State-Space Model

The main references for the derivation presented in this section are [34], [36] and [37]. Consider the discrete stochastic difference equation for the state $\mathbf{x}$

$$\mathbf{x}_{k+1} = \mathcal{M}_{k+1}(\mathbf{x}_k) + \Gamma_k(\mathbf{x}_k)\mathbf{w}_{k+1}, \qquad k = 0, 1, 2, ... \tag{2.1}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the state at time t_k, $\mathcal{M}_{k+1}(*)$ is an $n-$vector valued function (linear or nonlinear) that transitions the state $\mathbf{x}_k$ to $\mathbf{x}_{k+1}$, Γ_k is an $n \times r$ matrix function called the state-disturbance-loading matrix describing how the states $\mathbf{x}$ at period t_k combine with the state errors at period t_k and $\mathbf{w}_k \in \mathbb{R}^r$ is a sequence of white Gaussian vectors with $\mathbf{w}_k \sim N(\mathbf{0}, \mathbf{Q}_k)$, for each k that describe the state errors at time t_k. The distribution of the initial condition $\mathbf{x}_0$ is assumed to be known and furthermore, statistically independent of the process $\{\mathbf{w}_k\}_{k\geq 1}$. Given $\mathbf{x}_k$, the state $\mathbf{x}_{k+1}$ depend only on $\mathbf{w}_{k+1}$, which is independent of $\mathbf{x}_{k-1}, ..., \mathbf{x}_0$. Therefore, the solution $\{\mathbf{x}_k\}_{k\geq 0}$ to (2.1) is a Markov process.

Let $\mathbf{y}_k \in \mathbb{R}^m$ represent (denote) discrete, noisy observations taken on the state $\mathbf{x}_k$ at time t_k and suppose that the model for the observations is given by

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k, \qquad k = 1, 2, ... \tag{2.2}$$

where $\mathbf{h}_k()$ is an $m-$vector valued (observation operator) function describing how the observations relate to the states at time t_k and $\{\mathbf{v}_k\}_{k\geq 1}$ is a vector sequence of Gaussian white noise, with $\mathbf{v}_k \sim N(\mathbf{0}, \mathbf{R}_k)$ for each k. For simplicity, it is assumed that $\{\mathbf{v}_k\}_{k\geq 1}$ and $\{\mathbf{w}_k\}_{k\geq 1}$ are statistically independent, such that, there is no correlation amongst these two stochastic processes.

Now, consider having a set of realizations of the noisy observations $\mathbf{y}_k$. That is, consider having the information set $F_l = \{\mathbf{y}_1, ..., \mathbf{y}_l\}$ given by measurements of the system. The problem of computing the estimates $\mathbf{x}_k$ given the information set F_l, can be classified as follows:

1. The problem is called a **discrete smoothing** problem, if $k < l$, for each k.

2. The problem is called a **discrete filtering** problem, if $k = l$, and

3. The problem is called a **discrete prediction** problem, if $k > l$.

2.2 Probabilistic Approach

In this section, the posterior probability distribution is derived for the discrete problem (2.1) - (2.2) which will subsequently be a key ingredient to the optimal state estimation theory. Consider again the problem (2.1) - (2.2) and the information set $F_N = \{\mathbf{y}_1, ..., \mathbf{y}_N\}$. The solution of the estimation problem is a sequence $\{\hat{\mathbf{x}}_0, ..., \hat{\mathbf{x}}_N\}$ that maximizes the conditional probability density function $p(\mathbf{x}_0, ..., \mathbf{x}_N | \mathbf{y}_1, ..., \mathbf{y}_N)$. Using Bayes' rule for conditional densities

$$p(\mathbf{x}_0, ..., \mathbf{x}_N | \mathbf{y}_1, ..., \mathbf{y}_N) = \frac{p(\mathbf{y}_1, ..., \mathbf{y}_N | \mathbf{x}_0, ..., \mathbf{x}_N) p(\mathbf{x}_0, ..., \mathbf{x}_N)}{p(\mathbf{y}_1, ..., \mathbf{y}_N)} \tag{2.3}$$

Recall that, by assumption, $\mathbf{w}_k \sim N(\mathbf{0}, \mathbf{Q}_k)$, $\mathbf{v}_k \sim N(\mathbf{0}, \mathbf{R}_k)$, where $\mathbf{Q}_k, \mathbf{R}_k > 0$ and $\{\mathbf{w}_k\}_{k \geq 1}$ is statistically independent and $\{\mathbf{v}_k\}_{k \geq 1}$ is also statistically independent. Therefore, the probability distribution of the observations given the states is given by

$$p(\mathbf{y}_1, ..., \mathbf{y}_N | \mathbf{x}_0, ..., \mathbf{x}_N) = \prod_{k=1}^{N} p_{\mathbf{v}_k}(\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k)) \tag{2.4}$$

Where $p_{\mathbf{v}_k}(*)$ is the normal probability distribution of $\mathbf{v}_k$. Next, by the product law of probability and the Markov property of the state equation,

$$\begin{aligned} p(\mathbf{x}_0, ..., \mathbf{x}_N) &= p(\mathbf{x}_n | \mathbf{x}_1, ..., \mathbf{x}_{N-1}) \cdot p(\mathbf{x}_{N-1} | \mathbf{x}_1, ..., \mathbf{x}_{N-2}) \cdot ... \cdot p(\mathbf{x}_2 | \mathbf{x}_0, \mathbf{x}_1) \cdot p(\mathbf{x}_1 | \mathbf{x}_0) p(\mathbf{x}_0) \\ &= p(\mathbf{x}_0) \prod_{k=1}^{N} p_{\Gamma_k(\mathbf{x}_k)\mathbf{w}_k}(\mathbf{w}_k - \mathcal{M}_k(\mathbf{x}_{k-1})) \end{aligned} \tag{2.5}$$

Where $p_{\Gamma_k(\mathbf{x}_k)\mathbf{w}_k}(*)$ is the normal probability distribution of $\mathbf{w}_k$ scaled by the matrix $\Gamma_k(\mathbf{x}_k)$. Equality in equation (2.5) was obtained through the assumption that the

states are independent (see equation (5.28) in Jazwinski [34]) Since the probability distributions are normal, as well as the initial condition is also assumed to be normally distributed with $\mathbf{x}_0 \sim N(\mathbf{0}, \mathbf{P}_0)$, equation (2.3) becomes (up to the normalization constant)

$$\begin{aligned} p(\mathbf{x}_0, ..., \mathbf{x}_N | F_N) &= c \exp\Big\{ -\frac{1}{2}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)\mathbf{P}_0^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T \\ &- \frac{1}{2}\sum_{k=1}^{N}(\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k)\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k))^T \\ &- \frac{1}{2}\sum_{k=1}^{N}(\mathbf{x}_k - \mathcal{M}_k(\mathbf{x}_{k-1}))(\Gamma_k \mathbf{Q}_k \Gamma_k^T)^{-1}(\mathbf{x}_k - \mathcal{M}_k(\mathbf{x}_{k-1}))^T \Big\} \end{aligned} \quad (2.6)$$

Taking the negative log of (2.6) such that the problem becomes one of minimizing the loglikelihood (Bayesian) function, the functional to minimize is given by

$$\begin{aligned} \min_{\{\mathbf{x}_0,...,\mathbf{x}\}} J_N &= \frac{1}{2}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)\mathbf{P}_0^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T \\ &\frac{1}{2}\sum_{k=1}^{N}(\mathbf{y}_k - \mathbf{h}_k(x_k)\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{h}_k(x_k))^T \\ &\frac{1}{2}\sum_{k=1}^{N}(\mathbf{x}_k - \mathcal{M}_k(\mathbf{x}_{k-1}))(\Gamma_k \mathbf{Q}_k \Gamma_k^T)^{-1}(\mathbf{x}_k - \mathcal{M}_k(\mathbf{x}_{k-1}))^T \Big\} \end{aligned} \quad (2.7)$$

Equation (2.7) has a close relationship to the weak 4D-Var problem, as it will be shown in later sections. Variants of (2.7) also give rise to different data assimilation problems, such as the Kalman filter and the 4D-Var.

2.3 The Kalman Filter

Mathematician Rudolf Kalman [36] solved the estimation problem of State-Space models when the equations are linear. The celebrated solution is called the Kalman filter. In this section, the case where the equations are linear are considered to show

how the Kalman filter solution solves a maximum posterior probability problem.

Consider the linear State-Space model

$$\mathbf{x}_{k+1} = \mathbf{M}_{k+1}\mathbf{x}_k + \Gamma_k \mathbf{w}_{k+1}, \qquad k = 0, 1, ... \tag{2.8}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the state at time t_k, $\mathbf{M}_k \in \mathbb{R}^{n\times n}$ is a nonsingular *state transition matrix*, $\Gamma_k \in \mathbb{R}^{n\times r}$ and $\{\mathbf{w}_k\}_{k\geq 1}$ is an r-dimensional vector sequence of white Gaussian noises with $w_k \sim N(\mathbf{0}, \mathbf{Q}_k)$ and $\mathbf{Q}_k > 0$. The discrete, linear observations are given by

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k, \qquad k = 1, 2... \tag{2.9}$$

where $\mathbf{H}_k \in \mathbb{R}^{m\times n}$ is nonrandom matrix that relates the states to the observations and $\{\mathbf{v}_k\}_{k\geq 1}$ is an m-dimensional vector sequence of white Gaussian noises, with $\mathbf{v}_k \sim N(\mathbf{0}, \mathbf{R}_k)$ and $\mathbf{R}_k > 0$. Furthermore, the distribution of the initial condition $\mathbf{x}_0$, at initial time t_0, is Gaussian with $\mathbf{x}_0 \sim N(\hat{\mathbf{x}}_0, \mathbf{P}_0)$ and $\mathbf{x}_0, \{\mathbf{v}_k\}_{k\geq 1}, \{\mathbf{w}_k\}_{k\geq 1}$ are all assumed to be statistically independent.

Direct computations from equations (2.8)-(2.9) give the conditional mean and conditional variance. From equation (2.8), given the information set $F_k = \{\mathbf{y}_1, ..., \mathbf{y}_k\}$, the conditional mean becomes

$$\hat{\mathbf{x}}_{k+1|k} = \mathcal{E}\{\mathbf{x}_{k+1}|F_k\} = \mathbf{M}_{k+1}\hat{\mathbf{x}}_{k|k},$$

where the notation $\hat{\mathbf{x}}_{l|k} = \mathcal{E}\{\mathbf{x}_l|F_k\}$ is adopted to denote the conditional expectation of $\mathbf{x}_l$ given the information set F_k. From equation (2.8), the conditional variance may also be derived by subtracting the mean, taking the square and computing the

expectation

$$\begin{aligned}
\mathbf{P}_{k+1|k} &= \mathcal{E}\{(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})^T | F_k\} \\
&= \mathcal{E}\{[\mathbf{M}_{k+1}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}) + \Gamma_{k+1}\mathbf{w}_{k+1}][\mathbf{M}_{k+1}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}) + \Gamma_{k+1}\mathbf{w}_{k+1}]^T | F_k\} \\
&= \mathbf{M}_{k+1}\mathbf{P}_{k|k}\mathbf{M}_{k+1}^T + \Gamma_{k+1}\mathbf{Q}_{k+1}\Gamma_{k+1}^T
\end{aligned}$$

The discrete linear State-Space model is summarized in the following theorem [34] and it is the much celebrated Kalman filter.

Theorem 2.3.1. *The optimal (minimum variance) filter for the discrete linear State-Space model of equations (2.8)-(2.9) consists of the following difference equations of the conditional mean and conditional variance. Between observations,*

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{M}_{k+1}\hat{\mathbf{x}}_{k|k}, \tag{2.10}$$

$$\mathbf{P}_{k+1|k} = \mathbf{M}_{k+1}\mathbf{P}_{k|k}\mathbf{M}_{k+1}^T + \Gamma_k\mathbf{Q}_{k+1}\Gamma_k^T. \tag{2.11}$$

At observations,

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_{k|k-1}), \tag{2.12}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} \tag{2.13}$$

where

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T[\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k]^{-1} \tag{2.14}$$

*is the **Kalman Gain**. Prediction for $t_l > t_k$ $(\hat{\mathbf{x}}_{l|k}, \mathbf{P}_{l|k})$ is accomplished via equation (2.8) with initial condition $(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k})$.*

In his original publication, Kalman derived the Kalman filter by making use of orthogonal projections. There are many other ways of deriving the Kalman filter,

see for example [23] where multivariate normal distribution relationships are used to derive the filter.

In order to give the presentation of the Kalman filter as a minimization problem, it may be considered as a maximization problem of a one loop iteration of the posterior conditional distribution

$$p(\mathbf{x}_k|F_k) \tag{2.15}$$

It is noted that many have used maximum likelihood estimation to derive the Kalman filter including Ho [30], Schmidt [48] and Jazwinski [34]. The basic ideas shall be presented while preserving a probabilistic perspective. At each observation $\mathbf{x}_k$, the posterior distribution is given by

$$p(\mathbf{x}_k|F_k) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|F_{k-1})}{p(\mathbf{y}_k|F_{k-1})} \tag{2.16}$$

Where the corresponding distributions are given by

$$\mathbf{y}_k|\mathbf{x}_k \sim N(\mathbf{H}_k\mathbf{x}_k, \mathbf{R}_k) \tag{2.17}$$

$$\mathbf{x}_k|F_{k-1} \sim N(\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \tag{2.18}$$

$$\mathbf{y}_k|F_{k-1} \sim N(\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}, \underbrace{\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k}_{\mathbf{v}_{k|k-1}}) \tag{2.19}$$

Notice that because all of the distributions are normal, it is straight forward to obtain their corresponding distributions as they are characterized by their first and second moments, which it has been done above.

Using these equations we obtain (up to a constant) the following posterior distribution

$$
\begin{aligned}
p(\mathbf{x}_k|F_k) &= c' \exp\Big\{ -\frac{1}{2}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k)^T\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k) \\
&\quad -\frac{1}{2}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T(\mathbf{P}_{k|k-1})^{-1}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \\
&\quad +\frac{1}{2}(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})^T(\mathbf{V}_{k|k-1})^{-1}(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})\Big\}
\end{aligned}
$$

Maximization of the posterior distribution is the same as minimization (with respect to $\mathbf{x}_k$) of the following cost functional

$$
J = \frac{1}{2}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k)^T\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k) + \frac{1}{2}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T(\mathbf{P}_{k|k-1})^{-1}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \tag{2.20}
$$

Setting the gradient of J with respect to $\mathbf{x}_k$ equal to zero gives

$$
-\mathbf{H}_k^T\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k) + (\mathbf{P}_{k|k-1})^{-1}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) = \mathbf{0}
$$

The solution to the optimality condition is denoted by $\hat{\mathbf{x}}_{k|k}$ and it is the same as equation (2.12). Once the optimal state has been obtained, direct computation of $\mathbf{P}_{k|k} = \mathcal{E}\{(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^T|F_k\}$ gives equation (2.13). It has been shown that the Kalman filter is a solution to a maximum posterior probability problem and from equation (2.20), it can be seen that when the errors are normally distributed, the Kalman filter is indeed the solution to an optimization problem. Notice that this variational cost functional resembles that of equation (2.7) where the functional here assumes no model errors ($\mathbf{w}_k = \mathbf{0}$) and only one iteration is taken into account during the optimization process.

2.4 The Kalman Smoother

The Kalman smoother is a variation of the Kalman filter. While the Kalman filter in equations (2.12)-(2.13) uses an update equation and a one step ahead forecast as the information becomes available, the Kalman smoother uses all possible information already available and attempts to recover the state variables based on the batch of information given. More specifically, the Kalman filter uses equation (2.12)-(2.13) to filter through the state $\hat{\mathbf{x}}_{k|k}$ as the information set F_k becomes available for each k, and then uses the model equations to get the one step ahead forecast $\hat{\mathbf{x}}_{k+1|k}$ of the state variable. It iteratively does this for $k = 1, ..., N$, where t_N is the final forecast horizon. The Kalman smoother, on the other hand, uses all of the available information in the information set F_N, to recover the states $\hat{\mathbf{x}}_{k|N}$, for $k = N, N-1, ..., 1$.

Following Tsay [53], the smoothed state variables are defined as

$$\hat{\mathbf{x}}_{k|N} = \mathcal{E}\{\mathbf{x}_k|F_N\} = \mathcal{E}\{\mathbf{x}_k|\mathbf{y}_1, ..., \mathbf{y}_N\} \tag{2.21}$$

and they are given by

$$\hat{\mathbf{x}}_{k|N} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{P}_{k|k-1}\mathbf{q}_{k-1} \tag{2.22}$$

Where $\mathbf{q}_k$ is defined as

$$\begin{aligned}
\mathbf{q}_N &= \mathbf{0} \\
\mathbf{q}_{N-1} &= \mathbf{H}_N^{\mathrm{T}}\mathbf{V}_N^{-1}\mathbf{r}_N \\
\mathbf{q}_{N-2} &= \mathbf{H}_{N-1}^{\mathrm{T}}\mathbf{V}_{N-1}^{-1}\mathbf{r}_{N-1} + \mathbf{L}_{N-1}^{\mathrm{T}}\mathbf{H}_N^{\mathrm{T}}\mathbf{V}_N^{-1}\mathbf{r}_N \\
&\vdots \\
\mathbf{q}_{k-1} &= \sum_{s=k}^{N}\left(\prod_{j=k}^{s-1}\mathbf{L}_j^{\mathrm{T}}\right)\mathbf{H}_s^{\mathrm{T}}\mathbf{V}_s^{-1}\mathbf{r}_s, \qquad \text{for } k = N-2, N-3, ..., 1
\end{aligned}$$

Where $\mathbf{r}_k = \mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}$, $\mathbf{V}_k = \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k$ and $\mathbf{L}_k = \mathbf{M}_k - \mathbf{K}_k\mathbf{H}_k$.

The Kalman smoother algorithm therefore proceeds as follows. Compute the Kalman filter estimates $\hat{\mathbf{x}}_{k|k}$ using equations (2.12)-(2.13) for each $t_k \in [t_1, t_N]$ during the data assimilation window. Then, compute the Kalman smoother using the backwards recursion

$$\mathbf{q}_{k-1} = \mathbf{H}_k^{\mathrm{T}}\mathbf{V}_k^{-1}\mathbf{r}_k + \mathbf{L}_k^{\mathrm{T}}\mathbf{q}_k \tag{2.23}$$

$$\hat{\mathbf{x}}_{k|N} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{P}_{k|k-1}\mathbf{q}_{k-1} \tag{2.24}$$

for $k = N, N-1, ..., 1$ where $\mathbf{q}_N = \mathbf{0}$. It is noted that in the data assimilation literature, this algorithm is referred to as the *fixed interval smoothing* (see for example de Jong [21]).

For completeness, the state covariance matrix under the Kalman smoother algorithm is also presented. The smoothed covariance is denoted as

$$\mathbf{P}_{k|N} = \mathcal{E}\{(\mathbf{x}_k - \hat{\mathbf{x}}_{k|N})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|N})^{\mathrm{T}}|F_N\} \tag{2.25}$$

and the smoothed state covariance can be calculated as follow:

$$\mathbf{Z}_{k-1} = \mathbf{H}_k^{\mathrm{T}}\mathbf{V}_k^{-1}\mathbf{H}_k + \mathbf{L}_k^{\mathrm{T}}\mathbf{Z}_k\mathbf{L}_k \tag{2.26}$$

$$\mathbf{P}_{k|N} = \mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{Z}_{k-1}\mathbf{P}_{k|k-1} \tag{2.27}$$

for $k = N, N-1, ..., 1$ with $\mathbf{Z}_N = \mathbf{0}$. As with the state smoothing algorithm, one first obtains a forward pass of the state covariances using the Kalman filter equations then uses equations (2.26)-(2.27) to obtain a backward pass of the smoothed state covariances.

2.5 The Extended Kalman Filter

The goal of the Extended Kalman filter is to overcome the problems of nonlinearity in the equations by replacing the nonlinear functions in the state and observation equations by its corresponding linearization. The books of Jazwinski [34] and Anderson [3] provide a complete derivation of the Extended Kalman filter.

Consider the nonlinear State-Space model

$$\mathbf{x}_{k+1} = \mathcal{M}_{k+1}(\mathbf{x}_k) + \Gamma_k(\mathbf{x}_k)\mathbf{w}_{k+1}, \qquad k = 0, 1, ... \tag{2.28}$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \qquad k = 1, 2, ... \tag{2.29}$$

where $\mathcal{M}(*)$, $\Gamma(*)$ and $\mathbf{h}(*)$ are now nonlinear functions. The idea of the Extended Kalman filter is to replace these nonlinear functions with their corresponding linear approximations and proceed as in the linear case to use the Kalman filter. Proceeding formally, the linear approximations about the conditional mean $\hat{\mathbf{x}}_{k|k}$ and $\hat{\mathbf{x}}_{k|k-1}$ are

$$\begin{aligned} \mathcal{M}_{k+1}(\mathbf{x}_k) &\approx \mathcal{M}_{k+1}(\hat{\mathbf{x}}_{k|k}) + \mathbf{M}_{k+1}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}) \\ \mathbf{h}_k(\mathbf{x}_k) &\approx \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1}) + \mathbf{H}_k(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \end{aligned}$$

were $\mathbf{M}_{k+1}$ and $\mathbf{H}_k$ are the Jacobian matrix of first order derivatives of $\mathcal{M}_{k+1}(*)$ and $\mathbf{h}_k(*)$ of the model and observation operator, respectively. That is, the Jacobian matrices are calculated as

$$\mathbf{M}_{k+1} = \frac{\partial \mathcal{M}_{k+1}}{\partial \mathbf{x}}\Big|_{\hat{\mathbf{x}}_{k|k}} \tag{2.30}$$

$$\mathbf{H}_k = \frac{\partial \mathbf{H}_k}{\partial \mathbf{x}}\Big|_{\hat{\mathbf{x}}_{k|k-1}} \tag{2.31}$$

$$\tag{2.32}$$

Substituting these in, the nonlinear model becomes

$$\mathbf{x}_{k+1} = \mathbf{M}_{k+1}\mathbf{x}_k + \Gamma_k \mathbf{w}_{k+1} + \mathbf{u}_k, \qquad k = 0, 1, ... \tag{2.33}$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k + \mathbf{z}_k, \qquad k = 1, 2, ... \tag{2.34}$$

where $\Gamma_k = \Gamma_k(\hat{\mathbf{x}}_{k|k})$, $\mathbf{u}_k = \mathcal{M}_{k+1}(\hat{\mathbf{x}}_{k|k}) - \mathbf{M}_{k+1}\hat{\mathbf{x}}_{k|k}$ and $\mathbf{z}_k = \mathbf{h}_k(\hat{\mathbf{x}}_{k|k-1}) - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}$ are calculated online. Thus, the nonlinear models were transformed to a linear model and the Kalman filter algorithm may be implemented to solve for the conditional mean and covariance matrix.

2.6 4D-Var

Consider the nonlinear State-Space model

$$\mathbf{x}_{k+1} = \mathcal{M}_{k+1}(\mathbf{x}_k) \qquad k = 0, 1, ... \tag{2.35}$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \qquad k = 1, 2... \tag{2.36}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ and $\mathbf{y}_k \in \mathbb{R}^m$ are the state and observation at time $t_k \in [t_0, t_N]$ and $\mathcal{M}_{k+1}(*)$ and $\mathbf{h}_k(*)$ are nonlinear functions. It is assumed that the state equation has no errors and that the observation errors are a vector white noise sequence with $\mathbf{v}_k \sim N(0, \mathbf{R}_k)$. It is also assumed that the initial condition is normally distributed with $\mathbf{x}_0 \sim N(\hat{\mathbf{x}}_0^b, \mathbf{B})$, where $\mathbf{B} \in \mathbb{R}^{n \times n}$ is called the background covariance matrix and $\hat{\mathbf{x}}_0^b$ is the state background estimate.

It was shown in Section 2.2 that the solution of the estimates $\mathbf{x}_k$ may be obtained via maximum likelihood (Bayesian) estimation. Using the model equations (2.35) as

constraints, the analysis is obtained by the minimization of

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \hat{\mathbf{x}}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0^b) + \frac{1}{2}\sum_{k=0}^{N}(\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k))^T \mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k)) \quad (2.37)$$

The solution to this minimization problem is denoted by $\mathbf{x}_0^a$ and is called the analysis of $\mathbf{x}_0$. The remaining analysis estimates $\mathbf{x}_k^a$ can be obtained via equation (2.35) by evaluating the model since the model is "perfect". Minimization of the cost functional (2.37) amounts to the minimization of the fit of the observation data and the prior estimate $\hat{\mathbf{x}}_0^b$. The method of obtaining the best estimates through a minimization problem is called **variational method** and the minimization of the cost functional (2.37) is called **four dimensional variational data assimilation** (4D-Var).

2.7 Weak 4D-Var

Consider the same nonlinear State-Space model described in the 4D-Var problem but now, it is assumed that the model is not perfect and hence, there exists state modeling errors,

$$\mathbf{x}_{k+1} = \mathcal{M}_{k+1}(\mathbf{x}_k) + \mathbf{w}_{k+1} \qquad k = 0, 1, ... \quad (2.38)$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \qquad k = 1, 2, ... \quad (2.39)$$

where the errors $\{\mathbf{w}_k\}_{k\geq 1}$ is an n-vector sequence of Gaussian white noise with $\mathbf{w}_k \sim N(\mathbf{0}, \mathbf{Q}_k)$. Taking once again the probabilistic approach outlined in Section 2.2, the

minimization problem now becomes

$$J(\mathbf{x}_0, ..., \mathbf{x}_N) = \frac{1}{2}(\mathbf{x}_0 - \hat{\mathbf{x}}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \hat{\mathbf{x}}_0^b) + \frac{1}{2}\sum_{k=1}^{N}(\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k))^T \mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k)) + \frac{1}{2}\sum_{k=1}^{N}(\mathbf{x}_k - \mathcal{M}_k(\mathbf{x}_{k-1}))^T \mathbf{Q}_k^{-1}(\mathbf{x}_k - \mathcal{M}_k(\mathbf{x}_{k-1})) \quad (2.40)$$

Notice that since the model (2.38) is no longer perfect, minimization with respect to $\mathbf{x}_0^b$ and using the model equation (2.38) to obtain the remaining estimates is no longer valid. If one proceeds this way, each new iteration will carry out an error from the previous iteration and soon enough errors add up. Therefore, it is necessary that the minimization of J be done with respect to the state sequence $\{\mathbf{x}_0, ..., \mathbf{x}_N\}$.

The solution to the minimization problem (2.40) gives the analysis of the states $\{\mathbf{x}_0^a, ..., \mathbf{x}_N^a\}$. Prediction to the next step t_{N+1} is done via equation (2.38) with $w_{N+1} = 0$. That is, the best prediction occurs when there are no model errors. Finally, the method of minimizing the cost functional $J(\mathbf{x}_0, ..., \mathbf{x}_N)$ is called the **weak 4D-Var**, since the model equations $\mathbf{x}_{k+1} = \mathcal{M}_{k+1}(\mathbf{x}_k) + \mathbf{w}_{k+1}$ are now imposed "weakly" into the cost functional.

2.8 Properties of Variational Data Assimilation Methods

In the publication of Li and Navon [39], they showed the properties of the variational data assimilation methods as well as its relationship to the Kalman filter and smoother. In this section, we will provide a brief overview to the solutions of the 4D-Var and weak 4D-Var and how the 4D-Var is intimately related to the Kalman solution. For more information, we direct the reader to the paper of Li and Navon.

To this end, consider again the linear State-Space model

$$\mathbf{x}_{k+1} = \mathbf{M}_{k+1}\mathbf{x}_k + \Gamma_k \mathbf{w}_{k+1}, \qquad k = 0, 1, 2, ... \tag{2.41}$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \qquad k = 1, 2, ... \tag{2.42}$$

where $\mathbf{x}_k$ is an $n-$vector of states, $\mathbf{M}_{k+1}$ is the $n \times n$ transition matrix that transitions the states from t_k to t_{k+1}, Γ_k is an $n \times r$ matrix and $\{\mathbf{w}_k\}_{k\geq 1}$, $\{\mathbf{v}_k\}_{k\geq 1}$ are r- vector and m- vector, respectively, white sequences such that

$$\begin{aligned} \mathbf{w}_k &\sim N(\mathbf{0}, \mathbf{Q}_k), \\ \mathbf{v}_k &\sim N(\mathbf{0}, \mathbf{R}_k) \end{aligned}$$

where $\mathbf{Q}_k > 0$, $\mathbf{R}_k > 0$ are the covariance matrices, respectively. It is also assumed that the initial condition has distribution

$$\mathbf{x}_0 \sim N(\mathbf{x}_0^b, \mathbf{B}_0)$$

where $\mathbf{x}_0^b$ is the background state estimate and $\mathbf{B}_0$ is the background error covariance matrix at time t_0. We also assume that $\{\mathbf{w}_k\}_{k\geq 1}$ and $\{\mathbf{v}_k\}_{k\geq 1}$ are statistically independent and that $\{\mathbf{w}_k\}_{k\geq 1}$ is independent of $\mathbf{x}_0$.

Assume first a perfect model, i.e., there are no model errors ($\mathbf{w}_k = \mathbf{0}$). As shown before, the cost function in a 4D Variational data assimilation set up during the assimilation window $[t_0, t_N]$ is given by

$$J = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{k=1}^{N}(\mathbf{H}_k\mathbf{x}_k - \mathbf{y}_k)^T \mathbf{R}_k^{-1}(\mathbf{H}_k\mathbf{x}_k - \mathbf{y}_k) \tag{2.43}$$

In the standard 4D-Var analysis, the minimization of the cost function is done with

respect to $\mathbf{x}_0$. As such, one instead has

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{k=1}^{N}(\mathbf{H}_k\mathbf{M}(k,0)\mathbf{x}_0 - \mathbf{y}_k)^T \mathbf{R}_k^{-1}(\mathbf{H}_k\mathbf{M}(k,0)\mathbf{x}_0 - \mathbf{y}_k) \tag{2.44}$$

where $\mathbf{M}(k,i) = \mathbf{M}(k,k-1)...\mathbf{M}(i+1,i)$ is obtained since

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{M}(k+1,k)\mathbf{x}_k \\ &= \mathbf{M}(k+1,k)\mathbf{M}(k,k-1)\mathbf{x}_{k-1} \\ &\vdots \\ &= \underbrace{\mathbf{M}(k+1,k)\mathbf{M}(k,k-1)...\mathbf{M}(1,0)}_{\mathbf{M}(k,0)}\mathbf{x}_0 \end{aligned}$$

and where the notation $\mathbf{M}_{k+1} = \mathbf{M}(k+1,k)$ is implemented to denote the state transition matrix that transitions the state from t_k to t_{k+1}. Lorenc [40] and Thepaut and Courtier [52] presented the analytical solution to equation (2.44). It is important to emphasize that the analysis covariance matrix is given by the inverse of the Hessian, $\mathcal{H}$, of the cost function (see Rabier and Courtier [47]).

$$\mathbf{P}_0^a = \mathcal{E}\{(\mathbf{x}_0^a - \mathbf{x}_0^t)(\mathbf{x}_0^a - \mathbf{x}_0^t)^T\} = \mathcal{H}_{0,N}^{-1} \tag{2.45}$$

The solution to the optimization problem (4D-Var) is denoted by $\hat{\mathbf{x}}_0$ and it satisfies the optimality condition of equation (2.44) $\nabla J(\hat{\mathbf{x}}_0) = \mathbf{0}$. This solution is expressed as

$$\hat{\mathbf{x}}_0 = \mathbf{x}_0^b - \mathcal{H}_{0,N}^{-1}\nabla_{\mathbf{x}_0} J(\mathbf{x}_0^b) \tag{2.46}$$

where $\mathcal{H}_{0,N}$ is the Hessian of the cost function (2.44) evaluated at $\mathbf{x}_0$ and it is given

by

$$\mathcal{H}_{0,N} = \mathbf{B}_0^{-1} + \sum_{k=1}^{N} \mathbf{M}^T(k,0)\mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{H}_k\mathbf{M}(k,0) \tag{2.47}$$

and where $\nabla_{\mathbf{x}_0} J(\mathbf{x}_0^b)$ is the gradient of the of cost function (2.44) evaluated at $\mathbf{x}_0^b$ and it is given by

$$\nabla_{\mathbf{x}_0} J(\mathbf{x}_0^b) = \sum_{k=1}^{N} \mathbf{M}^T(k,0)\mathbf{H}_k^T\mathbf{R}_k^{-1}(\mathbf{y}_k^b - \mathbf{y}_k) \tag{2.48}$$

Here $\mathbf{y}_k^b$ is defined as

$$\mathbf{y}_k^b = \mathbf{H}_k\mathbf{x}_k^b = \mathbf{H}_k\mathbf{M}(k,0)\mathbf{x}_0^b \tag{2.49}$$

Furthermore, it is useful to use the Sherman-Morrison-Woodbury formula found in Jazwinski [34] (see [39]) to obtain the solution in equation (2.46) in terms of the Kalman Gain matrix. This solution, in terms of the Kalman Gain, is the analytical solution found in Lorenc [40] and Thepaut and Courtier [52] and it presents a similar format to the Kalman filter solution.

For a model with model errors (weak 4D-Var), the minimization is done on the cost function

$$J = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T\mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{k=1}^{N}(\mathbf{H}_k\mathbf{x}_k - \mathbf{y}_k)^T\mathbf{R}_k^{-1}(\mathbf{H}_k\mathbf{x}_k - \mathbf{y}_k) + \frac{1}{2}\sum_{k=1}^{N}\mathbf{w}_k^T\mathbf{Q}_k^{-1}\mathbf{w}_k \tag{2.50}$$

There are three ways in which one could solve the minimization problem (2.50):

1. Carrying out the minimization with respect to the sequence $(\mathbf{x}_0, ..., \mathbf{x}_N, \mathbf{w}_1, ..., \mathbf{w}_N)$. This approach is usually done for theoretical purposes and it has been exclusively used to prove the equivalence between Kalman smoothers and 4D-Var (see Bryson and Ho [13]; Bennett and Budgell [8]; Menard and Daley [42]; Zhu et al. [56])

2. Carrying out an unconstrained optimizing by explicitly inputing the values of $\mathbf{w}_i$ in the cost function and minimizing with respect to the sequence $(\mathbf{x}_0, ..., \mathbf{x}_N)$. In which case, the new minimization is done on the functional described in the weak 4D-Var and the solution can be interpreted as the optimal trajectory solution, rather than the optimal initial condition found in the 4D-Var problem.

3. Carrying out the minimization with respect to $(\mathbf{x}_0, \mathbf{w}_1, ..., \mathbf{w}_N)$ on the functional (see Zupanski [58] for more details)

$$\begin{aligned} J &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) \\ &+ \frac{1}{2}\sum_{k=1}^{N}\left(\mathbf{H}_k\left(\mathbf{M}(k,0)\mathbf{x}_0 + \sum_{j=1}^{k}\mathbf{M}(k,j)\Gamma(j-1)\mathbf{w}_j\right) - \mathbf{y}_k\right)^T \\ &\times \mathbf{R}_k^{-1}\left(\mathbf{H}_k\left(\mathbf{M}(k,0)\mathbf{x}_0 + \sum_{j=1}^{k}\mathbf{M}(k,j)\Gamma(j-1)\mathbf{w}_j\right) - \mathbf{y}_k\right) \\ &+ \frac{1}{2}\sum_{k=1}^{N}\mathbf{w}_k\mathbf{Q}_k^{-1}\mathbf{w}_k \end{aligned} \tag{2.51}$$

Note the solution to the difference equation (2.41) is

$$\mathbf{x}_k = \mathbf{M}(k,0)\mathbf{x}_0 + \sum_{j=1}^{k}\mathbf{M}(k,j)\Gamma(j-1)\mathbf{w}_j \tag{2.52}$$

In order to provide an optimal solution of $\{\mathbf{x}_0, \mathbf{w}_1, ..., \mathbf{w}_N\}$ to equation (2.51), an augmented vector $\mathbf{Z} = (\mathbf{x}_0^T, \mathbf{w}_1^T, ..., \mathbf{w}_N^T)$ is created. Then, equation (2.51) can be transformed to

$$J = \frac{1}{2}(\mathbf{Z}_0 - \mathbf{Z}_0^b)^T \mathbf{B}_{\mathbf{Z}_0}^{-1}(\mathbf{Z} - \mathbf{Z}_0^b) + \frac{1}{2}\sum_{k=1}^{N}(\mathbf{H}_k\mathbf{C}_k\mathbf{Z}_0 - \mathbf{y}_k)^T\mathbf{R}_k^{-1}(\mathbf{H}_k\mathbf{C}_k\mathbf{Z}_0 - \mathbf{y}_k) \tag{2.53}$$

where $\mathbf{Z}_0 = ((\mathbf{x}_0^b)^T, \mathbf{0}, ..., \mathbf{0})$, $\mathbf{B}_{\mathbf{Z}_0}$ is a $(1+N)\times(1+N)$ block diagonal matrix whose

block diagonals are $\mathbf{B}_0, \mathbf{Q}_1, ..., \mathbf{Q}_N$ and

$$\mathbf{C}_k = \{\mathbf{M}(k,0), \mathbf{M}(k,1)\Gamma(0), ..., \mathbf{M}(k,k)\Gamma(k-1), 0, ..., 0\} \tag{2.54}$$

are $n \times (n + rN)$ matrices.

It is noted that for the case of a perfect model, equation (2.53) reduces to equation (2.44). The optimal solution is therefore

$$\hat{\mathbf{Z}}_0 = \mathbf{Z}_0^b + \mathbf{H}_{\mathbf{Z}_0}^{-1} \nabla_{\mathbf{Z}_0} J(\mathbf{Z}_0^b) \tag{2.55}$$

where the Hessian matrix is given by

$$\mathbf{H}_{\mathbf{Z}_0} = \mathbf{B}_{\mathbf{Z}_0}^{-1} + \sum_{k=1}^{N} \mathbf{C}_k^T \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \mathbf{C}_k \tag{2.56}$$

and the gradient of J evaluated at $\mathbf{Z}_0^b$ is given by

$$\begin{aligned} \nabla_{\mathbf{Z}_0} J(\mathbf{Z}_0^b) &= \sum_{k=1}^{N} \mathbf{C}_k^T \mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{C}_k \mathbf{Z}_0^b - \mathbf{y}_k) \\ &= \sum_{k=1}^{N} \mathbf{C}_k^T \mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{M}(k,0) \mathbf{x}_0^b - \mathbf{y}_k) \end{aligned} \tag{2.57}$$

Finally, it is noted that the calculation of the error covariance matrix of $\hat{\mathbf{Z}}_0$ is given as the inverse of the Hessian matrix

$$\mathbf{P}_{\mathbf{Z}_0}^a = \mathbf{H}_{\mathbf{Z}_0}^{-1} \tag{2.58}$$

It was shown that the 4D-Var and the weak 4D-Var methods can have solutions in terms of an analytic expression. The reader is referred to Courtier et al. [15] for an incremental algorithm that is useful for obtaining numerical solutions of the 4D-Var

problem when the model and observation equations are nonlinear. It is noted that following the steps in the section, one could also implement the incremental algorithm of Courtier et al. to obtain a numerical (approximate) solution of the nonlinear weak 4D-Var.

2.8.1 Optimality of Variational Data Assimilation Methods

For a perfect model, the joint posterior conditional probability density is (see Lorenc [40])

$$\begin{aligned} p(\mathbf{x}_0, ..., \mathbf{x}_N | \mathbf{y}_1, ..., \mathbf{y}_N) &= c \exp \Big\{ -\frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}_0^{-1} (\mathbf{x}_0 - \mathbf{x}_0^b) \\ &- \frac{1}{2} \sum_{k=1}^{N} (\mathbf{H}_k \mathbf{x}_k - \mathbf{y}_k)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{x}_k - \mathbf{y}_k) \Big\} \end{aligned} \tag{2.59}$$

Thus, the maximization of the posterior distribution $p(\mathbf{x}_0, ..., \mathbf{x}_l | \mathbf{y}_1, ..., \mathbf{y}_l)$ with respect to $(\mathbf{x}_0, ..., \mathbf{x}_l)$ is equivalent to the minimization of the cost functional (2.43). Therefore, the solution to the 4D-Var can be interpreted as the joint maximum likelihood (Bayesian) estimate.

If instead, it is only of concern to find the optimal estimate $\mathbf{x}_k (0 \leq k \leq N)$ for a fixed k only, yet still consider the full information set $Y_N = \{\mathbf{y}_1, ..., \mathbf{y}_N\}$, the marginal posterior density becomes

$$p(\mathbf{x}_k | \mathbf{y}_1, ..., \mathbf{y}_l) \tag{2.60}$$

Since the model is perfect, $\mathbf{x}_k = \mathbf{M}(k, 0)\mathbf{x}_0$ and by Theorem 2.7 found in Jazwinski (see [34])

$$\|\mathbf{M}(k, 0)\| p(\mathbf{x}_0 | Y_l) = p(\mathbf{x}_k | Y_l)$$

where $\|\mathbf{M}(k, 0)\|$ is the absolute value of the determinant of the matrix $\mathbf{M}(k, 0)$ and

where

$$p(\mathbf{x}_0|Y_N) = c\exp\Big\{-\frac{1}{2}(\mathbf{x}_0-\mathbf{x}_0^b)^T\mathbf{B}_0^{-1}(\mathbf{x}_0-\mathbf{x}_0^b) - \frac{1}{2}\sum_{k=1}^{N}(\mathbf{H}_k\mathbf{M}(k,0)\mathbf{x}_0-\mathbf{y}_k)^T\mathbf{R}_k^{-1}(\mathbf{H}_k\mathbf{M}(k,0)\mathbf{x}_0-\mathbf{y}_k)\Big\} \quad (2.61)$$

Therefore, maximization of $p(\mathbf{x}_k|Y_N)$ with respect to $\mathbf{x}_k$ is equivalent to maximization of $p(\mathbf{x}_0|Y_N)$ with respect to $\mathbf{x}_0$. On the other hand, maximization of $p(\mathbf{x}_0|Y_N)$ with respect to $\mathbf{x}_0$ is equivalent to minimization of (2.44) with respect to $\mathbf{x}_0$. Due to the consistency property of the 4D-Var, it can be further seen that maximizing $p(\mathbf{x}_0|Y_N)$ with respect to $\mathbf{x}_0$ is equivalent to maximizing $p(\mathbf{x}_0,...,\mathbf{x}_N|Y_N)$ with respect to the state sequence $\{\mathbf{x}_0,...,\mathbf{x}_N\}$.

For a model with errors, the posterior probability density of $p(\mathbf{x}_0,\mathbf{w}_1,...,\mathbf{w}_N|Y_N)$ is given by (see Li [39])

$$\begin{aligned} p(\mathbf{x}_0,\mathbf{w}_1,...,\mathbf{w}_N|Y_N) &= c\exp\Big\{-\frac{1}{2}(\mathbf{x}_0-\mathbf{x}_0^b)^T\mathbf{B}_0^{-1}(\mathbf{x}_0-\mathbf{x}_0^b) \\ &- \frac{1}{2}\sum_{k=1}^{N}\left(\mathbf{H}_k\left(\mathbf{M}(k,0)\mathbf{x}_0+\sum_{j=1}^{k}\mathbf{M}(k,j)\Gamma(j-1)\mathbf{w}_j\right)-\mathbf{y}_k\right)^T \\ &\times \mathbf{R}_k^{-1}\left(\mathbf{H}_k\left(\mathbf{M}(k,0)\mathbf{x}_0+\sum_{j=1}^{k}\mathbf{M}(k,j)\Gamma(j-1)\mathbf{w}_j\right)-\mathbf{y}_k\right) \\ &- \frac{1}{2}\sum_{k=1}^{N}\mathbf{w}_k\mathbf{Q}_k^{-1}\mathbf{w}_k\Big\} \end{aligned} \quad (2.62)$$

Therefore, maximizing $p(\mathbf{x}_0,\mathbf{w}_1,...,\mathbf{w}_N|Y_N)$ with respect to $\{\mathbf{x}_0,\mathbf{w}_1,...,\mathbf{w}_N\}$ is equivalent to minimizing the functional (2.51) found in "step 3". One could also show that maximizing $p(\mathbf{x}_k,\mathbf{w}_1,...,\mathbf{w}_N|Y_N)$ with respect to $\{\mathbf{x}_k,\mathbf{w}_1,...,\mathbf{w}_N\}$ for any $0 \leq k \leq N$ is equivalent to maximizing $p(\mathbf{x}_0,\mathbf{w}_1,...,\mathbf{w}_N|Y_N)$ with respect to the sequence $\{\mathbf{x}_0,\mathbf{w}_1,...,\mathbf{w}_N\}$. Therefore, the 4D-Var solution is optimal with respect to the model

trajectory $\{\mathbf{x}_0, ..., \mathbf{x}_N\}$ and with respect to a single state $\mathbf{x}_k$, even though the model contains errors.

2.8.2 Relationship Between the 4D-Var and the Kalman Filter

It has been long known that for a perfect model with linear observation and model equations, the 4D-Var solution and the Kalman filter yield the same values at the end of the assimilation window (see Kalman [36]). It was shown in the previous sections that the solution of the 4D-Var is the maximum (Bayesian) likelihood solution and that this solution is identical to that of the Kalman filter at the end of the time window for both, perfect and imperfect models.

This equivalence can also be verified by directly comparing the 4D-Var and Kalman filter solutions. For one iteration of the Kalman filter, we have at the $(k-1)$ step that

$$\mathbf{x}_{k-1} \sim N(\hat{\mathbf{x}}_{k-1}, \mathbf{P}^a_{k-1}) \tag{2.63}$$

and the Kalman filter can be used to obtain the estimate $\hat{\mathbf{x}}_k$ given new data. If the set up of the cost function is given as

$$J_k = \frac{1}{2}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})^T \mathbf{B}^{-1}_{k-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}) + \frac{1}{2}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k)^T \mathbf{R}^{-1}_k(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k) + \frac{1}{2}\mathbf{w}^T_k \mathbf{Q}_k \mathbf{w}_k \tag{2.64}$$

with the optimal solution denoted as $\hat{\mathbf{x}}_k$, it will be identical to that of the Kalman filter (see Bryson and Ho [13]).

2.9 Particle Filters

Particle filters are Sequential Monte Carlo algorithms that attempt to construct the posterior distribution $p(\mathbf{x}_0, ..., \mathbf{x}_N | \mathbf{y}_1, ..., \mathbf{y}_N)$ by taking many samples (called parti-

cles) that approximate the true posterior. This approach allows for the relaxation of the linear and Gaussian assumption in the State-Space model. Arulampalam et al. [5] provide a tutorial on particle filters for online nonlinear and non-Gaussian Bayesian tracking. This section will provides an outline of the Particle filter algorithm and its application to nonlinear State-Space problems. Further details and recent advances in the theory and practical implementation may be found in references [4] and [22].

The problem at hand is extended to allow for nonlinearity and non-Gaussian noise. That is, it is assumed the State-Space model to be

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{v}_k) \tag{2.65}$$

$$\mathbf{x}_{k+1} = \mathcal{M}_k(\mathbf{x}_k, \mathbf{w}_k) \tag{2.66}$$

where the model and observation operators are allowed to be nonlinear functions of the states and the corresponding noises are allowed to have non-Gaussian distributions. To demonstrate the general algorithm, let $\{\mathbf{x}_0^i, ..., \mathbf{x}_k^i, w_k^i\}_{i=1}^{N_s}$, denote a *random measure* that characterizes the posterior probability density function (pdf) $p(\mathbf{x}_0, ..., \mathbf{x}_N|\mathbf{y}_1, ..., \mathbf{y}_N)$where N_s is the number of samples (particles). Here, the sequence $\{\mathbf{x}_0^i, ..., \mathbf{x}_k^i\}_{i=1}^{N_s}$ is a set of support points and the sequence $\{w_k^i\}_{i=1}^{N_s}$ are the associated normalized weights with $\sum_i w_k^i = 1$. Then, the posterior probability density can be approximated discretely by

$$p(\mathbf{x}_0, ..., \mathbf{x}_N|\mathbf{y}_1, ..., \mathbf{y}_N) \approx \sum_{i=1}^{N_s} w_k^i \delta((\mathbf{x}_0, ..., \mathbf{x}_N) - (\mathbf{x}_0^i, ..., \mathbf{x}_N^i)) \tag{2.67}$$

Therefore, a discrete approximation to the true posterior distribution can be written as $p(\mathbf{x}_0, ..., \mathbf{x}_N|\mathbf{y}_1, ..., \mathbf{y}_N)$ in terms of weights and support points. Thus, Sequential Monte Carlo algorithms that are based on particle filters vary in two ways. First way

is to choose appropriate weights and the second way, is to choose appropriate support points that have meaningful weights such that the approximation is as accurate as possible. The weights can be chosen using the principle of *importance sampling*, see for example [9] and [22]. This principle assumes that $p(\mathbf{x}) \propto \pi(\mathbf{x})$ is a probability density that is difficult to draw samples from but for which $\pi(\mathbf{x})$ can be evaluated, as well as $p(\mathbf{x})$ up to a proportionality constant. Furthermore, if $\mathbf{x}_i \sim q(\mathbf{x})$ for $i = 1, ..., N_s$ are samples that are easily generated from a proposal density $q(\mathbf{x})$, called the *importance density*, then, the weights for the approximation

$$p(\mathbf{x}) \approx \sum_{i=1}^{N_s} w^i \delta(\mathbf{x} - \mathbf{x}^i) \tag{2.68}$$

can be computed by

$$w^i \propto \frac{\pi(\mathbf{x}^i)}{q(\mathbf{x}^i)} \tag{2.69}$$

This same principle can be applied to the filtering approach. To see this, suppose the samples $\{\mathbf{x}_0^i, ..., \mathbf{x}_N^i\}_{i=1}^{N_s}$ are drawn from an importance density $q(\mathbf{x}_0, ..., \mathbf{x}_N|\mathbf{y}_1, ..., \mathbf{y}_N)$, then the weights in (2.67) can be computed by

$$w_k^i \propto \frac{p(\mathbf{x}_0^i, ..., \mathbf{x}_N^i|\mathbf{y}_1, ..., \mathbf{y}_N)}{q(\mathbf{x}_0^i, ..., \mathbf{x}_N^i|\mathbf{y}_1, ..., \mathbf{y}_N)} \tag{2.70}$$

Suppose now that the data is coming in sequentially (online) and that an approximation of $p(\mathbf{x}_0, ..., \mathbf{x}_{N-1}|\mathbf{y}_1, ..., \mathbf{y}_{N-1})$ has been calculated. The need is now to approximate $p(\mathbf{x}_0, ..., \mathbf{x}_N|\mathbf{y}_1, ..., \mathbf{y}_N)$ with a new set of samples. If the importance density is chosen such that it can be factorized as

$$q(\mathbf{x}_0, ..., \mathbf{x}_N|\mathbf{y}_1, ..., \mathbf{y}_N) = q(\mathbf{x}_N|\mathbf{x}_0, ..., \mathbf{x}_{N-1}, \mathbf{y}_1, ...\mathbf{y}_N)q(\mathbf{x}_0, ..., \mathbf{x}_{N-1}|\mathbf{y}_1, ..., \mathbf{y}_{N-1}) \tag{2.71}$$

then, a set of samples $\mathbf{x}_N^i \sim q(\mathbf{x}_0, ..., \mathbf{x}_N | \mathbf{y}_1, ..., \mathbf{y}_N)$ can be obtained by augmenting the existing samples $\{\mathbf{x}_0^i, ..., \mathbf{x}_{N-1}^i\}_{i=1}^{N_s} \sim q(\mathbf{x}_0, ..., \mathbf{x}_{N-1} | \mathbf{y}_1, ..., \mathbf{y}_{N-1})$ with the new state $\mathbf{x}_N^i \sim q(\mathbf{x}_N | \mathbf{x}_0, ..., \mathbf{x}_{N-1}, \mathbf{y}_1, ...\mathbf{y}_N)$. The weight update equation can be derived to obtain an approximation of $p(\mathbf{x}_0, ..., \mathbf{x}_N | \mathbf{y}_1, ..., \mathbf{y}_N)$. Furthermore, if the importance density is such that

$$q(\mathbf{x}_N | \mathbf{x}_0, ..., \mathbf{x}_{N-1}, \mathbf{y}_1, ..., \mathbf{y}_N) = q(\mathbf{x}_N | \mathbf{x}_{N-1}, \mathbf{y}_1, ...\mathbf{y}_N) \tag{2.72}$$

then the importance density is dependent only on $\mathbf{x}_{N-1}, \mathbf{y}_1, ..., \mathbf{y}_N$. This is of importance in the case where only a filtered estimate is required at each time step from $p(\mathbf{x}_N | \mathbf{y}_1, ..., \mathbf{y}_N)$. This assumption is a standard assumption in the sampling approach. Therefore, the approximation is given by

$$p(\mathbf{x}_N | \mathbf{y}_1, ..., \mathbf{y}_N) \approx \sum_{i=1}^{N_s} w_k^i \delta(\mathbf{x}_N - \mathbf{x}_N^i) \tag{2.73}$$

where the weights are defined as

$$w_N^i \propto w_{N-1}^i \frac{p(\mathbf{y}_N | \mathbf{x}_N^i) p(\mathbf{x}_N^i | \mathbf{x}_{N-1}^i)}{q(\mathbf{x}_k^i | \mathbf{x}_{N-1}^i, \mathbf{y}_N)} \tag{2.74}$$

Thus, the *Sequential Importance Sampling* algorithm consists of recursive propagation of weights and support points as each measurement is given online. The corresponding procedure for the SIS Particle filter is presented in Algorithm 2.1 and can be found as Algorithm 1 in [5].

Algorithm 2.1: SIS Particle Filter

Input: State sample $\{\mathbf{x}_{k-1}^i\}_{i=1}^{N_s}$, weights $\{w_{k-1}^i\}_{i=1}^{N_s}$ and observation $\mathbf{y}_k$

Output: Updated states sample and weights $\{\mathbf{x}_k^i, w_k^i\}_{i=1}^{N_s}$

1: **procedure** SIS-Particle Filters($\{\mathbf{x}_{k-1}^i, w_{k-1}^i\}_{i=1}^{N_s}, \mathbf{y}_k$)

2: **for** i = 1,2,...,N_s **do**

3: Draw $\mathbf{x}_k^i \sim q(\mathbf{x}_k|\mathbf{x}_{k-1}^i, \mathbf{y}_k)$ ▷ $q(*)$ is the importance density

4: Assign the particle weight w_k^i according to equation (2.74)

5: **end for**

6: **end procedure**

This algorithm represents the essence that is involved in any Sequential Monte Carlo algorithm. There are many ways to avoid *degeneracy* in the algorithm by implementing a resampling method and a better choice of the importance function. Degeneracy is phenomenon in the SIS algorithm where, after many iterations, some particles will have very negligible weights. This implies that there will be some particles that will not have a meaningful contribution to the construction of the posterior density. For a complete tutorial on Particle filters, Arulampalam et al. [5] provides a consistent article. See Andrieu et al. [4] for a comprehensive discussion.

Chapter 3

Volatility Models in Financial Mathematics

This chapter provides an overview of the Multivariate Stochastic Volatility model, including fundamental properties. Estimation of unknown matrices and the optimal Bayesian solution of conditional volatility is presented based on an extended Kalman filter. The pricing of European options under this stochastic volatility model for the univariate case is also provided. The literature surrounding stochastic volatility is vast and quickly emerging as the need to estimate time-varying volatility of financial variables presents an alternative formulae to understanding their randomness.

The earliest comments on time-varying volatility dates back to the works of Mandelbrot [41] and Fama [26], while the break-through works in continuous-time finance of Black and Scholes [10] made it clear that there is evidence of non-stationarity in the variance of these financial variables. Taylor [51] provided the first publication that explicitly deals with the Stochastic Volatility model under the univariate case, while Johnson and Shanno [35] provided the earliest applications of the Stochastic Volatility model to study option pricing using time-varying volatility. A well known publication of continuous-time stochastic volatility modeling is that of Hull and White [32], where they allowed the volatility of spot prices of assets to follow a general diffusion process. The univariate Stochastic Volatility model presented in Ghysels et al. [27], Broto and Ruiz [12] and Shephard [1] form the general basis for building Multivariate Stochastic Volatility models.

Harvey et al. [29], Danielsson [20], Smith and Pitts [49] and Chan et al. [14] provided an analysis of the Multivariate Stochastic Volatility model with no-leverage effects (no correlation between prices and their volatility), while Yu [55] and Omori et al. [45] provided considerable evidence that measurements and volatility innovations are correlated (have leverage effects) for returns on stocks. The work of Chan et al. [14] provided a reasonable framework for modeling the leverage effects of financial time series. For a comprehensive treatment on Multivariate Stochastic Volatility modeling the reader is referred to Andersen et al. [2].

3.1 Introduction

The Black-Scholes model for pricing assumes that the volatility of the underlying asset is constant. That is, if S_t is the price of the underlying asset, then the Black-Scholes model for option pricing assumes the lognormal process

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t \tag{3.1}$$

where dW_t denotes a standard Wiener process and σ is the constant volatility of S_t. This assumption of constant volatility indeed is not the case, as it can be easily seen from any financial time series plot that the variations in the data points are sporadic around the mean. In this section, the Black-Scholes model for pricing European options is extended under stochastic volatility models. For more details on the Black-Scholes model for option pricing, see Appendix B and for extending the assumptions of the Black-Scholes model, see Wilmott [54].

Suppose that the underlying asset price S_t follows the same geometric distribution

of equation (3.1) as originally assumed by Black and Scholes [10]

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t^{(1)} \tag{3.2}$$

where $dW_t^{(1)}$ denotes a standard Wiener process. However, it is further assumed that the volatility σ of the underlying asset S_t is not constant, but that it follows its own general stochastic process

$$d\sigma_t = p(S_t, \sigma_t, t)dt + q(S_t, \sigma_t, t)dW_t^{(2)} \tag{3.3}$$

where $dW_t^{(2)}$ denotes another standard Wiener process. One can further assume that the two Wiener processes in (3.2) and (3.3) are correlated with correlation parameter ρ. The choice of the drift p and diffusion q are what define the different types of stochastic volatility models.

Consider now the value of an option under the stochastic volatility model (3.3). This means that the value of the option $V = V(S_t, \sigma_t, t)$ and is now also dependent on the realized values of the stochastic volatility and of the price of the underlying asset. As with the original derivation of the Black-Scholes PDE, we can create a portfolio containing one option with value $V(S_t, \sigma_t, t)$, a quantity $-\Delta$ of the underlying asset and quantity $-\Delta_1$ of another option with value $V_1(S_t, \sigma_t, t)$. Under this scenario, the portfolio has value

$$\Pi = V - \Delta S - \Delta_1 V_1 \tag{3.4}$$

Taking the stochastic derivative via Ito's Lemma, the change in the portfolio over a

dt period of time is given by

$$\begin{aligned} d\Pi &= \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \rho\sigma q S \frac{\partial^2 V}{\partial S \partial \sigma} + \frac{1}{2} q^2 \frac{\partial^2 V}{\partial \sigma^2}\right) dt \\ &- \Delta_1 \left(\frac{\partial V_1}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V_1}{\partial S^2} + \rho\sigma q S \frac{\partial^2 V_1}{\partial S \partial \sigma} + \frac{1}{2} q^2 \frac{\partial^2 V_1}{\partial \sigma^2}\right) dt \\ &+ \left(\frac{\partial V}{\partial S} - \Delta_1 \frac{\partial V_1}{\partial S} - \Delta\right) dS \\ &+ \left(\frac{\partial V}{\partial \sigma} - \Delta_1 \frac{\partial V_1}{\partial \sigma}\right) d\sigma \end{aligned} \tag{3.5}$$

As with the original derivation of the Black-Scholes PDE, the goal is to eliminate all randomness from the portfolio. This means that eliminating dS and $d\sigma$ from equation (3.5) is desired. This can be accomplished if

$$\begin{aligned} \frac{\partial V}{\partial S} - \Delta_1 \frac{\partial V_1}{\partial S} - \Delta &= 0 \\ \frac{\partial V}{\partial \sigma} - \Delta_1 \frac{\partial V_1}{\partial \sigma} &= 0 \end{aligned} \tag{3.6}$$

Furthermore, in order to get consistent pricing, it is assumed that there are no arbitrage opportunities within the portfolio. This means that the return on the portfolio value is equal to the risk-free rate of return of the portfolio over the dt period. That is,

$$d\Pi = r\Pi dt \tag{3.7}$$

After solving for the portfolio weights Δ and Δ_1 in equation (3.6) and using the

no-arbitrage condition (3.7)

$$\begin{aligned} d\Pi &= \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \rho\sigma q S \frac{\partial^2 V}{\partial S \partial \sigma} + \frac{1}{2} q^2 \frac{\partial^2 V}{\partial \sigma^2}\right) dt \\ &- \left(\frac{\partial V}{\partial \sigma} / \frac{\partial V_1}{\partial \sigma}\right)\left(\frac{\partial V_1}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V_1}{\partial S^2} + \rho\sigma q S \frac{\partial^2 V_1}{\partial S \partial \sigma} + \frac{1}{2} q^2 \frac{\partial^2 V_1}{\partial \sigma^2}\right) dt \\ &= r\left(V - \left(\frac{\partial V}{\partial S} - \frac{\partial V}{\partial \sigma} / \frac{\partial V_1}{\partial \sigma}\frac{\partial V_1}{\partial S}\right) S - \left(\frac{\partial V}{\partial \sigma} / \frac{\partial V_1}{\partial \sigma}\right) V_1\right) dt \end{aligned} \tag{3.8}$$

This equation involves the two unknowns V and V_1. So, collecting all V terms into the left side and all V_1 terms into the right side, we get

$$\frac{\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \rho\sigma S q \frac{\partial^2 V}{\partial S \partial \sigma} + \frac{1}{2} q^2 \frac{\partial^2 V}{\partial \sigma^2} + rS\frac{\partial V}{\partial S} - rV}{\frac{\partial V}{\partial \sigma}}$$

$$= \frac{\frac{\partial V_1}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V_1}{\partial S^2} + \rho\sigma S q \frac{\partial^2 V_1}{\partial S \partial \sigma} + \frac{1}{2} q^2 \frac{\partial^2 V_1}{\partial \sigma^2} + rS\frac{\partial V_1}{\partial S} - rV_1}{\frac{\partial V_1}{\partial \sigma}}$$

Notice that the left hand side equation is a function of V alone and that the right hand side equation is a function of V_1 alone. Since the two options V and V_1 will typically have different structures (payoffs, strikes, expiries, etc.) and depend only on the independent variables S, σ, and t, one can only obtain equality on both sides of the equation if each is independent of the contract structure. Thus, there exists a function $\lambda(S, \sigma, t)$ such that the left hand side and the right hand side are equal to λ. Therefore, the pricing equation for the option contract V under a stochastic volatility model and an extension of the Black-Scholes option pricing model is

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \rho\sigma S q \frac{\partial^2 V}{\partial S \partial \sigma} + \frac{1}{2} q^2 \frac{\partial^2 V}{\partial \sigma^2} + rS\frac{\partial V}{\partial S} + (p - \lambda q)\frac{\partial V}{\partial \sigma} - rV = 0 \tag{3.9}$$

where the function $\lambda(S, \sigma, t)$ is called the *market price of volatility risk*.

3.1.1 Univariate Stochastic Volatility Model

It was shown in the previous section that the model for option pricing can be extended to include a stochastic volatility model of the underlying asset. Inclusion of a stochastic volatility model in option pricing can help the problem become more realistic than the simplistic assumptions taken by the Black-Scholes model. The financial industry has taken advantage of this approach and has gravitated towards assuming a stochastic volatility model when pricing options, in particular, in the pricing of swaption contracts. A swaption contract is a financial derivative that allows the holder of the swaption to enter into a swap contract with a counterparty at a prescribed strike price, called, the swap interest rate. For more information on swaptions, we refer the reader to Hull [31].

One of the greatest successes of stochastic volatility modeling in swaption pricing is the Stochastic Alpha Beta Rho (SABR) model. This model assumes dynamics

$$\begin{aligned} dS_t &= (S_t)^{\beta}\sigma_t dW_t^{(1)} \\ d\sigma_t &= \nu dW_t^{(2)}, \qquad \sigma_0 = \alpha \\ \mathcal{E}\{dW_t^{(1)} dW_t^{(2)}\} &= \rho \end{aligned} \tag{3.10}$$

Where the unknown parameters of the model are $\{\alpha, \beta, \rho, \nu\}$ and are usually calibrated by minimizing the square errors between the implied volatility obtained by the SABR model and the implied volatility observed on the market. Note that the implied volatility is the volatility such that the price in the Black-Scholes model is recovered and matches market observed prices. For more information on the SABR model and calibration of the unknown parameters, the reader is referred to Hagan et al. [28].

Another financial industry standard on modeling volatility from market observed prices is the well-known Generalized Autoregressive Conditional Heteroscedastic model (GARCH) developed by Engle [24] and Bollerslev [11]. This model has been extensively studied and its applications have covered a wide footprint in the econometrics and financial econometrics literature; see for example Zivot [57]. Although the GARCH model has found many applications in the financial industry and has been proven to be useful, it has been shown through many empirical studies that the Stochastic Volatility model considered in this dissertation provides a basis for more accurate forecasts of volatilities than those provided by the GARCH model; see for example Koopman et al. [38]. Moreover, the Stochastic Volatility model is considered to be more consistent with financial theory as opposed to the GARCH model.

The Stochastic Volatility model that is considered in this dissertation is similar to the SABR model and the GARCH model in that the process attempts to model the lognormal distribution of the underlying assets and volatilities. Assume again that the price of the underlying asset is given by S_t and that it follows the stochastic differential equation

$$\begin{aligned} d\log(S_t) &= \mu_t dt + \sigma_t dW_t^{(1)} \\ d\log(\sigma_t^2) &= \{\gamma + (\phi - 1)\log(\sigma_t^2)\}dt + \sigma_\eta dW_t^{(2)} \end{aligned} \tag{3.11}$$

where μ_t is the drift term, σ_t is the stochastic volatility of the asset price S_t and dW^i for $i = 1, 2$ are two Brownian processes. The model has the unknown parameters $\{\gamma, \phi, \sigma_\eta\}$, where σ_η represents the volatility of volatility (volvol) parameter. In many financial time series, it is observed that μ_t is very small. Thus, it is common practice to set $\mu_t = 0$.

By applying a direct Euler discretization scheme with $\Delta = 1$, we can arrive at the

Stochastic Volatility (SV) model considered in this dissertation

$$\begin{aligned} y_t &= \mu_t + \exp\{\hat{\sigma}_t/2\}\epsilon_t, \qquad \epsilon_t \sim N(0,1) \\ \hat{\sigma}_t &= \gamma + \phi\hat{\sigma}_t + \sigma_\eta \eta_t, \qquad \eta_t \sim N(0,1) \end{aligned} \tag{3.12}$$

where y_t are the log-returns of the asset price; that is, $y_t = \log S_{t+1} - \log S_t$, which means that the model assumes a constant mean μ. To derive equation (3.12), the transformation $\hat{\sigma}_t = \log \sigma_t^2$ is made. It is also assumed that the log-volatility process $\hat{\sigma}_t$ is a stationary process but persistent, meaning that $0 < \phi < 1$ is typically greater than 0.8 when estimated. The unconditional mean of the log-volatility process $\hat{\sigma}_t$ is given by $(1-\phi)^{-1}\gamma$ and is interpreted as the long-term log-variance of the asset return series y_t. The unconditional variance of the log-volatility process $\hat{\sigma}_t$ is given by $(1-\phi^2)^{-1}\sigma_\eta^2$ and it is sometimes referred to as the volatility of volatility of the log-variance. Furthermore, the stochastic time-varying variance of the log returns y_t conditional on $\hat{\sigma}_t$ is given by

$$\sigma_t^2 = \mathcal{E}\{(y_t - \mu)^2 | \hat{\sigma}_t\} = \exp \hat{\sigma}_t \tag{3.13}$$

In the next section, a multivariate Stochastic Volatility model (MSV) is derived as a natural extension to the Stochastic Volatility Model in equation (3.12).

3.2 Multivariate Stochastic Volatility Model

There has been a considerable amount of attention focused on the empirical modeling of conditional volatility, in particular, in the multivariate extensions of the univariate GARCH model; for example, see Bauwens et al. [7] for a recent survey. On the other hand, the financial econometrics literature has also focused on the development

of multivariate stochastic volatility models based on the multivariate extension of equation (3.12). For a comprehensive treatment of the various extensions of the MSV model, the reader is referred to Andersen et al. [2]. This dissertation focuses on two types of multivariate Stochastic Volatility models, one with leverage effects and one with non-leverage effects.

3.2.1 No-Leverage Effects Model

Consider a time series of p-financial variables $\mathbf{y}_t = (y_t^{(1)}, ..., y_t^{(p)})^T$ observed at time t and let $\mathbf{x} = (x_t^{(1)}, ..., x_t^{(p)})^T$ be the corresponding vector of log-volatilities at time t. A direct extension of the univariate Stochastic Volatility model (3.12) is given by

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\gamma} + \mathbf{V}_t^{1/2}\boldsymbol{\epsilon}_t \\ \mathbf{x}_{t+1} &= \boldsymbol{\mu} + \mathbf{M}(\mathbf{x}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \\ \mathbf{x}_0 &\sim N_p(\boldsymbol{\mu}, \mathbf{P}_0) \end{aligned} \tag{3.14}$$

where $\mathbf{x}_0 \in \mathbb{R}^p$ is the initial prior, $\mathbf{P}_0 \in \mathbb{R}^{p\times p}$ is a covariance matrix, $\mathbf{M} \in \mathbb{R}^{p\times p}$ and

$$\begin{aligned} \mathbf{V}_t^{1/2} &= \text{diag}(\exp(x_t^{(1)}/2), ..., \exp(x_t^{(p)}/2)) \\ \boldsymbol{\mu} &= (\mu^{(1)}, ..., \mu^{(p)}), \ \boldsymbol{\gamma} = (\gamma^{(1)}, ..., \gamma^{(p)}) \end{aligned} \tag{3.15}$$

Furthermore,

$$\boldsymbol{\epsilon}_t \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_{\epsilon\epsilon}), \qquad \boldsymbol{\eta}_t \sim N_p(\mathbf{0}, \mathbf{Q}) \tag{3.16}$$

where the problem of having no leverage effects amounts to having no correlation between the two white noise processes $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$. In order to model the conditional volatilities, the Multivariate Stochastic Volatility (MSV) model needs to be estimated based on historical observations to provide parameter estimates of the unknowns.

The unknown parameters of the MSV model are $\{\boldsymbol{\gamma}, \boldsymbol{\mu}, \mathbf{M}, \boldsymbol{\Sigma}_{\epsilon\epsilon}, \mathbf{Q}\}$ as well as the log-volatility series $\{\mathbf{x}_1, ..., \mathbf{x}_N\}$ for a data assimilation window $[t_0, t_N]$. As mentioned in Andersen et al. [2], for identification purposes, the diagonal elements of $\boldsymbol{\Sigma}_{\epsilon\epsilon}$ must be 1, which implies that the matrix $\boldsymbol{\Sigma}_{\epsilon\epsilon}$ is a correlation matrix. It is further noted that the MSV model is a State-Space model with a linear evolution of the state equation, as a Markov process, and a nonlinear evolution of the measurement equation (because $\mathbf{x}_t$ enters the model in a multiplicative way).

3.2.2 Leverage-Effects Model

The Leverage-Effects MSV model is another extension of the MSV model that allows for correlation between $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$. In this manner, the Leverage-Effects MSV models the correlations between the log-returns time series and the conditional volatility time series. This is of particular importance since the existence of correlations between stocks log-returns and their volatility innovations has been shown; see for example Yu [55] and Omori et al. [45]. The model considered by Chan et al. [14] is given as

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\gamma} + \mathbf{V}_t^{1/2}\boldsymbol{\epsilon}_t \\ \mathbf{x}_{t+1} &= \boldsymbol{\mu} + \underbrace{\mathrm{diag}(\phi_1, ..., \phi_p)}_{\mathbf{M}}(\mathbf{x}_t - \boldsymbol{\mu}) + \boldsymbol{\Psi}^{1/2}\boldsymbol{\eta}_t \\ \mathbf{x}_0 &\sim N_p(\boldsymbol{\mu}, \boldsymbol{\Psi}^{1/2}\mathbf{P}_0\boldsymbol{\Psi}^{1/2}) \end{aligned} \tag{3.17}$$

where

$$\begin{aligned} \mathbf{V}_t^{1/2} &= \mathrm{diag}(\exp(x_t^{(1)}/2), ..., \exp(x_t^{(p)}/2)) \\ \boldsymbol{\mu} &= (\mu^{(1)}, ..., \mu^{(p)}), \;\; \boldsymbol{\gamma} = (\gamma^{(1)}, ..., \gamma^{(p)}) \end{aligned}$$

The observation and state innovations continue to be normally distributed with covariance matrix $\mathbf{\Sigma}_{\epsilon\epsilon}$ and $\mathbf{\Sigma}_{\eta\eta}$, respectively. However, their process is modeled jointly with

$$\begin{bmatrix} \boldsymbol{\epsilon}_t \\ \boldsymbol{\eta}_t \end{bmatrix} \sim N_{2p}(\mathbf{0}, \mathbf{\Sigma}), \qquad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{\epsilon\epsilon} & \mathbf{\Sigma}_{\epsilon\eta} \\ \mathbf{\Sigma}_{\eta\epsilon} & \mathbf{\Sigma}_{\eta\eta} \end{bmatrix}$$

where $\mathbf{\Sigma}_{\epsilon\eta}$ denotes the correlation between $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$; that is, the returns of the asset and its volatility are now correlated. The state-disturbance loading matrix $\mathbf{\Psi}^{1/2}$ is defined as the diagonal matrix

$$\mathbf{\Psi}^{1/2} = \text{diag}(\sqrt{\psi_1^2}, ..., \sqrt{\psi_p^2})$$

whose diagonal entries ψ_i are unknown. Notice that $\mathbf{\Psi}$ can be interpreted as the volatility of the log-volatility process. Indeed, if $\mathbf{\Sigma}_{\eta\eta}$ is the identity matrix, then covariance of the innovation process $\hat{\boldsymbol{\eta}}_t = \mathbf{\Psi}^{1/2}\boldsymbol{\eta}_t$ is given by $\mathbf{\Psi}^{1/2}\mathbf{\Psi}^{1/2}$, implying that variance of each element in $\hat{\boldsymbol{\eta}}_t$ is given by the corresponding diagonal entry in $\mathbf{\Psi}^{1/2}\mathbf{\Psi}^{1/2}$.

Furthermore, the problem assumes that the state transition matrix $\mathbf{M}$ is defined as a diagonal matrix whose entries are the unknown parameters ϕ_i and the (i, j) element of $\mathbf{P}_0$ is the (i, j) element of $\mathbf{\Sigma}_{\eta\eta}$ divided by $1 - \phi_i\phi_j$ satisfying the stationary condition such that

$$\mathbf{P}_0 = \mathbf{M}\mathbf{P}_0\mathbf{M} + \mathbf{\Sigma}_{\eta\eta}$$

As with the no-leverage effects MSV model, this MSV model needs to be estimated from historical data. The unknown parameters in the model to estimate from data are $\{\boldsymbol{\gamma}, \boldsymbol{\mu}, \mathbf{M}, \mathbf{\Psi}, \mathbf{\Sigma}_{\epsilon\epsilon}, \mathbf{\Sigma}_{\eta\eta}, \mathbf{\Sigma}_{\epsilon\eta}\}$ as well as the log-volatility path $\{\mathbf{x}_0, ..., \mathbf{x}_N\}$ for a data assimilation window $[t_0, t_N]$.

3.3 Estimation of Model Parameters for Non-Leverage Effects

The model considered in this dissertation for estimation and numerical experiments is the MSV model with no leverage effects model of equation (3.14). Before estimates and forecasts of the log-volatility times series $\mathbf{x}_t$ can be obtained during a data assimilation window, note that all unknown matrices and vectors need to be consistently estimated. To this end, consider again the non-leverage effects MSV model

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\gamma} + \mathbf{V}_t^{1/2}\boldsymbol{\epsilon}_t \\ \mathbf{x}_{t+1} &= \boldsymbol{\mu} + \mathbf{M}(\mathbf{x}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \\ \mathbf{x}_0 &\sim N_p(\boldsymbol{\mu}, \mathbf{P}_0) \end{aligned} \tag{3.18}$$

To fit the model to historical time series data $\mathbf{y}_t$, linearization of the observation equation is performed as follows. Define the vector $\tilde{\mathbf{y}}_t = [\tilde{y}_t^{(1)}, ..., \tilde{y}_t^{(p)}]'$ where $\tilde{y}_t^{(i)} = \log\left(y_t^{(i)} - \gamma^{(i)}\right)^2$ for $i = 1, ..., p$. The nonlinear State-Space model (3.18) then becomes

$$\begin{aligned} \tilde{\mathbf{y}}_t &= (-1.27)\mathbf{1}_p + \mathbf{x}_t + \tilde{\boldsymbol{\epsilon}}_t \\ \mathbf{x}_{t+1} &= \boldsymbol{\mu} + \mathbf{M}(\mathbf{x}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \\ \mathbf{x}_0 &\sim N_p(\boldsymbol{\mu}, \mathbf{P}_0) \end{aligned} \tag{3.19}$$

where $\mathbf{1}_p = [1, ..., 1]' \in \mathbb{R}^p$ and $\tilde{\boldsymbol{\epsilon}}_t \in \mathbb{R}^p$ with $\tilde{\epsilon}_t^{(i)} = \log\left(\epsilon_t^{(i)}\right)^2 + 1.27$. Notice that observation equation is now linear and

$$\mathcal{E}\left\{\log\left(\epsilon_t^{(i)}\right)^2\right\} = -1.27, \qquad \text{Var}\left\{\log\left(\epsilon_t^{(i)}\right)^2\right\} = \frac{\pi^2}{2} \tag{3.20}$$

where the constant vector $(-1.27)\mathbf{1}_p$ is included in the observation equation to off-set the mean. However, the State-Space model (3.19) is now a Non-Gaussian State-Space

model since the new state-error stochastic process $\tilde{\boldsymbol{\epsilon}}_t$ no longer follows a multivariate Gaussian distribution. It was shown in Harvey et al. [29] that the (i,j) element of the covariance matrix $\mathbf{R}$ of $\tilde{\boldsymbol{\epsilon}}_t$ is given by $\left(\frac{\pi^2}{2}\right) r_{ij}$ where $r_{ii} = 1$ and

$$r_{ij} = \frac{\pi^2}{2} \sum_{n=1}^{\infty} \frac{(n-1)!}{\{\prod_{k=1}^{n}(1/2+k-1)\}\, n} \rho_{ij}^{2n} \tag{3.21}$$

where ρ_{ij} is the (i,j) element of the covariance matrix $\boldsymbol{\Sigma}_{\epsilon\epsilon}$ of $\boldsymbol{\epsilon}_t$. Although the linearized MSV model is no longer Gaussian, estimating the unknown parameters using a quasi-Maximum Likelihood Estimation (qMLE) methodology may be implemented by assuming that the new state-error stochastic process $\tilde{\boldsymbol{\epsilon}}_t$ is Gaussian.

Suppose that it is observed from the market a time series sequence $\{\mathbf{y}_1, ..., \mathbf{y}_N\}$ of the $p-$financial variables. Under the Gaussian assumption, the log-likelihood of the data given the model is given as

$$\ln\left(p(\tilde{\mathbf{y}}_1, ..., \tilde{\mathbf{y}}_N | \{\hat{\mathbf{x}}_i\}_{i=1}^N)\right) = \sum_{t=1}^{N} \ln\left(\phi(\tilde{\mathbf{y}}_t; \bar{\mathbf{y}}_{t|t-1}, \mathcal{V}_{t|t-1} | \{\hat{\mathbf{x}}_i\}_{i=1}^N)\right) \tag{3.22}$$

Where $\phi(\tilde{\mathbf{y}}_t; \bar{\mathbf{y}}_{t|t-1}, \mathcal{V}_{t|t-1} | \{\hat{\mathbf{x}}_i\}_{i=1}^N)$ is the multivariate normal distribution evaluated at $\tilde{\mathbf{y}}_t$ with mean $\bar{\mathbf{y}}_{t|t-1}$ and covariance matrix $\mathcal{V}_{t|t-1}$ given the state sequence $\{\hat{\mathbf{x}}_{i=1}^N\}$.

In order to describe the q-MLE procedure, suppose that the following State-Space model with unknown parameters $\{\boldsymbol{\gamma}, \boldsymbol{\mu}, \mathbf{M}, \mathbf{R}, \mathbf{Q}\}$ is given

$$\begin{aligned} \tilde{\mathbf{y}}_t &= \boldsymbol{\gamma} + \mathbf{H}\mathbf{x}_t + \mathbf{R}^{1/2}\tilde{\boldsymbol{\epsilon}}_t \\ \mathbf{x}_t &= \boldsymbol{\mu} + \mathbf{M}\mathbf{x}_{t-1} + \mathbf{Q}^{1/2}\boldsymbol{\eta}_t \end{aligned} \tag{3.23}$$

where $\boldsymbol{\epsilon}_t, \boldsymbol{\eta}_t \sim N_p(\mathbf{0}, \boldsymbol{I}_{p\times p})$. It is noted that in equation (3.23), the model incorporates the square root decompositions $\mathbf{R}^{1/2}$ and $\mathbf{Q}^{1/2}$ such that the covariance of the

stochastic processes $\mathbf{R}^{1/2}\tilde{\boldsymbol{\epsilon}}_t$ and $\mathbf{Q}^{1/2}\boldsymbol{\eta}_t$ are given by $\mathbf{R}^{1/2}\mathbf{R}^{T/2} = \mathbf{R}$ and $\mathbf{Q}^{1/2}\mathbf{Q}^{T/2} = \mathbf{Q}$, respectively. The state sequence $\{\hat{\mathbf{x}}_{i=1}^{N}(\boldsymbol{\gamma}, \boldsymbol{\mu}, \mathbf{M}, \mathbf{R}, \mathbf{Q})\}$ and its state covariance matrix sequence $\{\mathbf{P}_{i=1}^{N}(\boldsymbol{\gamma}, \boldsymbol{\mu}, \mathbf{M}, \mathbf{R}, \mathbf{Q})\}$ can be obtained via the Kalman filter (or any data assimilation method) as a function of the given parameters and, therefore, a value for the log-likelihood function via equation (3.22) can also be obtained. If the parameters $(\boldsymbol{\gamma}, \boldsymbol{\mu}, \mathbf{M}, \mathbf{R}, \mathbf{Q})\}$ are all stacked into a single vector $\boldsymbol{\theta}$, the value of the function $\phi(\tilde{\mathbf{y}}_t; \bar{\mathbf{y}}_{t|t-1}, \mathcal{V}_{t|t-1} | \{\hat{\mathbf{x}}(\boldsymbol{\theta})_i\}_{i=1}^{N})$ is given by

$$\tilde{\mathbf{y}}_t | \hat{\mathbf{x}}_{t-1}; \boldsymbol{\theta} \sim N_p(\bar{\mathbf{y}}_{t|t-1}(\boldsymbol{\theta}), \mathcal{V}_{t|t-1}(\boldsymbol{\theta})) \tag{3.24}$$

where

$$\begin{aligned} \bar{\mathbf{y}}_{t|t-1}(\boldsymbol{\theta}) &= \boldsymbol{\gamma}(\boldsymbol{\theta}) + \mathbf{H}(\boldsymbol{\theta})\hat{\mathbf{x}}_{t|t-1}(\boldsymbol{\theta}) \\ \mathcal{V}_{t|t-1} &= \mathbf{H}(\boldsymbol{\theta})\boldsymbol{P}_{t|t-1}\mathbf{H}^T(\boldsymbol{\theta}) + \mathbf{R}^{1/2}(\boldsymbol{\theta})\mathbf{R}^{T/2}(\boldsymbol{\theta}) \end{aligned} \tag{3.25}$$

Furthermore, the value of the log-likelihood function of the data given the parameter vector $\boldsymbol{\theta}$ is calculated as

$$\begin{aligned} ll(\boldsymbol{\theta}) &= \sum_{t=1}^{N} \log \phi(\tilde{\mathbf{y}}_t; \bar{\mathbf{y}}_{t|t-1}, \mathcal{V}_{t|t-1} | \{\hat{\mathbf{x}}_t\}_{i=1}^{N}) \\ &= \frac{Np}{2}\log(2\pi) + \frac{1}{2}\sum_{t=1}^{N} \log\left(\det\{\mathcal{V}_{t|t-1}\}\right) \\ &+ \frac{1}{2}\sum_{t=1}^{N} \left(\tilde{\mathbf{y}}_t - \bar{\mathbf{y}}_{t|t-1}(\boldsymbol{\theta})\right)^T \mathcal{V}_{t|t-1}^{-1} \left(\tilde{\mathbf{y}}_t - \bar{\mathbf{y}}_{t|t-1}(\boldsymbol{\theta})\right) \end{aligned} \tag{3.26}$$

The q-MLE methodology can be implemented for the MSV model with no leverage effects by letting $\mathbf{H} = \mathbf{I}_{p\times p}$ and $\boldsymbol{\gamma} = (-1.27)\mathbf{1}_p$ in equation (3.23). The q-MLE methodology can be implemented in an optimization algorithm to minimize the *negative*

log-likelihood function of equation (3.22) to obtain optimal parameters (estimates). The corresponding procedure to estimate the unknown model parameters is given in Algorithm 3.1 and details underlying this procedure can be found in Chapter 50 of the Handbook of Econometrics [25].

Algorithm 3.1: q-MLE State-Space Estimator

Input: Sequence of Data $\{\mathbf{y}_t\}_{t=1}^N$ and initial parameter estimate $\boldsymbol{\theta}_0$

Output: Optimized parameter $\boldsymbol{\theta}^*$ and Kalman filter sequence $\{\hat{\mathbf{x}}_{t|t}(\boldsymbol{\theta}^*)\}_{t=1}^N$

1: **procedure** Q-MLE($\mathbf{y}, \boldsymbol{\theta}_0$)
2: **for** k = 1,2,... **do**
3: Search for $\boldsymbol{\theta}_k$ such that $-ll(\boldsymbol{\theta}^k) < -ll(\boldsymbol{\theta}^{k-1})$ ▷ see equation (3.22)
4: Obtain Kalman filter sequence $\{\hat{\mathbf{x}}_{t|t}(\boldsymbol{\theta}_k)\}_{t=1}^N$
5: Evaluate convergence of $\boldsymbol{\theta}_k$ via a tolerance function
6: **if** Convergence criteria for $\boldsymbol{\theta}_k$ is satisfied **then**
7: $\boldsymbol{\theta}^* = \boldsymbol{\theta}_k$ and get $\{\hat{\mathbf{x}}_{t|t}(\boldsymbol{\theta}^*)\}_{t=1}^N$
8: **else**
9: Continue to next iteration
10: **end if**
11: **end for**
12: **end procedure**

Chapter 4

Sensitivity Analysis

Model-based approach for stochastic volatility requires accurate specification of the unknown parameters used to represent the model. The q-MLE methodology provides a consistent estimation, based on statistical analysis, of the unknown matrices in the model, in particular, the state and observation error covariance specifications. However, volatility forecast errors can be attributed to misspecification of the "true" covariance matrices that underly the "true" stochastic volatility model as these volatility forecasts are dependent on the specification of the input parameters (estimated parameters). Assessing the forecast errors in variational data assimilation systems due to variations and misspecification of input parameters such as observations, state and error covariance matrices, have been extensively studied over the years. Baker and Daley [6] have shown that forecast errors due to online observations can be evaluated based on an *all at once* forecast sensitivity analysis derived from the adjoint of the data assimilation system (adjoint-DAS). The adjoint-DAS applications have been extended by Daescu [16] to incorporate forecast sensitivity analysis due to state and observation error covariance model specifications in a nonlinear four-dimensional variational data assimilation system (4D-Var) DAS. The practical ability to estimate the forecast error sensitivities due to observation and state error covariance matrix specifications was shown by Daescu and Todling [19] for a three-dimensional variational data assimilation system (3D-Var DAS). Daescu and Langland [18] presented a

complete review of the 4D-Var system and provided additional forecast error sensitivity equations for the 4D-Var system along with their properties for and applications with parameter tuning and impact assessment. They also presented numerical results with the Naval Research Laboratory Atmospheric Data Assimilation System - Accelerated Representer and the Navy Operational Global Atmospheric Prediction System (NAVDAS-AR/NOGAPS) to emphasize the use of forecast error sensitivity information for analyzing and diagnosing the DAS performance.

The work in this chapter extends the adjoint-based approach for sensitivity analysis and impact estimation of the 4D-Var DAS to the linearized Multivariate Stochastic Volatility model (MSV) and provides equations to evaluate and assess the volatility forecast error sensitivities with respect to input parameters. These sensitivities are used to diagnose the current state and observation error covariance matrix specifications obtained from the q-MLE estimation approach. Guidance tools for error covariance parameter online tuning and assessment will also be presented for the MSV model as an adaptive approach.

4.1 The Analysis Equation

Consider again the linearized Multivariate Stochastic Volatility (MSV) model discussed in Chapter 3

$$\begin{aligned} \tilde{\mathbf{y}}_t &= (-1.27)\mathbf{1}_p + \mathbf{x}_t + \tilde{\boldsymbol{\epsilon}}_t \\ \mathbf{x}_{t+1} &= \boldsymbol{\mu} + \mathbf{M}(\mathbf{x}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \\ \mathbf{x}_0 &\sim N_p(\boldsymbol{\mu}, \mathbf{P}_0) \end{aligned} \tag{4.1}$$

where $\mathbf{x}_t$ is the vector of log-volatilities, $\mathbf{x}_0$ is the prior (background) state estimate with prior covariance matrix $\mathbf{P}_0$, and $\tilde{\mathbf{y}}_t \in \mathbb{R}^p$ is the vector of observational data (as

defined in Chapter 3). It is also assumed that the errors are white noise processes that have the following distributions

$$\tilde{\boldsymbol{\epsilon}}_t \sim N_p(\mathbf{0}, \mathbf{R}) \qquad \boldsymbol{\eta}_t \sim N_p(\mathbf{0}, \mathbf{Q}) \tag{4.2}$$

As mentioned before, this MSV model belongs to a general class of linear State-Space models that have the following form

$$\begin{aligned} \tilde{\mathbf{y}}_t &= \boldsymbol{\gamma} + \mathbf{H}\mathbf{x}_t + \tilde{\boldsymbol{\epsilon}}_t \\ \mathbf{x}_t &= \boldsymbol{\mu} + \mathbf{M}\mathbf{x}_{t-1} + \boldsymbol{\eta}_t \end{aligned} \tag{4.3}$$

where the errors are given as in equation (4.2).

If the parameters $\{\boldsymbol{\gamma}, \boldsymbol{\mu}, \mathbf{M}, \mathbf{H}, \mathbf{R}, \mathbf{Q}\}$ are all known, then we can use a standard Kalman filter to obtain an analysis $\mathbf{x}_{t|t}$, given all information up to t, for each time iteration. Furthermore, notice that at the next iteration $t+1$, we can make a prediction $\mathbf{x}_{t+1|t}$ from $\mathbf{x}_{t|t}$. Thus, at $t+1$, the prediction $\mathbf{x}_{t+1|t}$ becomes the new prior state estimate for the analysis $x_{t+1|t+1}$. Therefore, to ease the notation in this chapter, the time index t is dropped and refer to $\mathbf{x}^a$ as the analysis of $\mathbf{x}_t$ at the current iteration with covariance matrix $\mathbf{P}^a$ and $\mathbf{x}^b$ as the prior state estimate with covariance matrix $\mathbf{B}$.

In general, variational data assimilation provides an analysis $\mathbf{x}^a$ to the true state $\mathbf{x}$ by minimizing the cost functional

$$J(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x} - \mathbf{x}^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x} - \mathbf{x}^b\right) + \frac{1}{2}\left(\mathbf{h}(\mathbf{x}) - \mathbf{y}\right)^T \mathbf{R}^{-1}\left(\mathbf{h}(\mathbf{x}) - \mathbf{y}\right) \tag{4.4}$$

where $\mathbf{h}(*)$ is the linear operator that maps the states into observations and it is

defined by

$$\mathbf{h}(\mathbf{x}) = \boldsymbol{\gamma} + \mathbf{H}\mathbf{x} \tag{4.5}$$

Recall that if the Kalman filter is used to obtain a state estimate, this estimate is equivalent to minimizing equation (4.4). By direct calculation of the Jacobian of (4.4) yields

$$\nabla J(\mathbf{x}) = \mathbf{B}^{-1}\left(\mathbf{x} - \mathbf{x}^b\right) + \mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{h}(\mathbf{x}) - \mathbf{y}\right) = \mathbf{0} \tag{4.6}$$

The solution to (4.6) is denoted as $\mathbf{x}^a$, representing a single outer loop iteration of the analysis, and it is given by

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}\left[\mathbf{y} - \mathbf{h}(\mathbf{x}^b)\right] \tag{4.7}$$

where $\mathbf{K}$ is the gain matrix and it is given by

$$\mathbf{K} = \left[\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}^T\mathbf{R}^{-1} = \mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1} \tag{4.8}$$

It is often the case where the analysis equation is performed in a two-stage procedure by first solving the linear system

$$[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]\mathbf{z} = \mathbf{y} - \mathbf{h}(\mathbf{x}^b) \tag{4.9}$$

for the vector $\mathbf{z}$ and then followed by a post-multiplication operation

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{B}\mathbf{H}^T\mathbf{z} \tag{4.10}$$

From equation (4.7), it is observed that the analysis equation is dependent on the

input parameters; that is,

$$\mathbf{x}^a = \mathbf{x}^a(\mathbf{y}, \mathbf{x}^b, \mathbf{B}, \mathbf{R}) \tag{4.11}$$

Therefore, the focus of this chapter is to provide sensitivity analysis on $\mathbf{x}^a$ due to variations in each input in order to assess model performance based on state forecasts. Furthermore, sensitivity analysis is extended to the linearized Multivariate Stochastic Volatility model in order to provide diagnostics to model performance of the forecasted volatilities as well as show how one can implement an adaptive tuning procedure to obtain improved volatility estimates.

4.2 Forecast Sensitivity

Once a forecast of the states have been made, the evaluation of the forecast sensitivity with respect to each input is done via a forecast score function. This forecast score is defined as a short-range forecast-error measure and it is usually presented as

$$e(\mathbf{x}^a) = \left(\mathbf{x}_f - \mathbf{x}_f^v\right)^T \mathbf{E} \left(\mathbf{x}_f - \mathbf{x}_f^v\right) \tag{4.12}$$

where $\mathbf{x}_f = \mathcal{M}_{t_k,t_f}(\mathbf{x}^a)$ is the state forecast at verification time t_f initiated from $\mathbf{x}^a$ at time of analysis t_k with the state model represented as

$$\mathcal{M}(\mathbf{x}) = \boldsymbol{\mu} + \mathbf{M}\mathbf{x} \tag{4.13}$$

The vector $\mathbf{x}_f^v$ is the verifying analysis at time t_f and serves as a proxy to the true state, and $\mathbf{E}$ is a diagonal matrix of weights. For our applications in stochastic volatility, $\mathbf{E} = \mathbf{I}_{p\times p}$, however, other appropriate definitions of $\mathbf{E}$ may include a covariance matrix representation used to weight the forecasts.

The first-order variation in the forecast aspect $e(\mathbf{x}^a)$ induced by the variation $\delta\mathbf{x}^a$

of the analysis is defined as

$$\delta e = \left\langle \frac{\partial e}{\partial \mathbf{x}^a}, \delta \mathbf{x}^a \right\rangle \tag{4.14}$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product of two vectors, $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v}$. By direct computation of equation (4.12), the forecast sensitivity to analysis is obtained as

$$\frac{\partial e}{\partial \mathbf{x}^a} = 2[\mathbf{M}_{k,f}^a]^T \mathbf{E} \left(\mathbf{x}_f - \mathbf{x}_f^v \right) \tag{4.15}$$

where $[\mathbf{M}_{k,f}^a]^T$ denotes the adjoint of the tangent linear model from time t_k to time t_f evaluated along the analysis trajectory; that is, if t_f is the $N-$step ahead forecast initiated from the analysis $\mathbf{x}^a$ at t_k, then

$$\mathbf{M}_{k,f}^a = \mathbf{M}^N \mathbf{M}^{N-1} \ldots \mathbf{M} \left(\mathbf{x}^a \right) \tag{4.16}$$

The analysis equations in (4.7) - (4.8) provide the basis in deriving forecast sensitivities with respect to the other input parameters and are used to express the first-order variation $\delta \mathbf{x}^a$ in (4.14) in terms of the other input variations; for example, in terms of variations in $\mathbf{R}$ and $\mathbf{B}$. Furthermore, it is noted that the forecast sensitivities with respect to a matrix $\mathbf{X} \in \mathbb{R}^{p \times p}$ is defined as the matrix of first-order partial derivatives

$$\frac{\partial e}{\partial \mathbf{X}} = \left[\frac{\partial e}{\partial \mathbf{X}_{ij}} \right]_{i,j=1:p} \in \mathbb{R}^{p \times p} \tag{4.17}$$

4.3 Forecast Sensitivities to Observations and Background

It was shown by Baker and Daley [6] that the forecast sensitivities with respect to observations and background within a linear analysis scheme are given as

$$\frac{\partial e}{\partial \mathbf{y}} = \mathbf{K}^T \frac{\partial e}{\partial \mathbf{x}^a} \in \mathbb{R}^p \tag{4.18}$$

$$\frac{\partial e}{\partial \mathbf{x}^b} = \left[\mathbf{I}_{n\times n} - \mathbf{H}^T\mathbf{K}^T\right] \frac{\partial e}{\partial \mathbf{x}^a} = \frac{\partial e}{\partial \mathbf{x}^a} - \mathbf{H}^T \frac{\partial e}{\partial \mathbf{y}} \in \mathbb{R}^n \tag{4.19}$$

Note from equation (4.18) that the identity matrix $\mathbf{I}_{n\times n}$ matches the dimension of the states $\mathbf{x}$ and the corresponding $\mathbf{x}^b-$sensitivity. On the other hand, the states (log-volatilities) in the MSV model have the same dimensions as the observations (log-returns) such that $n = p$. This need not be the case, thus, the sensitivities are presented as if the states had different dimensions as the observations; that is, it is assumed $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^p$. It is also noticed that the $\mathbf{x}^b-$sensitivity equation (4.19) is formally valid for a linear observation operator, $\mathbf{h}(\mathbf{x}) = \boldsymbol{\gamma} + \mathbf{H}\mathbf{x}$, since it neglects the dependence of $\mathbf{x}^b$ in the case of a linearized observation model

$$\mathbf{h}(\mathbf{x}) \approx \mathbf{h}(\mathbf{x}^b) + \mathbf{H}(\mathbf{x} - \mathbf{x}^b) \tag{4.20}$$

where

$$\mathbf{H} = \left[\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right]_{|\mathbf{x}=\mathbf{x}^b} \in \mathbb{R}^{p\times n} \tag{4.21}$$

Therefore, for non-linear models, equation (4.19) can be interpreted as a vector notation when deriving forecast sensitivities with respect to the background covariance matrix $\mathbf{B}$. For the case of nonlinear models, Daescu [16] presents the exact $\mathbf{x}^b-$sensitivity equations. Since the MSV model considered in this dissertation has both linear observation and state equations, equation (4.19) becomes valid for the

analysis.

4.4 Forecast R−Sensitivity

Daescu and Langland [17] derived the forecast sensitivity with respect to the observation error covariance matrix $\mathbf{R}$. In order to derive this forecast error sensitivity with respect to $\mathbf{R}$, the first-order variation $\delta\mathbf{x}^a$ induced by a perturbation $\delta\mathbf{R} \in \mathbb{R}^{p\times p}$ in the covariance model is calculated. From equation (4.7), the first-order variation with respect to a perturbation $\delta\mathbf{K}$ in the gain matrix is given by

$$\delta\mathbf{x}^a = \delta\mathbf{K}\left[\mathbf{y} - \mathbf{h}(\mathbf{x}^a)\right] \tag{4.22}$$

Now, the first-order variation $\delta\mathbf{K}$ induced by a perturbation $\delta\mathbf{R}$ in the observation covariance model can be calculated directly from equation (4.8) and it is given by

$$\delta\mathbf{K} = -\mathbf{B}\mathbf{H}^T\left[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right]^{-1}\delta\mathbf{R}\left[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right]^{-1} = -\mathbf{K}\delta\mathbf{R}\left[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right]^{-1} \tag{4.23}$$

where the identity $\delta\mathbf{X}^{-1} = -\mathbf{X}^{-1}\delta\mathbf{X}\mathbf{X}^{-1}$ was used. From equation (4.14), the forecast variation δe induced by a perturbation $\delta\mathbf{R}$ in the observation covariance model is given by

$$\delta e = -\left\langle \frac{\partial e}{\partial \mathbf{x}^a}, -\mathbf{K}\delta\mathbf{R}\left[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right]^{-1}\left[\mathbf{y} - \mathbf{h}(\mathbf{x}^b)\right]\right\rangle_{\mathbb{R}^n} \tag{4.24}$$

The adjoint $\mathbf{K}^T$ of the gain matrix may be used in equation (4.24) and equation (4.9) to get

$$\delta e = -\left\langle \mathbf{K}^T\frac{\partial e}{\partial \mathbf{x}^a}, \delta\mathbf{R}\left[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right]^{-1}\left[\mathbf{y} - \mathbf{h}(\mathbf{x}^b)\right]\right\rangle_{\mathbb{R}^p} = -\left\langle \underbrace{\mathbf{K}^T\frac{\partial e}{\partial \mathbf{x}^a}}_{\frac{\partial e}{\partial \mathbf{y}}}, \delta\mathbf{R}\mathbf{z}\right\rangle_{\mathbb{R}^p} \tag{4.25}$$

Therefore, the forecast $\mathbf{R}-$sensitivity is given by the rank-one matrix

$$\frac{\partial e}{\partial \mathbf{R}} = -\frac{\partial e}{\partial \mathbf{y}}\mathbf{z}^T \in \mathbb{R}^{p\times p} \tag{4.26}$$

4.4.1 R−Sensitivity with Scalar Inflation

The forecast sensitivity with respect to a parametric representation of the observation covariance matrix $\mathbf{R}$ is now considered in this section. When performing tuning of the covariance matrix $\mathbf{R}$, a practical approach is to define the parametric representation of the covariance matrix

$$\mathbf{R}(s^o) = s^o\mathbf{R} \tag{4.27}$$

where $s^o > 0$ is a scalar coefficient used to adjust the information provided by the observations $\mathbf{y}$. The variations $\delta\mathbf{R}$ in the covariance model due to variations δs^o in the parameter s^o can be represented as

$$\delta\mathbf{R} = \delta s^o\mathbf{R} \tag{4.28}$$

Substituting equation (4.28) into equation (4.25) we get

$$\delta e = -\left\langle \frac{\partial e}{\partial \mathbf{y}}, \delta s^o\mathbf{R}\mathbf{z} \right\rangle_{\mathbb{R}^p} \tag{4.29}$$

Therefore, the forecast sensitivity due to a variation in the covariance inflation parameter s^o is given by

$$\frac{\partial e}{\partial s^o} = -(\mathbf{R}\mathbf{z})^T\frac{\partial e}{\partial \mathbf{y}} \tag{4.30}$$

An alternative formulation to equation (4.30) can be obtain from equations (4.9)-(4.10). From equation (4.10)

$$\mathbf{H}(\mathbf{x}^a - \mathbf{x}^b) = \mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{z} \tag{4.31}$$

Which can then be substituted into equation (4.9) to obtain

$$\mathbf{R}\mathbf{z} = \mathbf{y} - \mathbf{h}(\mathbf{x}^b) + \mathbf{H}(\mathbf{x}^b - \mathbf{x}^a) \tag{4.32}$$

By definition, $\mathbf{h}(\mathbf{x}^b) = \boldsymbol{\gamma} + \mathbf{H}\mathbf{x}^b$. Thus, equation (4.32) becomes

$$\mathbf{R}\mathbf{z} = \mathbf{y} - \mathbf{h}(\mathbf{x}^a) \tag{4.33}$$

Substituting equation (4.33) into equation (4.30), the alternative formulation of the forecast sensitivity to the observation error covariance scaling is

$$\frac{\partial e}{\partial s^o} = (\mathbf{h}(\mathbf{x}^a) - \mathbf{y})^T \frac{\partial e}{\partial \mathbf{y}} \tag{4.34}$$

4.4.2 R−Sensitivity with Matrix Decomposition

New forecast sensitivity results with respect to matrix square root decompositions are derived in this section as an extension to the results found in the work of Daescu and Langland [17]. When conducting the q-MLE estimation procedure, it is common practice to estimate the square root of the covariance matrix rather than the covariance matrix itself. Therefore, it is useful to also obtain the forecast $\mathbf{R}-$sensitivity with respect to the matrix decomposition $\mathbf{R}^{1/2}$, where $\mathbf{R}^{1/2}$ satisfies $\mathbf{R} = \mathbf{R}^{1/2}\mathbf{R}^{T/2}$. From equation (4.23), the variation $\delta\mathbf{K}$ induced by a perturbation $\delta\mathbf{R}^{1/2}$ in the de-

composition $\mathbf{R} = \mathbf{R}^{1/2}\mathbf{R}^{T/2}$ is given by

$$\delta\mathbf{R} = \delta\mathbf{R}^{1/2}\mathbf{R}^{T/2} + \mathbf{R}^{1/2}(\delta\mathbf{R}^{1/2})^T \tag{4.35}$$

Recall that for any two vectors $\mathbf{u} \in \mathbb{R}^p$, $\mathbf{v} \in \mathbb{R}^n$, and any matrix $\mathbf{X} \in \mathbb{R}^{p\times n}$, the following relationship between inner products holds

$$\langle \mathbf{u}, \mathbf{X}\mathbf{v}\rangle_{\mathbb{R}^p} = \left\langle \mathbf{u}\mathbf{v}^T, \mathbf{X}\right\rangle_{\mathbb{R}^{p\times n}} \tag{4.36}$$

Thus, by using the relationship (4.36) in equation (4.25)

$$\begin{aligned}
\delta e &= -\left\langle \frac{\partial e}{\partial \mathbf{y}}\mathbf{z}^T, \delta\mathbf{R}\right\rangle_{\mathbb{R}^{p\times p}} \\
&= -\left\langle \frac{\partial e}{\partial \mathbf{y}}\mathbf{z}^T, \delta\mathbf{R}^{1/2}\mathbf{R}^{T/2} + \mathbf{R}^{1/2}(\delta\mathbf{R}^{1/2})^T\right\rangle_{\mathbb{R}^{p\times p}} \\
&= -\mathrm{Tr}\left(\frac{\partial e}{\partial \mathbf{y}}\mathbf{z}^T\mathbf{R}^{1/2}[\delta\mathbf{R}^{1/2}]^T\right) - \mathrm{Tr}\left(\frac{\partial e}{\partial \mathbf{y}}\mathbf{z}^T\delta\mathbf{R}^{1/2}\mathbf{R}^{T/2}\right) \\
&= -\mathrm{Tr}\left(\frac{\partial e}{\partial \mathbf{y}}\mathbf{z}^T\mathbf{R}^{1/2}[\delta\mathbf{R}^{1/2}]^T\right) - \mathrm{Tr}\left(\mathbf{R}^{T/2}\frac{\partial e}{\partial \mathbf{y}}\mathbf{z}^T\delta\mathbf{R}^{1/2}\right) \\
&= -\mathrm{Tr}\left(\left[\frac{\partial e}{\partial \mathbf{y}}\mathbf{z}^T\mathbf{R}^{1/2}\right][\delta\mathbf{R}^{1/2}]^T\right) - \mathrm{Tr}\left(\left[\mathbf{z}\left(\frac{\partial e}{\partial \mathbf{y}}\right)^T\mathbf{R}^{1/2}\right]^T[\delta\mathbf{R}^{1/2}]\right) \\
&= -\mathrm{Tr}\left(\left[\frac{\partial e}{\partial \mathbf{y}}\mathbf{z}^T\mathbf{R}^{1/2}\right][\delta\mathbf{R}^{1/2}]^T\right) - \mathrm{Tr}\left(\left[\mathbf{z}\left(\frac{\partial e}{\partial \mathbf{y}}\right)^T\mathbf{R}^{1/2}\right][\delta\mathbf{R}^{1/2}]^T\right)
\end{aligned} \tag{4.37}$$

Therefore, the variation in the forecast error induced by the perturbation $\delta\mathbf{R}^{1/2}$ is

$$\delta e = -\left\langle \left(\frac{\partial e}{\partial \mathbf{y}}\mathbf{z}^T + \mathbf{z}\left(\frac{\partial e}{\partial \mathbf{y}}\right)^T\right)\mathbf{R}^{1/2}, \delta\mathbf{R}^{1/2}\right\rangle_{\mathbb{R}^{p\times p}}$$

Thus, it is concluded that the $\mathbf{R}^{1/2}-$sensitivity is given by

$$\frac{\partial e}{\partial \mathbf{R}^{1/2}} = -\left(\frac{\partial e}{\partial \mathbf{y}}\mathbf{z}^T + \mathbf{z}\left(\frac{\partial e}{\partial \mathbf{y}}\right)^T\right)\mathbf{R}^{1/2} \tag{4.38}$$

4.5 Forecast B−Sensitivity

Daescu and Langland [17] derived the forecast sensitivity with respect to the background error covariance matrix $\mathbf{B}$ and their results are derived in this section. First, the calculation of the first-order variation $\delta\mathbf{x}^a$ induced by a perturbation $\delta\mathbf{B} \in \mathbb{R}^{n\times n}$ in the background error covariance model is needed. This in turn relies on deriving the first-order variation $\delta\mathbf{K}$ of the gain matrix in equation (4.8) induced by the first-order variation $\delta\mathbf{B}$ in the background covariance model and it is given by

$$\begin{aligned}\delta\mathbf{K} &= \delta\mathbf{B}\mathbf{H}^T\left[\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R}\right]^{-1} - \mathbf{B}\mathbf{H}^T\left[\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R}\right]^{-1}\mathbf{H}\delta\mathbf{B}\mathbf{H}^T\left[\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R}\right]^{-1} \\ &= \left[\mathbf{I}_{n\times n} - \mathbf{K}\mathbf{H}\right]\delta\mathbf{B}\mathbf{H}^T\left[\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R}\right]^{-1}\end{aligned} \tag{4.39}$$

where we once again made use of the identity $\delta\mathbf{X}^{-1} = -\mathbf{X}^{-1}\delta\mathbf{X}\mathbf{X}^{-1}$. From equation (4.14), equation (4.22) and equation (4.39), the forecast variation δe induced by a variation $\delta\mathbf{B}$ in the background covariance model is given by

$$\begin{aligned}\delta e &= \left\langle \frac{\partial e}{\partial \mathbf{x}^a}, \left[\mathbf{I}_{n\times n} - \mathbf{K}\mathbf{H}\right]\delta\mathbf{B}\mathbf{H}^T\left[\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R}\right]^{-1}\left[\mathbf{y}-\mathbf{h}(\mathbf{x}^b)\right]\right\rangle_{\mathbb{R}^n} \\ &= \left\langle \left[\mathbf{I}_{n\times n} - \mathbf{K}\mathbf{H}\right]^T\frac{\partial e}{\partial \mathbf{x}^a}, \delta\mathbf{B}\mathbf{H}^T\left[\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R}\right]^{-1}\left[\mathbf{y}-\mathbf{h}(\mathbf{x}^b)\right]\right\rangle_{\mathbb{R}^n}\end{aligned} \tag{4.40}$$

Where we have made use of the adjoint of the matrix $[\mathbf{I}_{n\times n} - \mathbf{KH}]$ between inner products. By using equation (4.9) and equation (4.19), equation (4.40) becomes

$$\delta e = \left\langle \frac{\partial e}{\partial \mathbf{x}^b}, \delta \mathbf{B}\mathbf{H}^T\mathbf{z} \right\rangle_{\mathbb{R}^n} \tag{4.41}$$

Therefore, the forecast $\mathbf{B}-$sensitivity becomes the rank-one matrix

$$\frac{\partial e}{\partial \mathbf{B}} = \frac{\partial e}{\partial \mathbf{x}^b}\left(\mathbf{H}^T\mathbf{z}\right)^T \in \mathbb{R}^{n\times n} \tag{4.42}$$

where the inner product property in equation (4.36) was once again employed.

4.5.1 B−Sensitivity with Scalar Inflation

Consider now the forecast sensitivity with respect to a parametric representation of the background covariance matrix $\mathbf{B}$. When performing tuning of the covariance matrix $\mathbf{B}$, a practical approach is to define the parametric representation of the covariance matrix

$$\mathbf{B}(s^b) = s^b\mathbf{B} \tag{4.43}$$

where $s^b > 0$ is a scalar coefficient used to adjust the information provided by the background information. The variations $\delta\mathbf{B}$ in the covariance model due to variations δs^b in the parameter s^b can be represented as

$$\delta\mathbf{B} = \delta s^b\mathbf{B} \tag{4.44}$$

Substituting equation (4.44) into equation (4.41) and with the aid of the analysis equation (4.10), the first-order variation in the forecast aspect δe can be expressed in

terms of the inflation parameter s^b as

$$\delta e = \delta s^b \left\langle \frac{\partial e}{\partial \mathbf{x}^b}, \mathbf{x}^a - \mathbf{x}^b \right\rangle_{\mathbb{R}^n} \tag{4.45}$$

Therefore, the forecast sensitivity with respects to the covariance inflation parameter s^b is given by

$$\frac{\partial e}{\partial s^b} = (\mathbf{x}^a - \mathbf{x}^b)^T \frac{\partial e}{\partial \mathbf{x}^b} \tag{4.46}$$

It is also noted that the following relationship can be derived from equation (4.6) by evaluating the Jacobian at $\mathbf{x}^a$; that is, $\nabla J(\mathbf{x}^a) = 0$ and using equation (4.18)-(4.19):

$$[\mathbf{h}(\mathbf{x}^a) - \mathbf{y}]^T \frac{\partial e}{\partial \mathbf{y}} + (\mathbf{x}^a - \mathbf{x}^b)^T \frac{\partial e}{\partial \mathbf{x}^b} = 0 \tag{4.47}$$

Equation (4.47) is an intrinsic property of the variational problem (4.4) and can be used to derive an alternative formulation of the s^b-sensitivity in terms of an observation space equation. Substituting equation (4.47) into equation (4.46), an observation space equation to the s^b-sensitivity can be written as

$$\frac{\partial e}{\partial s^b} = [\mathbf{y} - \mathbf{h}(\mathbf{x}^a)]^T \frac{\partial e}{\partial \mathbf{y}} \tag{4.48}$$

Consequently, from equation (4.34) and equation (4.48), the following identity is derived

$$\frac{\partial e}{\partial s^o} + \frac{\partial e}{\partial s^b} = 0 \tag{4.49}$$

and reflects the intrinsic property of the variational optimization problem: multiplication of all error covariances in the system by the same (positive) constant has no impact on the analysis.

4.5.2 B−Sensitivity with Matrix Decomposition

New forecast sensitivity results with respect to matrix square root decompositions are derived in this section as an extension to the results found in the work of Daescu and Langland [17]. The forecast sensitivity is derived in this section with respects to the square root decomposition of the background error covariance matrix $\mathbf{B}$; that is, with respect to decomposition $\mathbf{B}^{1/2}$ where $\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{T/2}$. The steps are similar to those in deriving the sensitivity with respects to $\mathbf{R}^{1/2}$. The variations in $\mathbf{B}$ to variations in the square root matrix $\mathbf{B}^{1/2}$ are given by

$$\delta\mathbf{B}^{1/2} = \delta\mathbf{B}^{1/2}\mathbf{B}^{T/2} + \mathbf{B}^{1/2}\left(\delta\mathbf{B}^{1/2}\right)^T \tag{4.50}$$

Substituting equation (4.50) into equation (4.41), with the aid of identity (4.36) and the properties of the trace operator

$$\begin{aligned}
\delta e &= \left\langle \frac{\partial e}{\partial \mathbf{x}^b}\left(\mathbf{H}^T\mathbf{z}\right)^T, \underbrace{\delta\mathbf{B}^{1/2}\mathbf{B}^{T/2} + \mathbf{B}^{1/2}\left(\delta\mathbf{B}^{1/2}\right)^T}_{\delta\mathbf{B}} \right\rangle_{\mathbb{R}^{n\times n}} \\
&= \mathrm{Tr}\left(\left[\frac{\partial e}{\partial \mathbf{x}^b}\left(\mathbf{H}^T\mathbf{z}\right)^T\right]\left[\delta\mathbf{B}^{1/2}\mathbf{B}^{T/2}\right]^T\right) + \mathrm{Tr}\left(\left[\frac{\partial e}{\partial \mathbf{x}^b}\left(\mathbf{H}^T\mathbf{z}\right)^T\right]\left[\mathbf{B}^{1/2}\left(\delta\mathbf{B}^{1/2}\right)^T\right]^T\right) \\
&= \mathrm{Tr}\left(\left[\frac{\partial e}{\partial \mathbf{x}^b}\left(\mathbf{H}^T\mathbf{z}\right)^T\mathbf{B}^{1/2}\right]\left[\delta\mathbf{B}^{1/2}\right]^T\right) + \mathrm{Tr}\left(\left[\mathbf{B}^{T/2}\frac{\partial e}{\partial \mathbf{x}^b}\left(\mathbf{H}^T\mathbf{z}\right)^T\right]\left(\delta\mathbf{B}^{1/2}\right)\right) \\
&= \mathrm{Tr}\left(\left[\frac{\partial e}{\partial \mathbf{x}^b}\left(\mathbf{H}^T\mathbf{z}\right)^T\mathbf{B}^{1/2}\right]\left[\delta\mathbf{B}^{1/2}\right]^T\right) + \mathrm{Tr}\left(\left[\left(\mathbf{H}^T\mathbf{z}\right)\left(\frac{\partial e}{\partial \mathbf{x}^b}\right)^T\mathbf{B}^{1/2}\right]^T\left(\delta\mathbf{B}^{1/2}\right)\right) \\
&= \mathrm{Tr}\left(\left[\frac{\partial e}{\partial \mathbf{x}^b}\left(\mathbf{H}^T\mathbf{z}\right)^T\mathbf{B}^{1/2}\right]\left[\delta\mathbf{B}^{1/2}\right]^T\right) + \mathrm{Tr}\left(\left[\left(\mathbf{H}^T\mathbf{z}\right)\left(\frac{\partial e}{\partial \mathbf{x}^b}\right)^T\mathbf{B}^{1/2}\right]\left[\delta\mathbf{B}^{1/2}\right]^T\right) \\
&= \left\langle \left[\frac{\partial e}{\partial \mathbf{x}^b}\left(\mathbf{H}^T\mathbf{z}\right)^T + \left(\mathbf{H}^T\mathbf{z}\right)\left(\frac{\partial e}{\partial \mathbf{x}^b}\right)^T\right]\mathbf{B}^{1/2}, \delta\mathbf{B}^{1/2}\right\rangle_{\mathbb{R}^{n\times n}}
\end{aligned} \tag{4.51}$$

Therefore, it is concluded that the forecast sensitivity with respects to the matrix decomposition $\mathbf{B}^{1/2}$ is given by

$$\frac{\partial e}{\partial \mathbf{B}^{1/2}} = \left[\frac{\partial e}{\partial \mathbf{x}^b} \left(\mathbf{H}^T \mathbf{z} \right)^T + \left(\mathbf{H}^T \mathbf{z} \right) \left(\frac{\partial e}{\partial \mathbf{x}^b} \right)^T \right] \mathbf{B}^{1/2} \tag{4.52}$$

4.6 Summary of Forecast Sensitivities

A summary of equations used to evaluate the forecast sensitivity with respect to various input parameters of a data assimilation system with a single outer loop iteration is provided in table 4.1.

Parameter	Significance	Dimension	Forecast Sensitivity
$\mathbf{y}$	Observation vector	$\mathbb{R}^p$	$\mathbf{K}^T \dfrac{\partial e}{\partial \mathbf{x}^a}$
$\mathbf{x}^b$	Background state vector Vector	$\mathbb{R}^n$	$\dfrac{\partial e}{\partial \mathbf{x}^a} - \mathbf{H}^T \dfrac{\partial e}{\partial \mathbf{y}}$
$\mathbf{R}$	Observation error covariance model	$\mathbb{R}^{p\times p}$	$-\dfrac{\partial e}{\partial \mathbf{y}} \mathbf{z}^T$
s^o	Observation error covariance weight	$\mathbb{R}^1$	$[\mathbf{h}(\mathbf{x}^a) - \mathbf{y}]^T \dfrac{\partial e}{\partial \mathbf{y}}$
$\mathbf{R}^{1/2}$	Observation error covariance square root	$\mathbb{R}^{p\times p}$	$-\left(\dfrac{\partial e}{\partial \mathbf{y}} \mathbf{z}^T + \mathbf{z}\left(\dfrac{\partial e}{\partial \mathbf{y}}\right)^T\right) \mathbf{R}^{1/2}$
$\mathbf{B}$	Background error covariance model	$\mathbb{R}^{n\times n}$	$\dfrac{\partial e}{\partial \mathbf{x}^b} (\mathbf{H}^T \mathbf{z})^T$
s^b	Background error covariance weight	$\mathbb{R}^1$	$[\mathbf{y} - \mathbf{h}(\mathbf{x}^a)]^T \dfrac{\partial e}{\partial \mathbf{y}}$
$\mathbf{B}^{1/2}$	Background error covariance square root	$\mathbb{R}^{n\times n}$	$\left[\dfrac{\partial e}{\partial \mathbf{x}^b} \left(\mathbf{H}^T \mathbf{z}\right)^T + \left(\mathbf{H}^T \mathbf{z}\right) \left(\dfrac{\partial e}{\partial \mathbf{x}^b}\right)^T\right] \mathbf{B}^{1/2}$

Table 4.1: Forecast sensitivity to various input parameters of a data assimilation system with a single outer loop iteration

4.6.1 Forecast Sensitivity of the MSV Model

New results are derived in this section for the forecast sensitivity analysis with respect to the various inputs for the Multivariate Stochastic Volatility (MSV) model of equation (4.1). If the parameters of the MSV model $\{\boldsymbol{\mu}, \mathbf{M}, \mathbf{R}, \mathbf{Q}, \mathbf{P}_0\}$ are all known, then a Kalman filter may be implemented to obtain estimates of the unobservable state sequence $\mathbf{x}_t$ of log-volatilities, for each time iteration t. Furthermore, this Kalman filter solves the data assimilation problem of equation (4.4) at each iteration t, therefore, the forecast sensitivities of the MSV model can be derived. To this end, define the following variables:

- Let $\mathbf{x}^b = \mathbf{x}_{t|t-1}$ be the one-step ahead state model prediction from $t-1$ and let $\mathbf{x}^a = \mathbf{x}_{t|t}$ be the analysis of the states (log-volatilities) obtained from the Kalman filter at each iteration t. See Section 2.3 for more information on the Kalman filter and its notation. Notice that, although $\mathbf{x}_{t|t-1}$ is the state model prediction from the previous time step $t-1$, at the new iteration t, $\mathbf{x}_{t|t-1}$ now becomes the background estimate of the states.

- Let $\mathbf{H} = \mathbf{I}_{p\times p}$ be the identity observation operator that translates the states to observations. Recall that within the MSV model, the states have the same dimension as the observations ($n = p$) since we seek the stochastic volatility of *each* financial variable.

- Let $\mathbf{B} = \mathbf{P}_{t|t-1}$ be the one-step ahead state error covariance matrix prediction from $t-1$ obtained from the Kalman filter. Notice that, although $\mathbf{P}_{t|t-1}$ is the state covariance prediction from the previous time step $t-1$, at the new iteration t, $\mathbf{P}_{t|t-1}$ now becomes the background covariance estimate of the states.

- The observation error covariance matrix $\mathbf{R}$ of the transformed time series $\tilde{\mathbf{y}}_t$ remains unchanged and it represents the *precision* of the normal distribution approximation to the multivariate log of χ^2 distribution.

The sensitivity analysis equations for the Multivariate Stochastic Volatility using the Kalman filter are give in Table 4.2. It is noted that the Kalman filter solution, at each iteration t, for the MSV model with known parameters and known prior distribution $\mathbf{x}_{1|0} \sim N_p(\mathbf{0}, \mathbf{P}_0)$ is given as follow:

Parameter	Significance	Dimension	Forecast Sensitivity
$\tilde{\mathbf{y}}_t$	Observation vector	$\mathbb{R}^p$	$\mathbf{K}^T \frac{\partial e}{\partial \mathbf{x}_{t\|t}}$
$\mathbf{x}_{t\|t-1}$	Background state vector Vector	$\mathbb{R}^n$	$\frac{\partial e}{\partial \mathbf{x}_{t\|t}} - \frac{\partial e}{\partial \tilde{\mathbf{y}}_t}$
$\mathbf{R}$	Observation error covariance model	$\mathbb{R}^{p\times p}$	$-\frac{\partial e}{\partial \tilde{\mathbf{y}}_t}\mathbf{z}_t^T$
s^o	Observation error covariance weight	$\mathbb{R}^1$	$\left[\mathbf{h}(\mathbf{x}_{t\|t}) - \tilde{\mathbf{y}}_t\right]^T \frac{\partial e}{\partial \tilde{\mathbf{y}}_t}$
$\mathbf{R}^{1/2}$	Observation error covariance square root	$\mathbb{R}^{p\times p}$	$-\left(\frac{\partial e}{\partial \tilde{\mathbf{y}}_t}\mathbf{z}_t^T + \mathbf{z}_t\left(\frac{\partial e}{\partial \tilde{\mathbf{y}}_t}\right)^T\right)\mathbf{R}^{1/2}$
$\mathbf{P}_{t\|t-1}$	Background error covariance model	$\mathbb{R}^{n\times n}$	$\frac{\partial e}{\partial \mathbf{x}_{t\|t-1}}\mathbf{z}_t^T$
s^b	Background error covariance weight	$\mathbb{R}^1$	$\left[\tilde{\mathbf{y}}_t - \mathbf{h}(\mathbf{x}_{t\|t})\right]^T \frac{\partial e}{\partial \tilde{\mathbf{y}}_t}$
$\mathbf{P}_{t\|t-1}^{1/2}$	Background error covariance square root	$\mathbb{R}^{n\times n}$	$\left[\frac{\partial e}{\partial \mathbf{x}_{t\|t-1}}\mathbf{z}_t^T + \mathbf{z}_t\left(\frac{\partial e}{\partial \mathbf{x}_{t\|t-1}}\right)^T\right]\mathbf{P}_{t\|t-1}^{1/2}$

Table 4.2: Forecast sensitivity to various input parameters of a MSV Model within a data assimilation system with a single time outer loop iteration

Between observations,

$$\mathbf{x}_{t+1|t} = \mathcal{M}(\mathbf{x}_{t|t}), \tag{4.53}$$

$$\mathbf{P}_{t+1|t} = \mathbf{M}\mathbf{P}_{t|t}\mathbf{M}^T + \mathbf{Q} \tag{4.54}$$

At observations,

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathbf{K}(\tilde{\mathbf{y}}_t - \mathbf{x}_{t|t-1}), \tag{4.55}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}\mathbf{P}_{t|t-1} \tag{4.56}$$

where the gain matrix is given by

$$\mathbf{K} = \mathbf{P}_{t|t-1}[\mathbf{P}_{t|t-1} + \mathbf{R}]^{-1} \tag{4.57}$$

Moreover, the vector $\mathbf{z}_t$ is defined as the solution at each iteration t to the linear system

$$\left[\mathbf{P}_{t|t-1} + \mathbf{R}\right] \mathbf{z}_t = \tilde{\mathbf{y}}_t - \mathbf{h}(\mathbf{x}_{t|t-1}) \tag{4.58}$$

where the observation operator is defined as $\mathbf{h}(\mathbf{x}) = (-1.27)\mathbf{1}_p + \mathbf{x}$. The forecast score at each iteration t, defined as the as the one-step ahead short-range forecast-error measure, is given as

$$e(\mathbf{x}_{t|t}) = \left(\mathbf{x}_{t+1|t} - \mathbf{x}^v_{t+1}\right)^T \left(\mathbf{x}_{t+1|t} - \mathbf{x}^v_{t+1}\right) \tag{4.59}$$

where $\mathbf{x}_{t+1|t} = \mathcal{M}(\mathbf{x}_{t|t})$ is the state forecast at verification time $t+1$ initiated from $\mathbf{x}_{t|t}$ at time of analysis t. As discussed before, the vector $\mathbf{x}^v_{t+1}$ is the verifying analysis at time $t+1$ and serves as a proxy to the true state of log-volatilities. The forecast sensitivity to analysis within this framework is therefore given by

$$\frac{\partial e}{\partial \mathbf{x}_{t|t}} = 2\mathbf{M}^T \left(\mathbf{x}_{t+1|t} - \mathbf{x}^v_{t+1}\right) \tag{4.60}$$

The forecast sensitivities can now be evaluated as in Table 4.2 given the information provided.

4.7 Adaptive Tuning of Covariance Parameters

The work of Song et al. [50] provide an adjoint-based approach to achieving adaptive tuning of the background error covariance matrix specification within a hybrid

ensemble Kalman filter. The forecast error sensitivity equations in Table 4.1 or the forecast error sensitivities of the MSV model in Table 4.2 not only provide a tool for diagnosing model performance, but also provide tools for tuning model error covariance matrices. The derivative information provided in the Table 4.2 provide guidance on reducing the forecast error e using a steepest descent direction in an update of the form

$$\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} - \alpha_k \frac{\partial e(\mathbf{x}^a)}{\partial \mathbf{X}}\bigg|_{\mathbf{X}=\mathbf{X}^{(k)}} \tag{4.61}$$

where $\mathbf{X}$ denotes the value of the parameter of interest; for example, $\mathbf{X} = \mathbf{R}$ if tuning of the observation error covariance matrix is desired. In our analysis, it is noted that the linearization of the MSV model is done in terms of the observation equation. This linearized observation equation will have a non-Gaussian probability distribution due to the log transformations. Thus, it is of interest to our work to provide an adaptive tuning of the observation error covariance matrix $\mathbf{R}$ since the accuracy of the log-volatility estimates $\mathbf{x}_{t|t}$ will be dependent on accurately modeling the covariance $\mathbf{R}$. This can be seen from the quasi-MLE method as this method tries to approximate the log of a χ^2 distribution with a normal distribution by optimally specifying $\mathbf{R}$. Although an all at once estimate of $\mathbf{R}$ can produce reasonably good volatility estimates within the quasi-MLE method, an adaptive tuning of $\mathbf{R}$ may provide improved volatility estimates during each iteration. In what follows, a practical implementation of adaptive tuning of the observation error covariance matrix $\mathbf{R}$ is presented. Note, however, that this can also be applied for tuning the background error covariance matrix with implications of updating prior estimates during each time iteration.

The quasi-MLE method is implemented for a time window $[t_0, t_N]$ and produces MLE estimates of the unknown parameters. From the estimated model (unknown parameters have been estimated), an analysis sequence for the time window $[t_0, t_N]$

can be provided based on the Kalman filter. The forecast score function can then be evaluated at each time iteration $t \in [t_0, t_N]$ to produce diagnostics of model performance. The proxy selected for evaluation of the forecast score function is subjective and left for the analyst to provide input on their real-world view of forecasted volatilities. Evaluation of the forecast score at each time iteration provides a sequence of forecasted errors from which the corresponding sensitivities can be evaluated in terms of the quasi-MLE estimates. Using the steepest descent algorithm, a new parameter estimate can be obtained via equation (4.61) during each time iteration t. The quality of the new parameter estimate is then evaluated via the forecast score function. If the forecast score is reduced during the time iteration t, then the step size α is updated and the parameter estimate is accepted as an improved parameter to obtain a new analysis based on improved parameters that will then be used for the next time iteration $t+1$. Otherwise, a new steepest descent iteration is defined based on a new step size α. It is noted that in practice, data assimilation is computationally expensive, thus, it is impractical to implement more than one iteration of the steepest descent algorithm during each time iteration t. Therefore, if during an iteration t the forecast score cannot be reduced, the update is rejected and the quasi-MLE estimates are accepted.

The selection of the step size α is initialized as $\alpha_0 = 1$, and is dynamically updated according [44]. Thus, if the forecast error is reduced and the new parameter estimate is accepted, the step-size is updated as

$$\alpha = \frac{2\left(e(\mathbf{x}^a) - e(\mathbf{x}_{\text{new}})\right)}{\left\| \frac{\partial e}{\partial \mathbf{X}} \right\|^2} \tag{4.62}$$

which is the step-size obtained from minimizing the quadratic function based on the data $e(\mathbf{x}^a)$, $e(\mathbf{x}_{\text{new}})$ and $\left\| \frac{\partial e}{\partial \mathbf{X}} \right\|^2$. Algorithm 4.1 provides the procedure for the new

results of adaptive tuning of the observation error covariance matrix $\mathbf{R}^{1/2}$ within the Multivariate Stochastic Volatility model.

Algorithm 4.1: Procedure to update the model error covariance matrix using sensitivities

Input: Estimated model parameters $\{\boldsymbol{\mu}, \boldsymbol{\gamma}, \mathbf{M}, \mathbf{H}, \mathbf{R}, \mathbf{Q}\}$, background state $\mathbf{x}^b$ and background covariance matrix $\mathbf{B}$ at current assimilation iteration and observations $\mathbf{y}$ at current assimilation iteration

Output: Updated error covariance $\mathbf{R}$ and updated analysis state $\mathbf{x}^a$

procedure UPDATE $\mathbf{R}$ COVARIANCE($\mathbf{R}, \alpha_0$)
  $\alpha = \alpha_0$ ▷ Initialize step-size
  $\mathbf{x}^a = \mathrm{KF}(\mathbf{x}^b, \mathbf{y}, \mathbf{B}, \mathbf{R}, \mathbf{Q})$ ▷ Use Kalman Filter to obtain analysis
  $\mathbf{p} = -\dfrac{\partial e}{\partial \mathbf{R}^{1/2}}$ ▷ Set the search direction in terms of square-root matrix
  $\mathbf{R}^{1/2}_{\text{New}} = \mathbf{R}^{1/2} + \alpha \mathbf{p}$ ▷ Calculate new square-root matrix
  $\mathbf{R}_{\text{New}} = \mathbf{R}^{1/2}_{\text{New}} \mathbf{R}^{\mathrm{T}/2}_{\text{New}}$ ▷ Calculate new covariance
  $\mathbf{x}^a_{\text{New}} = \mathrm{KF}(\mathbf{x}^b, \mathbf{y}, \mathbf{B}, \mathbf{R}_{\text{New}}, \mathbf{Q})$ ▷ Redo analysis with new specification
  $\epsilon = e(\mathbf{x}^a_{\text{New}}) - e(\mathbf{x}^a)$ ▷ Evaluate forecast error impact
  if $\epsilon < 0$ **then** ▷ Forecast error was decreased
    $\mathbf{R} = \mathbf{R}_{\text{New}}$ ▷ Update observation error covariance
    $\mathbf{x}^a = \mathbf{x}^a_{\text{New}}$ ▷ Update analysis state
    $\alpha = \dfrac{2\left(e(\mathbf{x}^a) - e(\mathbf{x}_{\text{new}})\right)}{\left\| \frac{\partial e}{\partial \mathbf{x}} \right\|^2}$ ▷ Update step-size for next assimilation iteration
  else ▷ Reject the updated observation error covariance
    $\alpha = 0.5\alpha$ ▷ Reduce the step-size for next assimilation iteration
  end if
  Continue to next assimilation iteration
end procedure

Chapter 5

Numerical Experiments

Numerical experiments with the Multivariate Stochastic Volatility model are presented in this section to test the q-MLE estimation methodology and to be able to provide model performance diagnostics in terms of sensitivity analysis. The sensitivity analysis is extended to include an adaptive tuning of the observation error covariance matrix $\mathbf{R}$ in order to provide improved volatility estimates. The validity and benefits of these approaches are shown in two numerical experiments. First numerical experiment provides a proof-of-concept where the simulation of volatility data and observations, in accordance to the Multivariate Stochastic Volatility model, is generated to estimate the model and perform sensitivity analysis and adaptive tuning. Diagnostics of the proof-of-concept experiments are performed in terms of the "true" simulated volatility to demonstrate superiority of the adaptive tuning procedure when estimating volatilities with the *correct* volatility proxy. The second numerical experiment extends these methodology to a set of foreign exchange rate time series data where the Multivariate Stochastic Volatility model is estimated based on the q-MLE methodology, and sensitivity analysis is performed to provide diagnostics of model performance. Since volatility data of foreign exchange rates are not observable, the selection of appropriate proxies is discussed for adaptive tuning of $\mathbf{R}$ to provide improved volatility estimates.

5.1 Proof-of-Concept

For the first numerical experiment, a time series of 4 financial variables ($p = 4$) $\mathbf{y}_t = (y_t^{(1)}, y_t^{(2)}, y_t^{(3)}, y_t^{(4)})^T$ is considered with their corresponding log-volatility time series $\mathbf{x}_t = (x_t^{(1)}, x_t^{(2)}, x_t^{(3)}, x_t^{(4)})^T$. The Multivariate Stochastic Volatility Model (5.1) is then used to generate the time series $\{\mathbf{y}_t\}$ and $\{\mathbf{x}_t\}$ for a training window and a test window each containing 900 daily observations for a total of 1800 days; that is, for a total of 5 years. The MSV model considered for the data generation is

$$\begin{aligned} \mathbf{y}_t &= \mathbf{V}_t^{1/2}\boldsymbol{\epsilon}_t \\ \mathbf{x}_{t+1} &= \mathbf{M}\mathbf{x}_t + \boldsymbol{\eta}_t, \\ \mathbf{x}_0 &\sim N_p(\mathbf{0}, \mathbf{P}_0) \end{aligned} \tag{5.1}$$

with $\mathbf{x}_0 \in \mathbb{R}^p$ as the initial prior, $\mathbf{P}_0 \in \mathbb{R}^{p\times p}$ as the initial state covariance matrix and

$$\begin{aligned} \mathbf{V}_t^{1/2} &= \operatorname{diag}(\exp(x_t^{(1)}/2), \exp(x_t^{(2)}/2), \exp(x_t^{(3)}/2), \exp(x_t^{(4)}/2)) \\ \mathbf{M} &= \mathbf{I}_{4\times 4} \end{aligned} \tag{5.2}$$

Furthermore, the random processes are defined as

$$\boldsymbol{\epsilon}_t \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_{\epsilon\epsilon}), \qquad \boldsymbol{\eta}_t \sim N_p(\mathbf{0}, \mathbf{Q}) \tag{5.3}$$

with covariance matrices taken as

$$\mathbf{\Sigma}_{\epsilon\epsilon} = \begin{bmatrix} 1 & 0.84 & 0.74 & 0.80 \\ 0.84 & 1 & 0.84 & 0.92 \\ 0.74 & 0.84 & 1 & 0.81 \\ 0.80 & 0.92 & 0.81 & 1 \end{bmatrix}, \qquad \mathbf{Q} = 10^{-3} \begin{bmatrix} 9.65 & 11.42 & 3.97 & 12.07 \\ 11.42 & 20.43 & 5.44 & 21.09 \\ 3.97 & 5.44 & 5.45 & 7.08 \\ 12.07 & 21.09 & 7.08 & 22.31 \end{bmatrix} \tag{5.4}$$

Since the work of Harvey et al. [29] provided the first applications of the MSV model, it is noted that the covariance matrices in (5.4) are selected to match the covariances estimated in their publication. Taking $\mathbf{x}_0 = \mathbf{0}$, the MSV model (5.1) with the known parameters is used to generate the 1800 observations $\mathbf{y}_t$ with their corresponding log-volatilities $\mathbf{x}_t$. The resulting simulated time series of volatilities for each financial variable, $i = 1, 2, 3, 4$, is presented in Figure 5.1 for the first 900 observations as the training set and in Figure 5.2 for the remaining 900 observations as the test set (validation set). Note that these represent the "true" volatilities conditioned on the true log-volatilities $\mathbf{x}_t^T$ simulated from the model with the $x-$axis in the figures representing the time period and the $y-$axis representing the volatility level measured in percentages.

Taking the simulated observations $\{\mathbf{y}_t\}_{t=1}^{900}$ generated from the simulated log-volatilities series $\{\mathbf{x}_t\}_{t=0}^{900}$ in Figure 5.1, we proceed to use the q-MLE method to estimate the linearized MSV model (5.5) with unknown parameters

$$\begin{aligned} \tilde{\mathbf{y}}_t &= (-1.27)\mathbf{1}_p + \mathbf{x}_t + \tilde{\boldsymbol{\epsilon}}_t, & \tilde{\boldsymbol{\epsilon}}_t &\sim N(\mathbf{0}, \mathbf{R}) \\ \mathbf{x}_{t+1} &= \mathbf{x}_t + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim N(\mathbf{0}, \mathbf{Q}) \\ \mathbf{x}_0 &\sim N_p(\mathbf{0}, \mathbf{P}_0) \end{aligned} \tag{5.5}$$

where the vector $\tilde{\mathbf{y}}_t = [\tilde{y}_t^{(1)}, ..., \tilde{y}_t^{(4)}]'$ with $\tilde{y}_t^{(i)} = \log\left(y_t^{(i)}\right)^2$, for $i = 1, 2, 3, 4$. It is

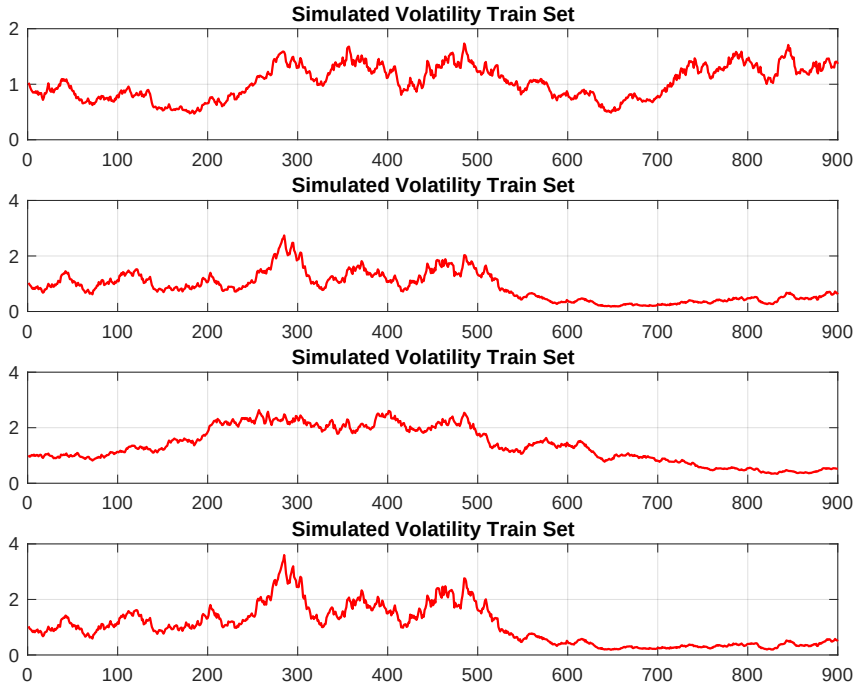

Figure 5.1: Simulated conditional $i^{th}-$volatility measured as $\exp\{\mathbf{x}_t^i/2\}$. These time series of volatilities serve as a placeholder for the "true" volatility observed during the training set.

important to note that the optimization scheme in the q-MLE algorithm is notorious for diverging from the optimal solution if the initial parameter estimates $\boldsymbol{\theta}_0$ is far from the optimal solution. Therefore, a new initial parameter estimation process for the MSV model is implemented to overcome this limitation. Taking $\mathbf{M} = \mathbf{I}_{4\times 4}$, the unknown parameters in the linearized MSV model (5.5) are the unknown coefficients in the matrices $\mathbf{R}$ and $\mathbf{Q}$ constrained to preserve the covariance structure. From equation (5.5), the expected value of the observation equation can be written as

$$\mathcal{E}\{\mathbf{x}_t\} = \tilde{\mathbf{y}}_t - (-1.27)\mathbf{1}_p = \hat{\mathbf{z}}_t \tag{5.6}$$

Thus, in order to provide initial parameter estimates $\boldsymbol{\theta}_0$ for the covariance matrices,

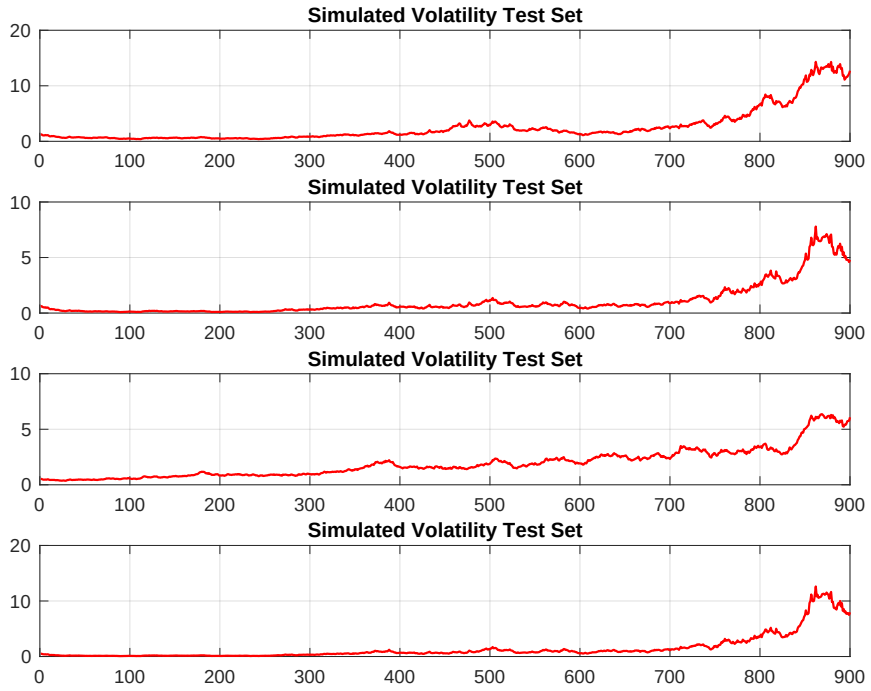

Figure 5.2: Simulated conditional $i^{th}-$volatility measured as $\exp\{\mathbf{x}_t^i/2\}$. These time series of volatilities serve as a placeholder for the "true" volatility observed during the test set.

we estimate the following vector AR(1) model using ordinary least squares (OLS)

$$\hat{\mathbf{z}}_t = \mathbf{M}^0 \hat{\mathbf{z}}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}^0) \tag{5.7}$$

This method will provide estimates for the unknown matrix $\mathbf{M}^0$ and the covariance matrix $\mathbf{Q}^0$. Notice that if the MSV model had $\boldsymbol{\mu} \neq 0$ and $\mathbf{M} \neq \mathbf{I}_{4\times 4}$, then this method may also provide an initial parameter estimate for $\mathbf{M}$ and $\boldsymbol{\mu}$ that can then be used for the q-MLE method for full estimation of the State-Space model. Proceeding to estimate the model (5.7) with the simulated time series observations $\mathbf{y}_t$ and unknown

diagonal matrix $\mathbf{M}^0$, the initial parameter estimates are

$$\mathbf{M}^0 = \begin{bmatrix} 0.9720 & 0 & 0 & 0 \\ 0 & 0.9924 & 0 & 0 \\ 0 & 0 & 0.9937 & 0 \\ 0 & 0 & 0 & 0.9939 \end{bmatrix} \tag{5.8}$$

$$\mathbf{Q}^0 = \begin{bmatrix} 0.0214 & 0.0082 & 0.0046 & 0.0068 \\ 0.0082 & 0.0206 & 0.0083 & 0.0115 \\ 0.0046 & 0.0083 & 0.0243 & 0.0074 \\ 0.0068 & 0.0115 & 0.0074 & 0.0244 \end{bmatrix} \tag{5.9}$$

It is also worth noting that in order to preserve stability in the estimates, equation (5.7) is fitted with the 30-day moving average of $\hat{\mathbf{z}}_t$ in lieu of $\hat{\mathbf{z}}_t$ itself. A consistent initial estimator for the matrix $\mathbf{R}$ is given by the correlation matrix of the 30-day moving average of $\hat{\mathbf{z}}_t$

$$\mathbf{R}^0 = \begin{bmatrix} 1.0000 & 0.3747 & 0.0808 & 0.3092 \\ 0.3747 & 1.0000 & 0.6881 & 0.9587 \\ 0.0808 & 0.6881 & 1.0000 & 0.8009 \\ 0.3092 & 0.9587 & 0.8009 & 1.0000 \end{bmatrix} \tag{5.10}$$

Now that the initial covariance parameters have all been estimated, we make the following adjustment to the model (5.5) before proceeding with the q-MLE procedure. The model used for the q-MLE procedure that incorporates the adjustment is

$$\begin{aligned} \tilde{\mathbf{y}}_t &= (-1.27)\mathbf{1}_p + \mathbf{x}_t + \mathbf{R}^{1/2}\tilde{\boldsymbol{\epsilon}}_t, \qquad \tilde{\boldsymbol{\epsilon}}_t \sim N(\mathbf{0}, \mathbf{I}) \\ \mathbf{x}_{t+1} &= \mathbf{x}_t + \mathbf{Q}^{1/2}\boldsymbol{\eta}_t, \qquad\qquad\quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{I}) \end{aligned} \tag{5.11}$$

where the covariance matrices are related by $\mathbf{R} = \mathbf{R}^{1/2}\mathbf{R}^{T/2}$ and $\mathbf{Q} = \mathbf{Q}^{1/2}\mathbf{Q}^{T/2}$. Since the decomposition of the matrices are not unique, the matrices $\mathbf{R}^{1/2}$ and $\mathbf{Q}^{1/2}$ are selected as the Cholesky decomposition of the covariance matrices. This means that the unknown parameters of the model (5.11) are the coefficients in the lower triangular matrices $\mathbf{R}^{1/2}$ and $\mathbf{Q}^{1/2}$ with their initial estimates taken as

$$(\mathbf{Q}^{1/2})^0 = \begin{bmatrix} 0.1461 & 0 & 0 & 0 \\ 0.0565 & 0.1319 & 0 & 0 \\ 0.0313 & 0.0493 & 0.1445 & 0 \\ 0.0466 & 0.0671 & 0.0181 & 0.1320 \end{bmatrix} \tag{5.12}$$

$$(\mathbf{R}^{1/2})^0 = \begin{bmatrix} 1.0000 & 0 & 0 & 0 \\ 0.3747 & 0.9272 & 0 & 0 \\ 0.0808 & 0.7095 & 0.7000 & 0 \\ 0.3092 & 0.9091 & 0.1870 & 0.2073 \end{bmatrix} \tag{5.13}$$

where the Cholesky decomposition of (5.9) and (5.10) was used. Therefore, the initial parameter estimate $\boldsymbol{\theta}_0$ that is used for the q-MLE procedure is taken as the vector stacked with the elements in the matrix $(\mathbf{R}^{1/2})^0$ and $(\mathbf{Q}^{1/2})^0$ as shown in Table 5.1. Recall that the MSV model considered for q-MLE estimation assumes $\mathbf{M} = \mathbf{I}_{4\times 4}$, so the matrix $\mathbf{M}^0$ is not included when stacking the vector $\boldsymbol{\theta}_0$.

Description	$\boldsymbol{\theta}_0$
Elements of $(\mathbf{R}^{1/2})^0$ in $\boldsymbol{\theta}_0$	0.1461
	0.0565
	0.0313
	0.0466
	0.1319
	0.0493
	0.0671
	0.1445
	0.0181
	0.1320
Elements of $(\mathbf{Q}^{1/2})^0$ in $\boldsymbol{\theta}_0$	1.0000
	0.3747
	0.0808
	0.3092
	0.9272
	0.7095
	0.9091
	0.7000
	0.1870
	0.2073

Table 5.1: Initial parameter estimate $\boldsymbol{\theta}_0$.

Given the initial parameter estimate $\boldsymbol{\theta}_0$, the q-MLE procedure is employed and the estimated results are given in Table 5.2 where the t-statistics and statistical significance based on the p-Values of the estimated parameters are also presented. From Table 5.2, it is observed that most parameters are statistically significant given their small p-Values with the exception of θ^3 and θ^{10} where their p-Values suggests that there is minimal statistical evidence that these two parameters should not be zero. Overall, the parameter estimate's t-statistics and p-values are within expectations.

Parameter	Estimate	Std Errr	t-Stat	Prob
θ^1	0.06873	0.01719	3.99822	0.00006
θ^2	0.10746	0.03667	2.93033	0.00339
θ^3	0.00429	0.02845	0.15073	0.88019
θ^4	0.09973	0.03896	2.56016	0.01046
θ^5	0.09483	0.02775	3.41787	0.00063
θ^6	0.04013	0.02859	1.40357	0.16045
θ^7	0.10269	0.02872	3.57515	0.00035
θ^8	0.07139	0.02019	3.53645	0.00041
θ^9	0.02455	0.01204	2.04005	0.04135
θ^{10}	0.00000	50.57849	0.00000	1.00000
θ^{11}	2.21535	0.04360	50.80757	0
θ^{12}	0.83326	0.07381	11.28991	0
θ^{13}	0.48292	0.07843	6.15721	0
θ^{14}	0.69605	0.08741	7.96341	0
θ^{15}	1.97658	0.04203	47.02932	0
θ^{16}	0.72530	0.08539	8.49399	0
θ^{17}	0.96241	0.08126	11.84411	0
θ^{18}	2.17536	0.02950	73.75218	0
θ^{19}	0.29103	0.06494	4.48137	0.00001
θ^{20}	1.98057	0.02657	74.53379	0

Table 5.2: Parameter Estimates for the MSV model (5.5) via q-MLE procedure

From Table 5.2, the q-MLE estimate of the state error covariance matrix $\mathbf{Q}$ is given by

$$\hat{\mathbf{Q}} = \hat{\mathbf{Q}}^{1/2}\hat{\mathbf{Q}}^{T/2} = \begin{bmatrix} 0.0047 & 0.0074 & 0.0003 & 0.0069 \\ 0.0074 & 0.0205 & 0.0043 & 0.0205 \\ 0.0003 & 0.0043 & 0.0067 & 0.0063 \\ 0.0069 & 0.0205 & 0.0063 & 0.0211 \end{bmatrix} \tag{5.14}$$

Where $\hat{\mathbf{Q}}^{1/2}$ is the estimated lower Cholesky-like covariance matrix from the q-MLE procedure and it is given by

$$\hat{\mathbf{Q}}^{1/2} = \begin{bmatrix} 0.0687 & 0 & 0 & 0 \\ 0.1075 & 0.0948 & 0 & 0 \\ 0.0043 & 0.0401 & 0.0714 & 0 \\ 0.0997 & 0.1027 & 0.0246 & 0.0000 \end{bmatrix} \tag{5.15}$$

Similarly, the estimate of the observation error covariance matrix $\mathbf{R}$ is given by

$$\hat{\mathbf{R}} = \hat{\mathbf{R}}^{1/2}\hat{\mathbf{R}}^{T/2} = \left(\frac{\pi^2}{2}\right)^{-1} \begin{bmatrix} 0.9945 & 0.3741 & 0.2168 & 0.3125 \\ 0.3741 & 0.9324 & 0.3721 & 0.5030 \\ 0.2168 & 0.3721 & 1.1128 & 0.3379 \\ 0.3125 & 0.5030 & 0.3379 & 1.0979 \end{bmatrix} \tag{5.16}$$

Where $\hat{\mathbf{R}}^{1/2}$ is the estimated lower Cholesky-like covariance matrix from the q-MLE procedure and it is given by

$$\hat{\mathbf{R}}^{1/2} = \begin{bmatrix} 2.2154 & 0 & 0 & 0 \\ 0.8333 & 1.9766 & 0 & 0 \\ 0.4829 & 0.7253 & 2.1754 & 0 \\ 0.6960 & 0.9624 & 0.2910 & 1.9806 \end{bmatrix} \tag{5.17}$$

It is noticed from the estimate $\hat{\mathbf{R}}$ that no restrictions on a correlation structure were imposed; that is, it is assumed only that the matrices have covariance structures but that $\mathbf{R}$ is not constraint to have diagonals as $\pi^2/2$. This assumption is shown to be appropriate as the estimated diagonal entries of $\hat{\mathbf{R}}$ are close to 1.

Now that the MSV model (5.5) has been fully estimated; that is, all matrices are known, the estimates of the log-volatility time series $\{\mathbf{x}_t\}$ can be obtained for each day in the training window via the Kalman filter. Figure 5.3 plots the conditional volatility time series derived from the Kalman filter estimates $\{\mathbf{x}_{t|t}\}$ defined as the diagonals of the matrix $\mathbf{V}_t^{1/2}$ in equation (5.1) and compares them to the "true" volatility. Recall that this "true" volatility is the volatility that was generated from the model and it is the same as the volatility in Figure 5.1. From Figure 5.3, it is noticed that the linearization of equation (5.5) is an appropriate model approximation to the MSV model (5.1) as the conditional volatility closely approximates the "true" volatility. These results are remarkable on their own since the linearization imposed a non-Gaussianity assumption in the model where the q-MLE method attempts to approximate the non-Gaussian distribution with a Normal distribution by appropriately specifying the model, in particular, $\mathbf{R}$. Thus, the Normal approximation sufficiently captures the "true" unobservable volatility.

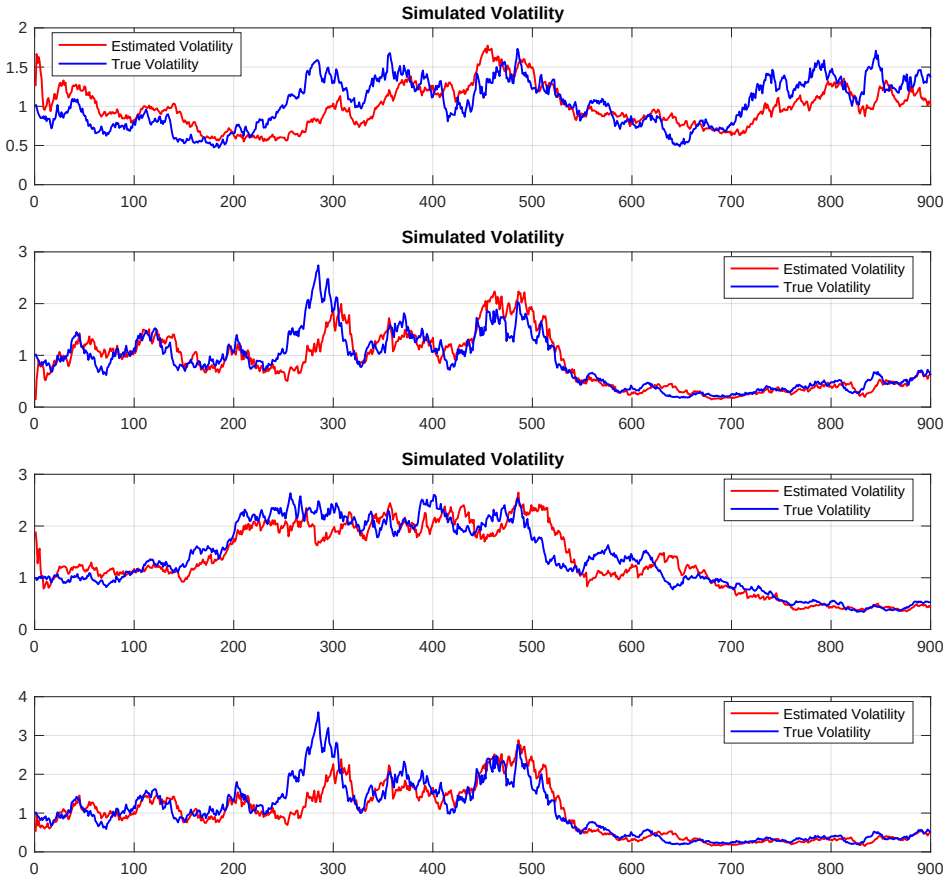

Figure 5.3: Estimated conditional $i^{th}-$volatility measured as $\exp\{x_t^i/2\}$, where $\mathbf{x}_t$ is the Kalman filter estimate.

5.1.1 Performance Diagnostics

Sensitivity analysis can be provided during the entire training set and test set for the estimated MSV model in terms of the short-range forecast error measure

$$e(\mathbf{x}_{t|t}) = \left(\mathbf{x}_{t+1|t} - \mathbf{x}_{t+1}^{v}\right)^{T} \left(\mathbf{x}_{t+1|t} - \mathbf{x}_{t+1}^{v}\right) \tag{5.18}$$

where $\mathbf{x}_{t+1|t}$ is the one-step ahead forecast produced from the Kalman filter estimate $\mathbf{x}_{t|t}$ and $\mathbf{x}_{t+1}^{v}$ is the verifying analysis at forecast date $t+1$. As mentioned before,

the verifying analysis is subjective and it provides the analyst with the freedom to input their own view of forecasted volatilities. For this numerical experiment, the verifying analysis is selected as the *fixed interval smoothed* log-volatilities obtained from the observations at the verification time (see Kalman smoother in Section 2.4). Using the forecast sensitivity equations derived for the MSV model in Section 4.6.1, impacts to the volatility forecast can be derived in terms of the estimated parameters and inputs.

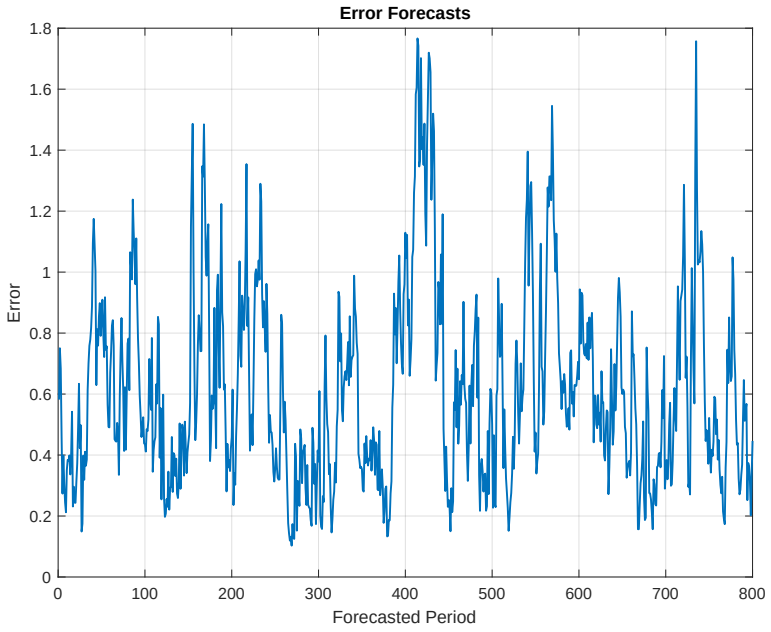

Figure 5.4: Standard Forecast Error Measure $\sqrt{e}$ as in equation (5.18)

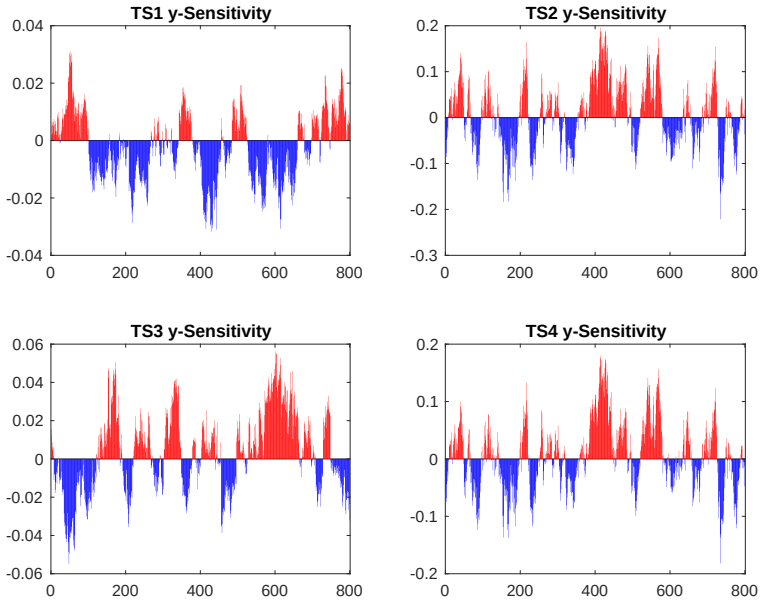

Figure 5.5: Forecast Error Sensitivity to Observations $\frac{\partial e}{\partial \tilde{\mathbf{y}}_t}$ for each time series (TS)

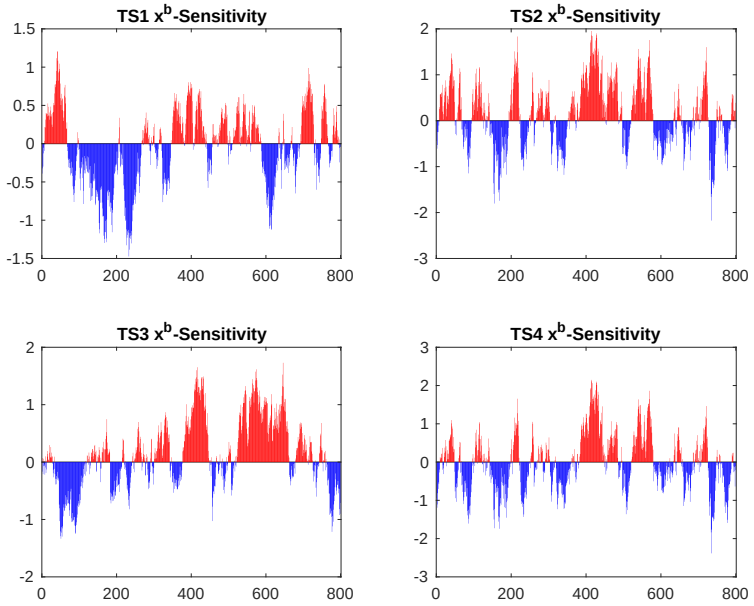

Figure 5.6: Forecast Error Sensitivity to background $\frac{\partial e}{\partial \mathbf{x}_{t|t-1}}$ for each time series (TS). Recall that the background vector is defined as the previous one-step ahead forecast.

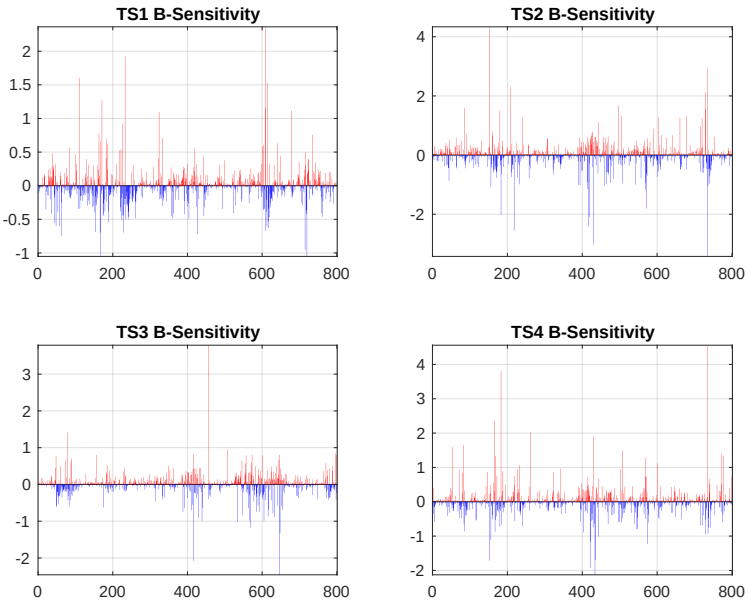

Figure 5.7: Forecast Error Sensitivity to Background Covariance Matrix for each time series (TS) taken as the diagonal entries of $\frac{\partial e}{\partial \mathbf{B}}$. Recall that the background covariance matrix is defined as the previous one-step ahead forecast.

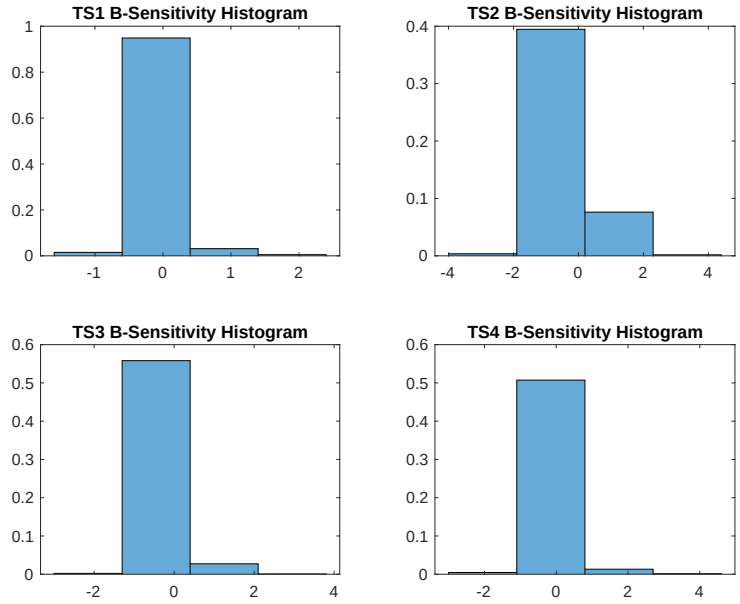

Figure 5.8: Histogram of the diagonal entries of $\frac{\partial e}{\partial \mathbf{B}}$ for each time series (TS).

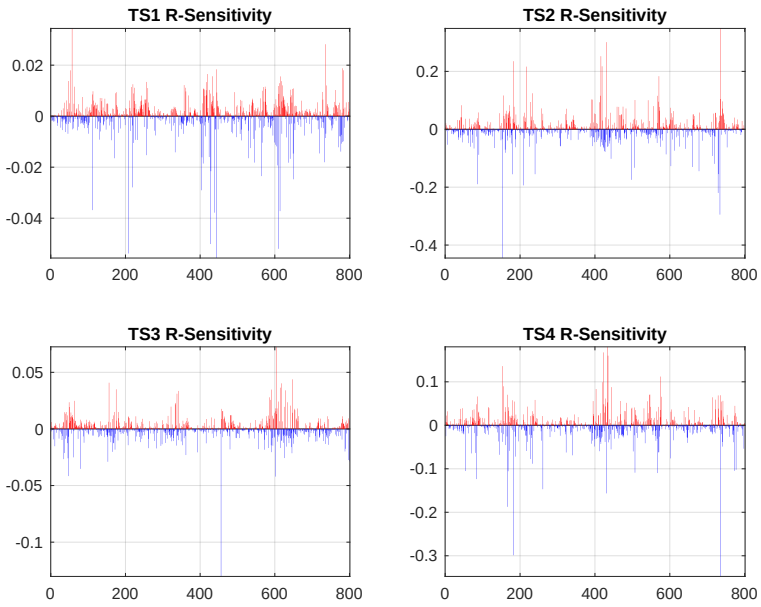

Figure 5.9: Forecast Error Sensitivity to Observations Error Covariance Matrix for each time series (TS) taken as the diagonal entries of $\frac{\partial e}{\partial \mathbf{R}}$

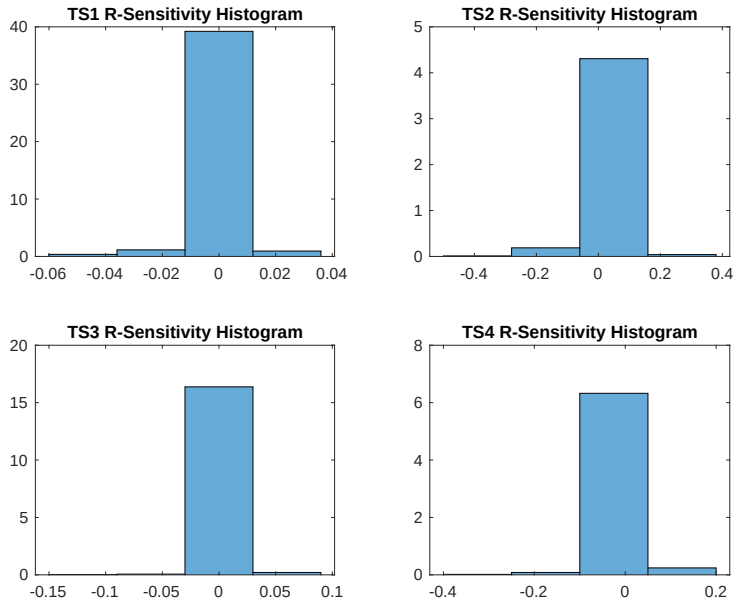

Figure 5.10: Histogram of the diagonal entries of $\frac{\partial e}{\partial \mathbf{R}}$ for each time series (TS).

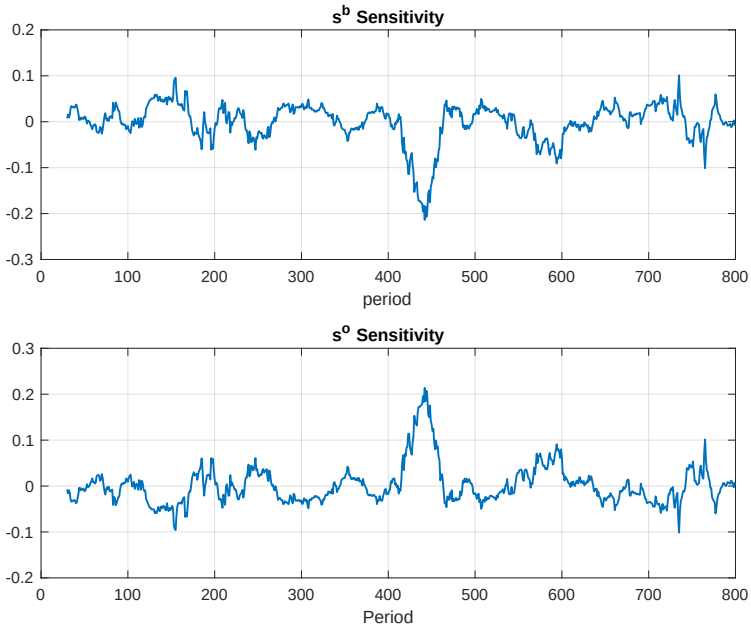

Figure 5.11: Forecast Error Sensitivity to Error Covariance Matrix inflation parameters $\frac{\partial e}{\partial s^o}$ and $\frac{\partial e}{\partial s^b}$ taken as their 30-day moving average.

Figure 5.4 presents the one-step ahead log-volatility forecast error score $\sqrt{e(\mathbf{x}_{t|t})}$ for the past 800 periods. Figure 5.5 presents the one-step ahead log-volatility forecast error sensitivity to the observations during each time iteration in the training set for the past 800 periods. This plot can be interpreted as a diagnostic tool to determine which financial variable is degrading the forecasted volatilities. Figure 5.6 and Figure 5.7 present the one-step ahead log-volatility forecast error sensitivity to the background estimate and to the background error covariance matrix, respectively. During each time iteration of the Kalman filter, these background estimates represent the previous one-step ahead log-volatility forecast and the previous one-step ahead forecast of the covariance matrix of the log-volatility estimates. For the Kalman filter, these background estimates during each time iteration will be dependent on the initial priors $\mathbf{x}_0 \sim N(\mathbf{0}, \mathbf{P}_0)$. Figure 5.9 represents the one-step ahead log-volatility forecast error sensitivity to the observation error covariance matrix. Figure 5.11 presents the one-step ahead log-volatility forecast error sensitivity to the matrix scaling factor s^0 for the observation error covariance and s^b to the state background error covariance matrix.

Each of the forecast sensitivities provide model performance diagnostics during each time iteration. It is observed from Figure 5.5 to Figure 5.11 that the largest impact on the one-step ahead log-volatility forecast is due to the background estimates. The initial background estimates, or the initial priors in statistical theory, are one of the main focuses on the study of Bayesian estimation as these background estimates can have large impacts on results. This is clearly seen in our simulated case as the sensitivities to $\mathbf{x}^b$ and $\mathbf{B}$ are capable of capturing the degradation in the forecasts by showing positive sensitivities. This is particularly true for the sensitivity to the background matrix $\mathbf{B}$ as the forecast degrades almost to 5 points during many cycles of the training set. The initial background covariance matrix was selected to have large

quantities on the diagonals ($1e^{10}$) so as to create a diffusion estimation problem. This selection is common in the econometric literature when little information is known about the priors, however, it may degrade model performance in the long run.

5.1.2 Adaptive Tuning of the Observation Error Covariance Matrix

The assumption with the q-MLE estimator is that the non-Gaussian distribution can be approximated by a Normal distribution with precision as good as the covariance matrix $\mathbf{R}$. Although the sensitivity diagnostics in Figure 5.9 seem to suggest that the q-MLE estimated covariance matrix $\mathbf{R}$ does not materially degrade forecasted log-volatilities, adaptive tuning of the matrix $\mathbf{R}$ is implemented to increase the accuracy of the forecasts such that the Normal approximation is more accurate.

Taking the search direction $\mathbf{p} = -\dfrac{\partial e}{\partial \mathbf{R}^{1/2}}$, an adaptive tuning step may be added during each assimilation iteration of the Kalman filter with

$$\mathbf{R}^{1/2}_{\text{NEW}} = \mathbf{R}^{1/2} + \alpha \mathbf{p} \tag{5.19}$$

where $\mathbf{R}^{1/2}$ is the lower Cholesky covariance decomposition of the matrix $\mathbf{R}$. If the log-volatility forecast error $e(\mathbf{x}_{t|t})$ is decreased under this new covariance matrix, the adaptive tuning step is accepted and a new analysis $\mathbf{x}_{t|t}$ is produced with a new forecast $\mathbf{x}_{t+1|t}$. The reason behind selecting the decomposition $\mathbf{R}^{1/2}$ as the lower Cholesky covariance is due to the fact that the q-MLE procedure was implemented using a Cholesky-like covariance decomposition. Thus, to be consistent with the q-MLE estimator, the adaptive procedure is implemented with $\mathbf{R}^{1/2}$ in lieu of $\mathbf{R}$.

Proceeding to include the adaptive step during each iteration of the Kalman filter using equation (5.19) and with the aid of Algorithm 4.1, updated volatility estimates are produced. Figure 5.12 presents the results of the Kalman filter volatility estimates

during the training set with the adaptive $\mathbf{R}^{1/2}$ procedure and compares these results to the "true" volatility. Since in this numerical experiment the "true" volatilities are available, Figure 5.13 presents the volatility error measured as the difference between the "true" volatility and Kalman filter with adaptive $\mathbf{R}^{1/2}$ volatility estimates. The Kalman filter with no adaptive tuning volatility errors are also presented in Figure 5.13 for comparison and Table 5.3 presents error statistics of each estimation error. Figure 5.14 presents the dynamic step-size implemented during each iteration of the Kalman filter with adaptive $\mathbf{R}^{1/2}$ model and it indicates the step-size taken during each iteration. It is noted that the adaptive procedure accepted 898 adaptive steps out of 900.

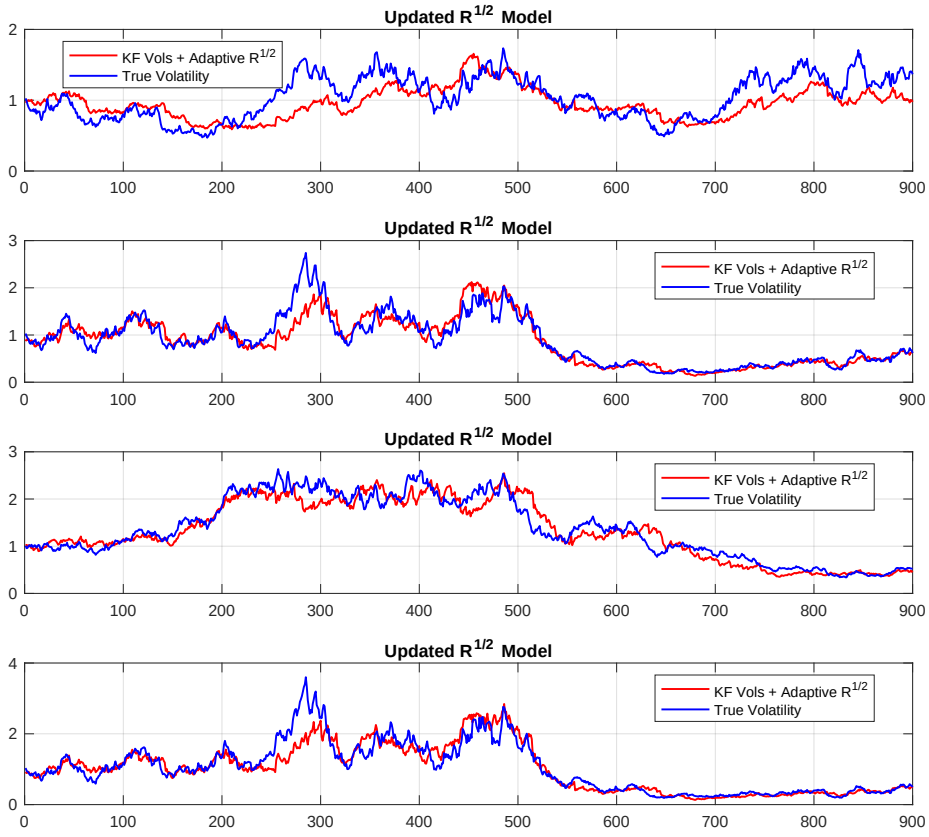

Figure 5.12: Kalman filter volatility estimation with adaptive $\mathbf{R}^{1/2}$ model during the training set.

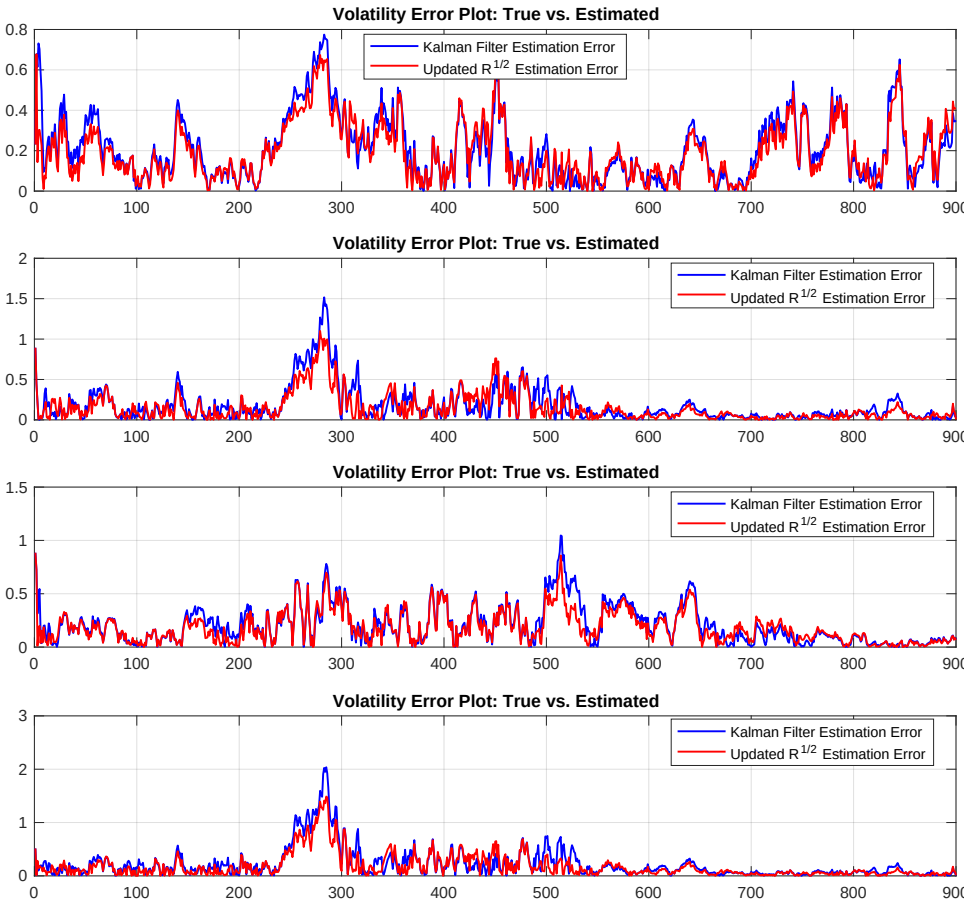

Figure 5.13: Volatility estimation error during training set measured as the difference between true volatility and estimated volatility. Kalman filter estimation error represents the difference between the true volatility and the Kalman filter volatility estimates. Updated $\mathbf{R}^{1/2}$ estimation error represents the difference between the true volatility and the $\mathbf{R}^{1/2}$ adaptive tuning step with Kalman filter volatility estimation.

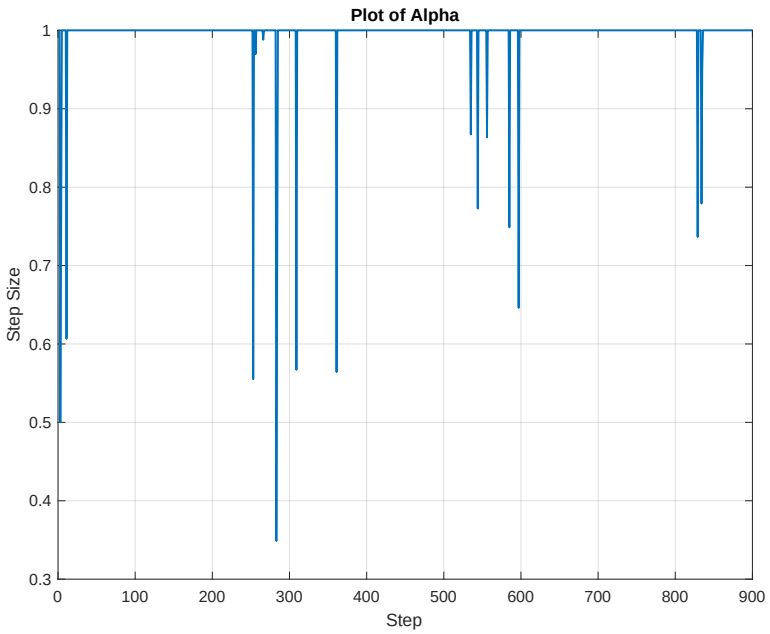

Figure 5.14: Step-size α during each iteration of the Kalman filter in the training set.

	MAE KF	MAE KF with Adaptivity	RSME KF	RSM KF with Adaptivity
TS1	0.20974	0.19117	0.26296	0.23758
TS2	0.17607	0.14747	0.27912	0.22837
TS3	0.20537	0.18218	0.2715	0.23454
TS4	0.21675	0.17613	0.35606	0.285

Table 5.3: Error statistics based on the volatility error during the training set.

The adaptive procedure for tuning the observation error covariance matrix $\mathbf{R}^{1/2}$ produces improved conditional volatility estimates over the plain Kalman filter (no adaptivity), as demonstrated by the improved volatility estimation errors in Figure 5.13 and the reduced error statistics of Table 5.3. For example, it is noticed from Figure 5.13 for the 4^{th} financial variable that the volatility estimate around the 280-300 time period is significantly improved from a volatility error of about 2% to a volatility error of about 1.5 %. Furthermore, it is worth mentioning that the adaptive procedure in Algorithm 4.1 is designed to only improve the volatility estimates; that is, including adaptive tuning of the covariance matrix $\mathbf{R}^{1/2}$ is no worse than the original Kalman filter volatility estimates, as seen in Figure 5.13.

It is a well known fact from econometric theory that models degrade over time due to the fact that model parameters are estimated during a training set and forecasts are employed during a test set. It is expected that using models over a long period of time, outside of the training set, will degrade in performance since the estimated parameters may become stale or irrelevant to new incoming data. A common solution is to re-develop the model to more relevant data. On the other hand, adaptivity of model parameters may be implemented during the test set to provide improved estimates without necessitating to re-estimate the model again.

Figure 5.15 presents the conditional volatility estimates during the test set using the Kalman filter with adaptive $\mathbf{R}^{1/2}$ matrix and compares them to the true volatility time series. Figure 5.16 presents the volatility estimation error during the test set measured as the difference between the true volatility and the Kalman filter with adaptive $\mathbf{R}^{1/2}$ procedure. The plot also shows the volatility estimation error between the true volatility and the plain Kalman filter (no adaptivity) for comparison, while Table 5.4 presents error statistics under each methodology. Figure 5.17 presents the dynamic plot of the step-size α during each iteration in the test set of the Kalman filter.

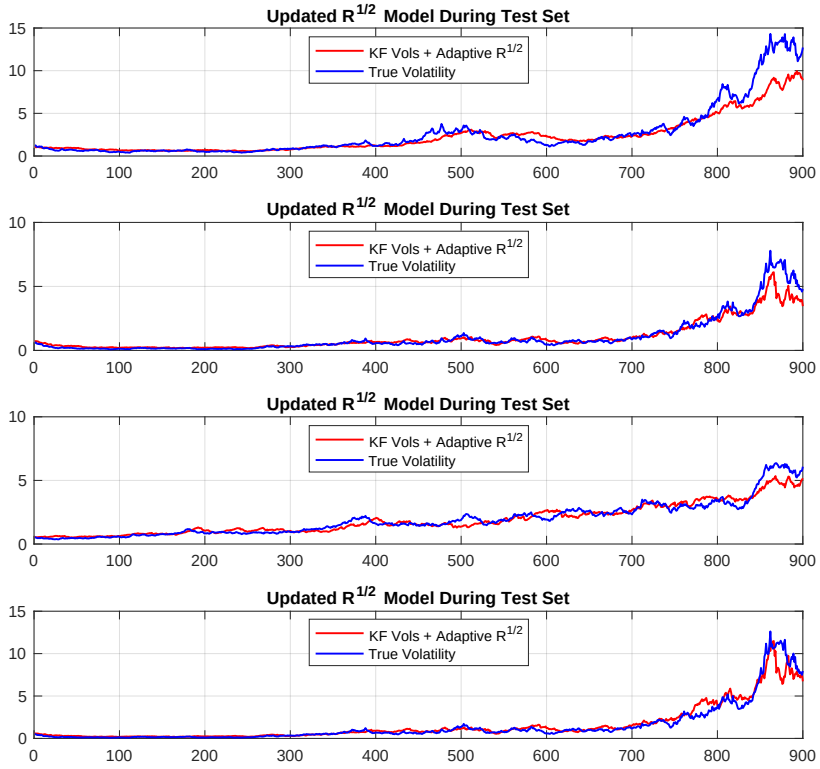

Figure 5.15: Kalman filter volatility estimation with adaptive $\mathbf{R}^{1/2}$ model during the test set.

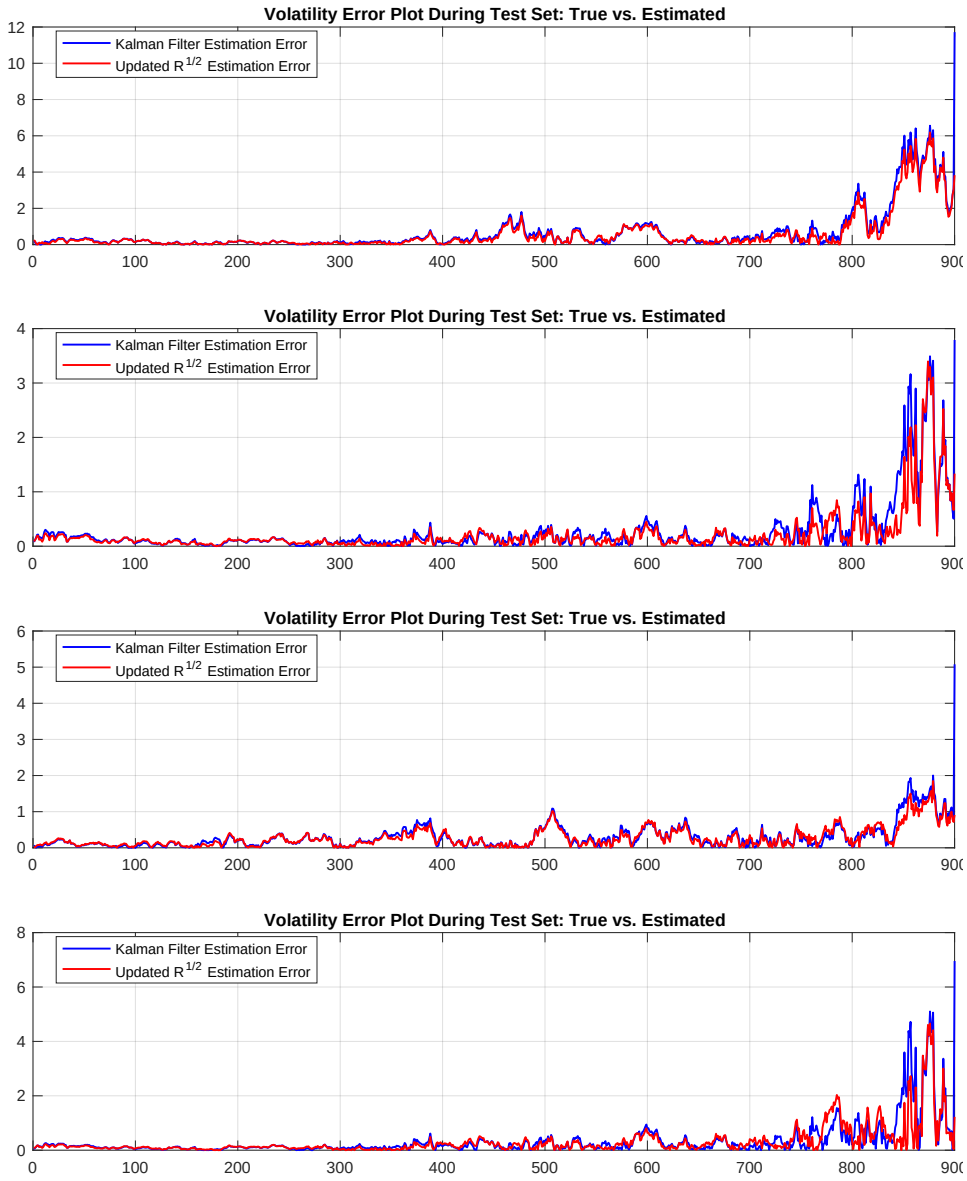

Figure 5.16: Volatility estimation error during test set measured as the difference between true volatility and estimated volatility. Kalman filter estimation error represents the difference between the true volatility and the Kalman filter volatility estimates. Updated $\mathbf{R}^{1/2}$ estimation error represents the difference between the true volatility and the $\mathbf{R}^{1/2}$ adaptive tuning step with Kalman filter volatility estimation.

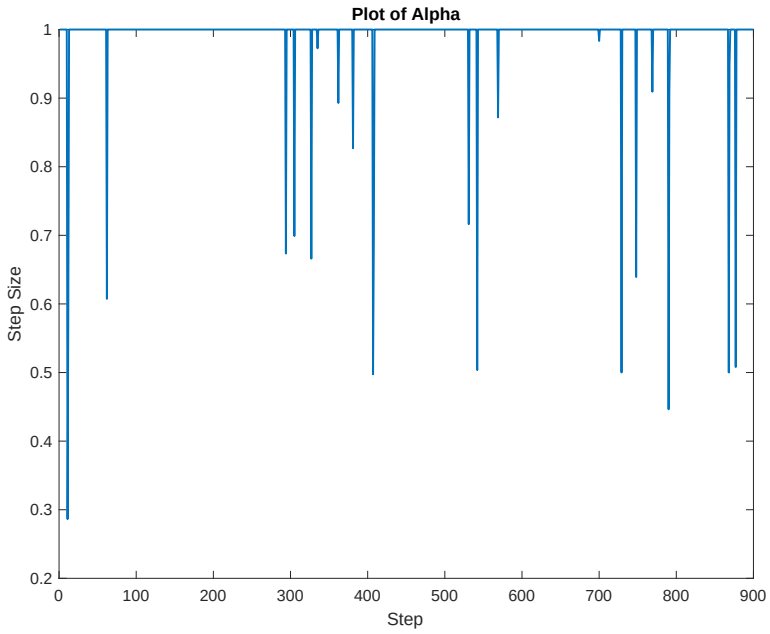

Figure 5.17: Step-size α during each iteration of the Kalman filter in the test set.

	MAE KF	MAE KF with Adaptivity	RSME KF	RSM KF with Adaptivity
TS1	0.48643	0.42818	1.0584	0.88234
TS2	0.21087	0.1916	0.46672	0.38638
TS3	0.27423	0.24373	0.40103	0.34848
TS4	0.33179	0.30456	0.83181	0.69963

Table 5.4: Error statistics based on the volatility error during the test set.

From Figure 5.16 it is observed that, overall, the adaptive $\mathbf{R}^{1/2}$ procedure continues to provide improved volatility estimates during the test set. Although there are a few instances were the updated volatility provides slightly worse results than the plain Kalman filter, this concern is mitigated from the fact the adaptivity provides improved results over all, specially seen after the 800 time period were volatility jumped up. Table 5.4 also demonstrates that the adaptivity procedure is capable

of reducing volatility error statistics. It is also worth mentioning that, although the updated volatility provided slightly worse results during a few cycles, it is due to the fact that the update is performed in terms of the proxy volatility $\mathbf{x}_t^v$ (a run of the Kalman smoother during the test set without adaptivity). Thus, it is the case that the forecast error ϵ was reduced in terms of this proxy, which depends on the linearized observations $\tilde{\mathbf{y}}_t$ (already an approximation introducing errors), suggesting that careful considerations of the proxy should be implemented. In practice, the "true" volatility cannot be observed, however, the proxy implemented in this dissertation provides sufficient evidence that it is an appropriate selection. Since the "true" volatilities are available in this experiment, adaptivity of the background estimates are performed in the following section in terms of the true volatility as proxy to demonstrate the predictive power of the adaptive procedure when the "correct" volatilities are implemented.

5.1.3 Adaptive Tuning of Background Error Covariance Matrix

To demonstrate the importance of the proxy selection and the predictive power of including an adaptive procedure within each iteration of the Kalman filter, adaptivity will be performed in terms of true volatility as the proxy. That is, using the volatility time series generated for the problem, the volatility forecast score function is defined as

$$e(\mathbf{x}_{t|t}) = \left(\mathbf{x}_{t+1|t} - \mathbf{x}_{t+1}^T\right)^T \left(\mathbf{x}_{t+1|t} - \mathbf{x}_{t+1}^T\right)^T \tag{5.20}$$

where $\mathbf{x}_{t+1|t}$ is the Kalman filter one-step ahead forecast log-volatility from the analysis at time t and $\mathbf{x}_{t+1}^T$ is the true log-volatility at time $t+1$. Using Algorithm 4.1 with the search direction $\mathbf{p}$ taken as $0.5(\hat{\mathbf{p}} + \hat{\mathbf{p}}^T)$, where $\hat{\mathbf{p}} = -\dfrac{\partial e}{\partial \mathbf{B}}$, an adaptive tuning

step is added during each assimilation iteration of the Kalman filter with

$$\mathbf{B}_{\text{NEW}} = \mathbf{B} + \alpha \mathbf{p} \tag{5.21}$$

where $\mathbf{B}$ is the background error covariance matrix obtained during each iteration of the Kalman filter. Notice that the search direction was selected to assure that Positive Definiteness is preserved during each iteration. If the log-volatility forecast score $e(\mathbf{x}_{t|t})$ is reduced under this new background covariance matrix $\mathbf{B}$, the adaptive step is accepted and updated log-volatility and forecasted log-volatilities are produced. Figure 5.18 presents the results of the Kalman filter with adaptive tuning of $\mathbf{B}$ volatility estimates measured as the diagonals of $\mathbf{V}_t^{1/2}$ and compares them to the true volatility time series. The results shown in Figure 5.18 are remarkable and demonstrate that adding the adaptive tuning step with the correct volatility proxy (in this case the true volatility) approximates the true volatility with high fidelity. Figure 5.19 presents the volatility estimation error measured as the difference between the true volatility and the estimated volatility. The Kalman filter estimation error represents the difference between the true volatility and the Kalman filter volatility estimates while the KF + Adaptive $\mathbf{B}$ estimation error represents the difference between the true volatility and the adaptive $\mathbf{B}$ tuning step with Kalman filter volatility estimation.

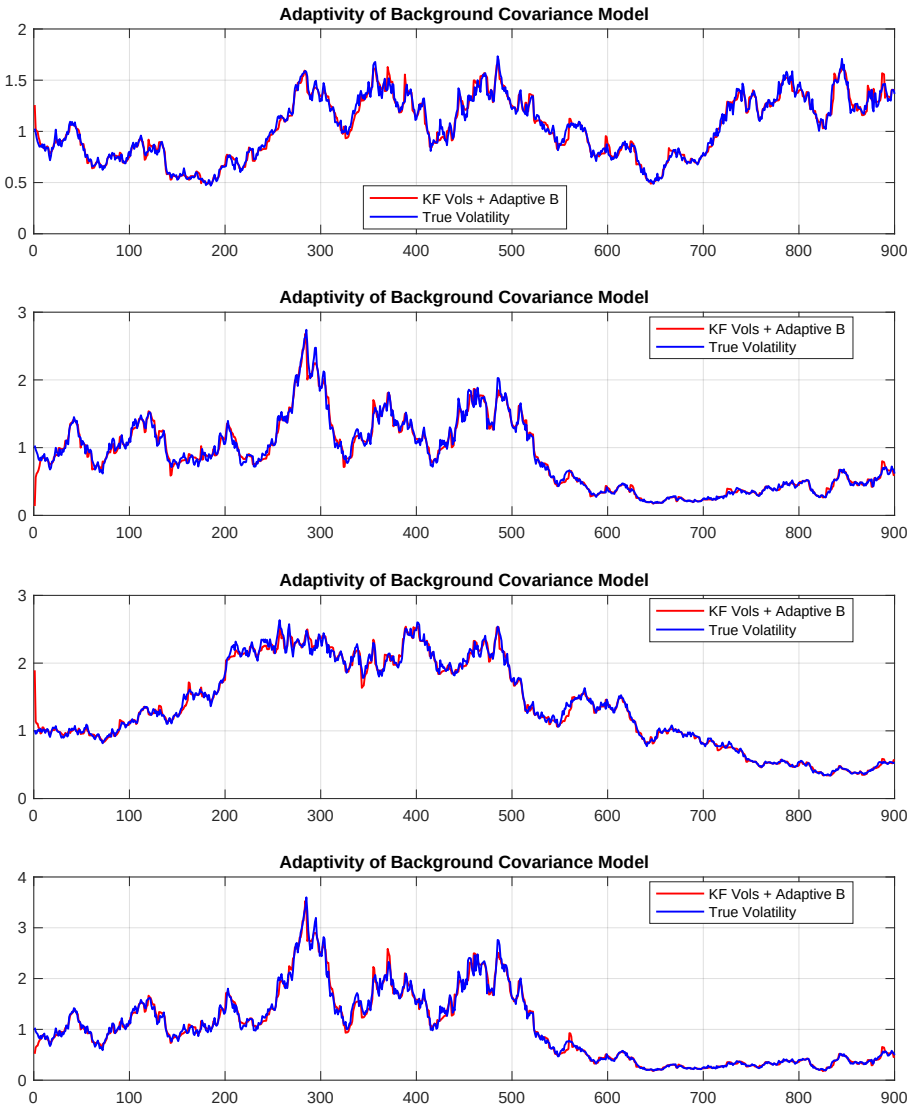

Figure 5.18: Kalman filter volatility estimation with adaptive **B** model during the training set.

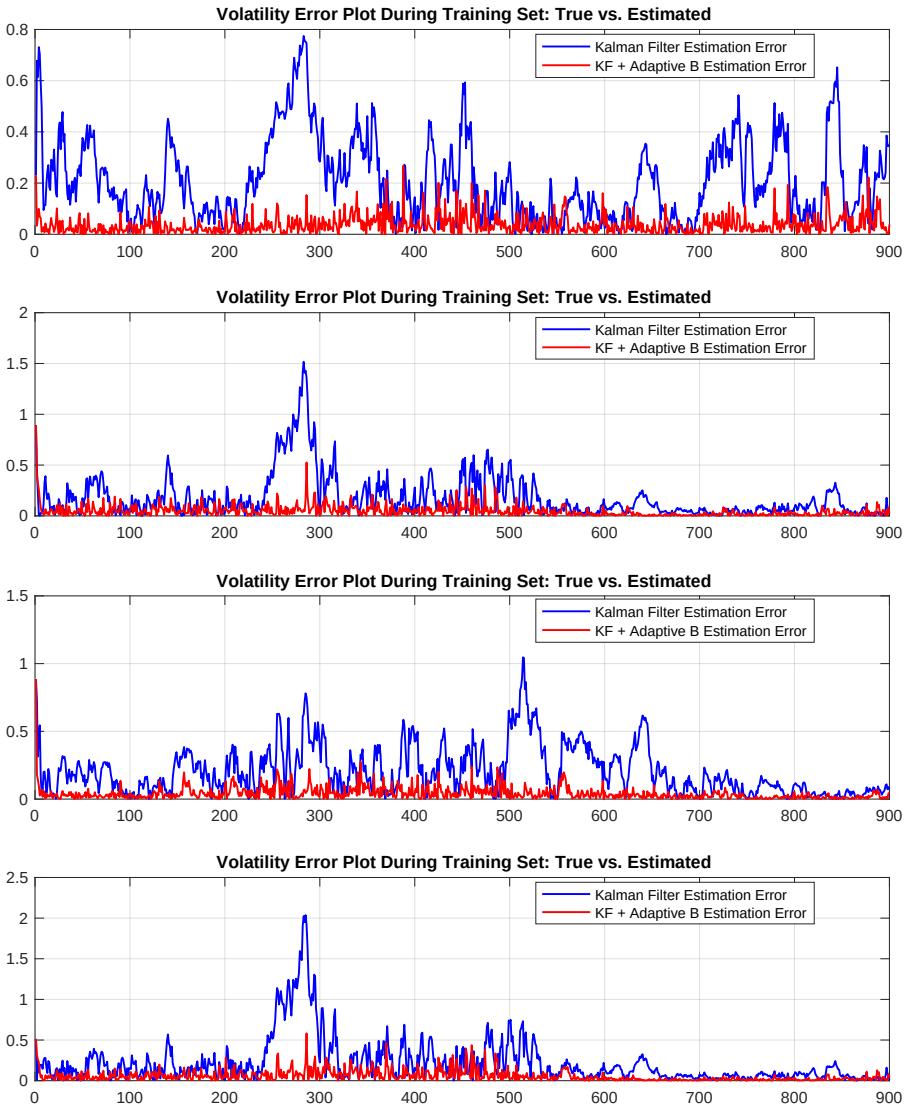

Figure 5.19: Volatility estimation error during training set measured as the difference between true volatility and estimated volatility. Kalman filter estimation error represents the difference between the true volatility and the Kalman filter volatility estimates. KF + Adaptive **B** estimation error represents the difference between the true volatility and the adaptive **B** tuning step with Kalman filter volatility estimation.

As shown in Figure 5.19, the volatility estimation error quickly goes towards zero within the first iteration of the Kalman filter with adaptivity. These improved volatility estimates provide evidence of the proof-of-concept behind the adaptivity procedure. Furthermore, as shown during the diagnostic analysis of Figure 5.6 and Figure 5.7, the priors $\mathbf{x}^b$ and $\mathbf{B}$ within each iteration of the Kalman filter materially impacted the log-volatility forecasts. Thus, providing the derivative information of the background error covariance matrix $\mathbf{B}$ within each iteration to update the analysis (and priors) provides the most impact to the estimates and it is in line with the diagnostic expectations.

The Kalman filter with adaptive tuning of $\mathbf{B}$ is also implemented during the test set to demonstrate the robustness of the adaptive procedure. Figure 5.20 provides the Kalman filter volatility estimation with adaptive $\mathbf{B}$ model during the test set and Figure 5.21 provides the volatility estimation error during the test set measured as the difference between the true volatility and the estimated volatility. These analyses further demonstrates the importance of selecting the correct proxy and the predictive power of adding an adaptive procedure to the input that material impacts the forecasted log-volatilities.

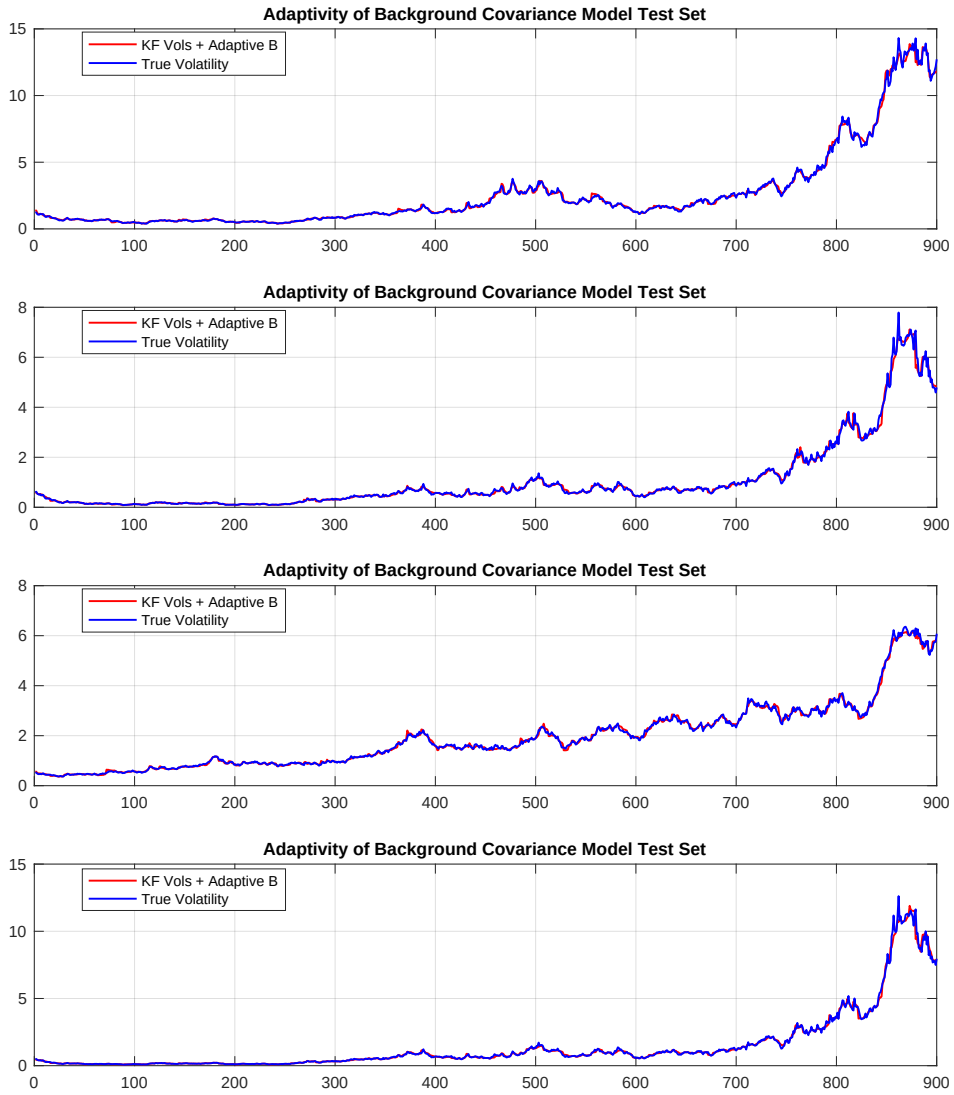

Figure 5.20: Kalman filter volatility estimation with adaptive **B** model during the test set.

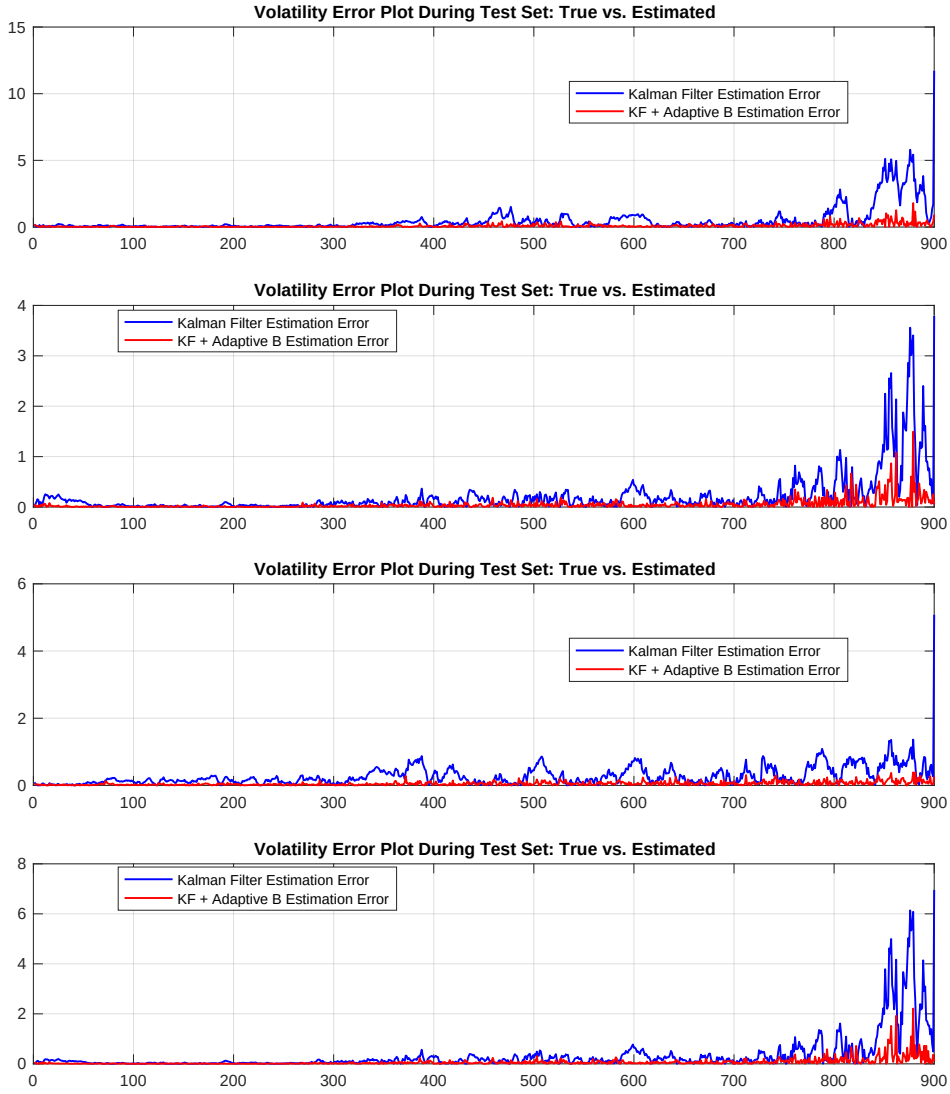

Figure 5.21: Volatility estimation error during test set measured as the difference between true volatility and estimated volatility. Kalman filter estimation error represents the difference between the true volatility and the Kalman filter volatility estimates. KF + Adaptive **B** estimation error represents the difference between the true volatility and the adaptive **B** tuning step with Kalman filter volatility estimation.

5.2 Foreign Exchange MSV Estimation

For the second set of numerical experiments, historical data on seven Foreign Exchange (FX) rates are used for analysis. The FX rate time series consists of observations obtained directly from and published by the Federal Reserve Bank of St. Louis Research Economic Data (https://www.stlouisfed.org). The FX rates selected for modeling include the Australian dollar (AUD), Canadian dollar (CAD), Chinese Yuan (CNY), European Euro (EUR), British Pound Sterling (GBP), Japanese Yen (JPY) and the Mexican Peso (MXN) all paired to the U.S. dollar. Historical observations of weekdays data from these time series are obtained for the period 10/20/2014 to 10/18/2019, for a total of 1,305 observations. The dataset is split into a training set and a test set where the training set contains observations from 10/20/2014 up to 01/01/2018, while the test set contains observations from 01/01/2018 to 10/18/2019.

5.2.1 Data Pre-Processing

It is observed from the dataset that certain dates have missing values. In order to overcome this data limitation, a univariate Kalman filter is used to predict the missing time series observation given all available data up to that date. In particular, the Kalman filter implemented for the data imputation is given by the procedure in Algorithm 5.1. It is further noted that the Algorithm 5.1 is only applicable for univariate time series in that it is necessary that there exist a time dependency in the data.

Figure 5.22 presents each FX time series data after applying the Kalman filter imputation procedure of Algorithm 5.1 to 56 missing observations. Note that by construction of the algorithm, the observations remain unchanged and only the missing entries are imputed.

Algorithm 5.1: Missing Observation Imputation

Input: Univariate Time Series Observations $\{y_t\}_{t=1}^N$ with missing values and initial estimates $\hat{y}_0$ and P_0.

Output: Univariate Time Series Observations $\{y_t\}_{t=1}^N$ with missing values predicted.

procedure MISSING VALUE IMPUTATION($\{y_t\}_{t=1}^N$)
  for t = 1,2,...N **do**
    if y_t = NaN **then** ▷ observation value is missing
      $v_t = 0$ and $K_t = 0$
      $\hat{y}_t = \hat{y}_{t-1} + K_t v_t$
      $P_t = P_{t-1}(1 - K_t) + 1$
      $y_t = \hat{y}_t$ ▷ Impute missing observation as KF analysis
    else ▷ No need for imputation
      $v_t = y_t - \hat{y}_t$
      $K_t = P_t^{-1}(P_t + 1)$
      $\hat{y}_t = \hat{y}_{t-1} + K_t v_t$ ▷ Continue with the KF
      $P_t = P_{t-1}(1 - K_t) + 1$
      $y_t = y_t$ ▷ Observation remains the same
    end if
  end for
end procedure

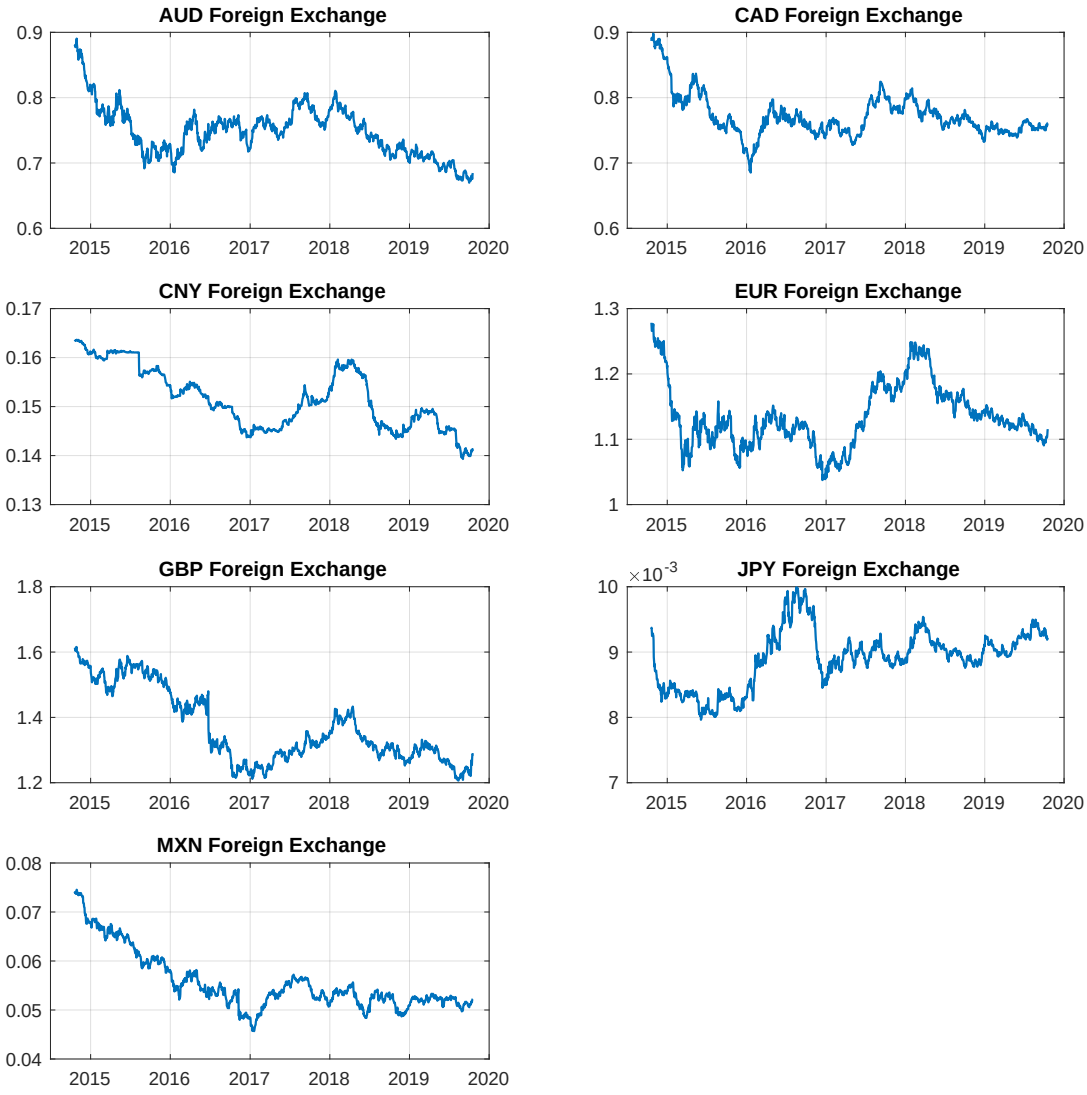

Figure 5.22: Foreign exchange time series data after Kalman filter imputation.

Next, for each FX rate time series, their corresponding returns are calculated as the log-returns. Recall that the log-returns of a financial time series is defined as the yield of the financial variable at time t and it is calculated as

$$\text{Log-Return}_t = \log \text{FX}_t - \log \text{FX}_{t-1} \tag{5.22}$$

Each of the FX rate returns are plotted in Figure 5.23 as the absolute value of equation (5.22). From these returns, the observations $\mathbf{y}_t$ that will be used for modeling are defined as

$$y_t^{(i)} = \text{Log-Return}_t^{(i)} - \mathcal{E}\left(\text{Log-Return}_t^{(i)}\right) \tag{5.23}$$

where, in this sense, the statistical operator $\mathcal{E}\{*\}$ is taken as the empirical average of the log-returns and the index i = AUD, CAD, CNY, EUR, GBP, JPY, MXN. The reason behind taking the empirical average in equation (5.23) is to ensure that there are no $y_t^{(i)}$'s identically equal to zero; this could create difficulties when taking the logarithms of $y_t^{(i)}$. Since the MSV model uses the transformation $\log\left(y_t^{(i)}\right)^2$ for each FX time series, it is of importance to be able to detect the presence of a unit-root in the FX data. Recall that if the FX data possess a unit-root, then the time lag of the FX series is required for modeling such that the observation equation includes one more difference in the data. In general, these tests take the form

$$\log\left(y_t^{(i)}\right)^2 = \phi_0 + \phi_1 \log\left(y_{t-1}^{(i)}\right)^2 + \text{Noise} \tag{5.24}$$

Common statistical tests to detect the presence of a unit-root ($\phi = 1$) in the data include the Augmented Dickey-Fuller (ADF) test, the Kwiatkowski-Philips-Schmidt-Shin (KPSS) test and the Philips-Perron (PP) tests. As a method of pre-processing the FX data, Table 5.5 presents each of the unit-root test conclusions at the 95% confidence level along with their p-Values for each FX time series under each statistical test. Notice from the table that all of the unit-root tests reject the null hypothesis that the FX rate data contains a unit-root. Therefore, the conclusion is that there is minimal statistical evidence that the transformed time series $\log\left(y_t^{(i)}\right)^2$ requires to be differenced further.

FX	ADF Test	KPSS Test	PP Test
AUD	0.0162	0.0100	0.0036
Unit-Root	No	No	No
CAD	0.0161	0.0100	0.0046
Unit-Root	No	No	No
CNY	0.0421	0.0100	0.0142
Unit-Root	No	No	No
EUR	0.0195	0.0100	0.0038
Unit-Root	No	No	No
GBP	0.0127	0.0325	0.0010
Unit-Root	No	No	No
JPY	0.0122	0.1000	0.0018
Unit-Root	No	Yes	No
MXN	0.0132	0.0293	0.0021
Unit-Root	No	No	No

Table 5.5: Unit-Root Test results at the 95% confidence level for $\log\left(y_t^2\right)$. The test results "No" indicate the rejection of the null hypothesis that there is a unit-root in the data.

5.2.2 FX MSV Model

Using the transformed FX data $\mathbf{y}_t \in \mathbb{R}^7$, the following Multivariate Stochastic Volatility model is considered for estimation

$$\begin{aligned} \mathbf{y}_t &= \mathbf{V}_t^{1/2}\boldsymbol{\epsilon}_t \\ \mathbf{x}_{t+1} &= \mathbf{x}_t + \boldsymbol{\eta}_t, \\ \mathbf{x}_0 &\sim N_p(\mathbf{0}, \mathbf{P}_0) \end{aligned} \tag{5.25}$$

Model (5.25) is selected with $\mathbf{M} = \mathbf{I}$ for the purpose of allowing the log-volatilities to follow a random walk; that is, it is expected that the volatility will be stochastic for some period but, overall, it should remain fairly constant. Linearization of this model is required for estimation and the resulting linear State-Space model is given by

$$\begin{aligned} \tilde{\mathbf{y}}_t &= (-1.27)\mathbf{1}_p + \mathbf{x}_t + \tilde{\boldsymbol{\epsilon}}_t, & \tilde{\boldsymbol{\epsilon}}_t &\sim N(\mathbf{0}, \mathbf{R}) \\ \mathbf{x}_{t+1} &= \mathbf{x}_t + \boldsymbol{\eta}_t, & \boldsymbol{\eta}_t &\sim N(\mathbf{0}, \mathbf{Q}) \\ \mathbf{x}_0 &\sim N_p(\mathbf{0}, \mathbf{P}_0) \end{aligned} \tag{5.26}$$

where $\tilde{y}_t^{(i)} = \log\left(y_t^{(i)}\right)^2$ and the index $i =$ AUD, CAD, CNY, EUR, GBP, JPY, MXN denoting each FX time series. In order to produce Maximum Likelihood estimates of the unknown covariance parameters $\{\mathbf{R}, \mathbf{Q}\}$, initial covariance parameter estimates are produced in a similar fashion as described in the Section 5.1. The resulting initial covariance parameters $\mathbf{R}^0$ and $\mathbf{Q}^0$ are presented below where an initial estimate of $\mathbf{M}$ is also given; however, ruling in favor of the identity matrix as described by equation (5.26). It is also worth mentioning that in the model, $\boldsymbol{\mu}$ was set to zero; that is, the model does not assume a long-term mean of log-volatilities. This assumption appears to be appropriate since the initial estimate of $\mathbf{M}$ is approximately an identity matrix and if $\boldsymbol{\mu} \neq \mathbf{0}$, then $\boldsymbol{\mu} + \mathbf{M}(\mathbf{x}_t - \boldsymbol{\mu})$ implies that $\boldsymbol{\mu} = \mathbf{0}$.

$$\mathbf{M}^0 = \begin{bmatrix} 0.9997 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.9998 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9998 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9998 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.9997 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.9999 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.9996 \end{bmatrix} \tag{5.27}$$

$$\mathbf{Q}^0 = \begin{bmatrix} 0.0224 & 0.0070 & 0.0014 & 0.0031 & 0.0026 & 0.0033 & 0.0051 \\ 0.0070 & 0.0223 & 0.0027 & 0.0023 & 0.0036 & 0.0034 & 0.0059 \\ 0.0014 & 0.0027 & 0.0296 & 0.0036 & 0.0007 & 0.0031 & 0.0016 \\ 0.0031 & 0.0023 & 0.0036 & 0.0228 & 0.0046 & 0.0053 & 0.0026 \\ 0.0026 & 0.0036 & 0.0007 & 0.0046 & 0.0260 & 0.0030 & 0.0028 \\ 0.0033 & 0.0034 & 0.0031 & 0.0053 & 0.0030 & 0.0280 & 0.0032 \\ 0.0051 & 0.0059 & 0.0016 & 0.0026 & 0.0028 & 0.0032 & 0.0251 \end{bmatrix} \tag{5.28}$$

$$\mathbf{R}^0 = \begin{bmatrix} 1.0000 & 0.5112 & -0.3481 & 0.5156 & 0.4731 & 0.0104 & 0.3796 \\ 0.5112 & 1.0000 & -0.1091 & 0.1762 & 0.5466 & 0.2252 & 0.5624 \\ -0.3481 & -0.1091 & 1.0000 & -0.3693 & 0.2113 & 0.0444 & 0.1682 \\ 0.5156 & 0.1762 & -0.3693 & 1.0000 & 0.0903 & -0.1427 & -0.0251 \\ 0.4731 & 0.5466 & 0.2113 & 0.0903 & 1.0000 & 0.1167 & 0.5311 \\ 0.0104 & 0.2252 & 0.0444 & -0.1427 & 0.1167 & 1.0000 & 0.2060 \\ 0.3796 & 0.5624 & 0.1682 & -0.0251 & 0.5311 & 0.2060 & 1.0000 \end{bmatrix} \tag{5.29}$$

Then, the Cholesky decomposition of the initial covariance parameters $\mathbf{R}^0$ and $\mathbf{Q}^0$ are produced and their elements are all stacked into a single vector $\boldsymbol{\theta}^0 \in \mathbb{R}^{56}$, as described in Section 5.1. This initial parameter vector $\boldsymbol{\theta}^0$ is then fed into the q-MLE procedure of Algorithm 3.1 to produce Maximum Likelihood Estimates (MLE) $\hat{\boldsymbol{\theta}}$ from which MLEs $\hat{\mathbf{R}}$ and $\hat{\mathbf{Q}}$ of the observation and state error covariance matrices are further obtained. Recall that the q-MLE procedure is possible only if the linear State-Space model (5.26) assumes a Gaussian distribution of the observation and state innovations.

The optimization results from the q-MLE procedure are provided below, where the standard error and p-Value of each estimate is omitted. However, we briefly summarize that most parameter estimates are statistically significant (p-Values < 0.1) with the exception of a few parameter estimates that are not statistically significant; that is, for these parameters, there is minimal statistical evidence that they should not be zero. Not having statistical significance of parameter estimates is of minimal concern within this context (as opposed to a regression problem) since these estimates represent the covariance explained between FX rates and not the predictive power explained between FX rates.

$$\hat{\mathbf{Q}} = \begin{bmatrix} 0.0040 & 0.0018 & 0.0050 & 0.0033 & 0.0051 & -0.0022 & 0.0040 \\ 0.0018 & 0.0025 & -0.0023 & -0.0006 & 0.0024 & 0.0008 & 0.0015 \\ 0.0050 & -0.0023 & 0.0894 & 0.0141 & 0.0252 & -0.0152 & 0.0213 \\ 0.0033 & -0.0006 & 0.0141 & 0.0060 & 0.0051 & -0.0045 & 0.0043 \\ 0.0051 & 0.0024 & 0.0252 & 0.0051 & 0.0118 & -0.0046 & 0.0095 \\ -0.0022 & 0.0008 & -0.0152 & -0.0045 & -0.0046 & 0.0044 & -0.0036 \\ 0.0040 & 0.0015 & 0.0213 & 0.0043 & 0.0095 & -0.0036 & 0.0079 \end{bmatrix} \tag{5.30}$$

$$\hat{\mathbf{R}} = \left(\frac{\pi^2}{2}\right) \begin{bmatrix} 1.0464 & 0.3216 & 0.0569 & 0.1431 & 0.1054 & 0.1669 & 0.2162 \\ 0.3216 & 1.0329 & 0.1292 & 0.1099 & 0.1645 & 0.1622 & 0.2736 \\ 0.0569 & 0.1292 & 1.2425 & 0.1494 & 0.0065 & 0.1623 & 0.0461 \\ 0.1431 & 0.1099 & 0.1494 & 1.0586 & 0.2010 & 0.2560 & 0.1100 \\ 0.1054 & 0.1645 & 0.0065 & 0.2010 & 1.2095 & 0.1465 & 0.1123 \\ 0.1669 & 0.1622 & 0.1623 & 0.2560 & 0.1465 & 1.2929 & 0.1500 \\ 0.2162 & 0.2736 & 0.0461 & 0.1100 & 0.1123 & 0.1500 & 1.1760 \end{bmatrix} \tag{5.31}$$

Now that the State-Space model (5.26) has been fully estimated (all parameters are known), the Kalman filter is employed to produce log-volatility estimates of each FX time series for each time period. Figure 5.23 presents the conditional volatility estimates, measured as the diagonals entries of $\mathbf{V}_t^{1/2}$, of each FX time series obtained from the Kalman filter estimates. These conditional volatilities are plotted along with the absolute value of the returns (log-returns) of each FX time series to provide insight on the full behavior of the FX series. For example, the GBP FX time series produces consistent returns over time; that is, for most periods, the returns have been relatively stable with minimal sudden jumps above the historicals. This observation is also in line with GBP conditional volatility estimates as the volatility produced by the Kalman filter shows relatively dormant volatility periods over time. On the other hand, the CNY FX time series shows periods in time where the returns have had sudden increases and decreases. This is also consistent with the conditional volatility estimates produced by the Kalman filter as the volatility series produces frequent increases and decreases in volatility estimates. Furthermore, it is noted that this observation is consistent with financial theory as it is well known that the returns on assets are mostly driven by sudden changes in volatilities; that is,

investors require higher returns on their investments as risks increases. Thus, the Multivariate Stochastic Volatility model makes it an attractive modeling choice for modeling multivariate conditional volatilities.

Although the volatility estimates are producing outputs with intuition consistent with financial theory, it is noted that the academic literature provides limited knowledge of a consistent methodology to diagnose model outputs and model performance of the Multivariate Stochastic Volatility model. This dissertation, therefore, extends this limitation by providing new diagnostic tools for model performance and the results are presented in the next section.

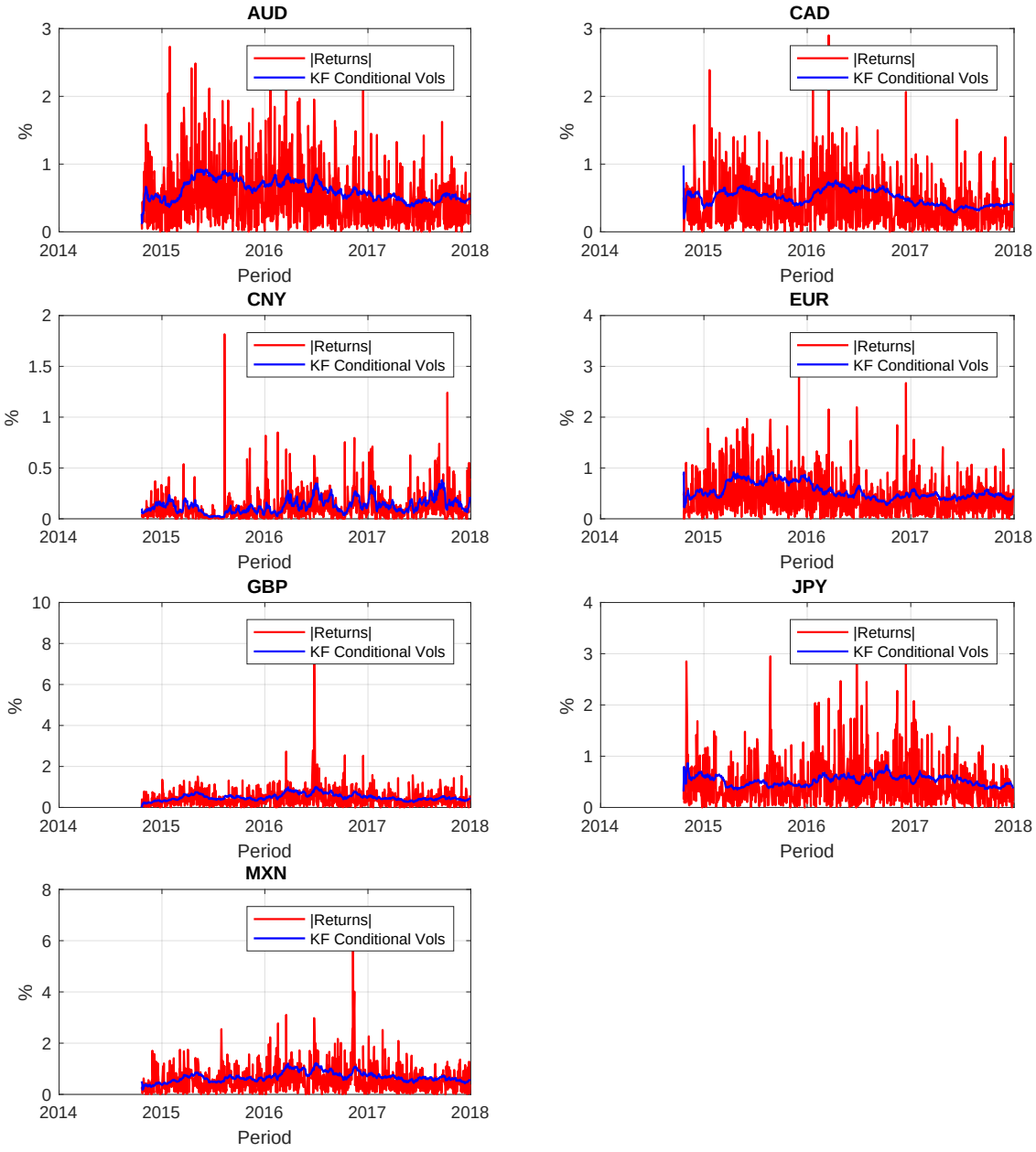

Figure 5.23: Foreign Exchange rate conditional volatilities. The absolute value of the log-returns of each FX rate time series is plotted along with their conditional volatilities obtained from the Kalman filter estimates.

5.2.3 Sensitivity Diagnostics

The MSV model provided conditional volatility estimates for each FX time series during each time period in the training set. Sensitivity analysis can be developed to provide diagnostic tools to assess the MSV's model performance during the training set. Once the MSV model (5.26) has been fully estimated, the Kalman filter may be implemented to provide the log-volatility state sequence $\{\mathbf{x}_{t|t}\}$ during each time period of the training set. It was shown that these log-volatilities estimates are now dependent on the input parameters from the model; that is, for each time period t,

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t}\left(\tilde{\mathbf{y}}_t, \mathbf{x}_{t|t-1}, \mathbf{P}_{t|t-1}, \mathbf{R}, \mathbf{Q}\right) \tag{5.32}$$

where $\mathbf{x}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$ are the background estimates at time t; i.e., the current Kalman filter forecast, $\mathbf{R}$ and $\mathbf{Q}$ are the observation and state error covariance matrix, respectively. The forecast score function is then defined as

$$e(\mathbf{x}_{t|t}) = \left(\mathbf{x}_{t+1|t} - \mathbf{x}^v_{t+1}\right)^T \left(\mathbf{x}_{t+1|t} - \mathbf{x}^v_{t+1}\right)^T \tag{5.33}$$

where $\mathbf{x}_{t+1|t}$ is the one-step ahead Kalman filter prediction of log-volatilities and $\mathbf{x}^v_{t+1}$ is the verifying analysis at time $t+1$ and serves as a proxy for the "true" log-volatilities. Since sensitivity analysis will be provided during each time iteration of the training set, this section of the dissertation implements the Kalman filter at time $t+1$, from observations $\tilde{\mathbf{y}}_{t+1}$, as the log-volatility proxy to the "true" states. Using the results derived in Table 4.2, the sensitivity analysis to input parameters are obtained.

Figure 5.24 presents the forecast score function $\sqrt{e}$ during each time iteration in the training set and represents the *all at once* short-range log-volatility forecast error impact. Figure 5.25 presents the error forecast sensitivity to observations $\tilde{\mathbf{y}}_t$

during each time iteration of the training set and represents the log-volatility forecast impact to daily changes of FX rates. Figure 5.26 and Figure 5.27 presents the log-volatility forecast error sensitivity to background error estimates and background error covariance estimates during each time iteration of the training set and represents the log-volatility forecast error impact to prior log-volatility estimates of each FX rate. Figure 5.29 presents the log-volatility forecast error sensitivity to the transformed observation error covariance matrix and represents the log-volatility forecast error impact to the accuracy of the linearized observation model. The plot of the distribution of the background and observation error covariance estimates is also present in Figure 5.28 and in Figure 5.30, respectively, to help identify the symmetry of the distribution of forecast degradations (positive values).

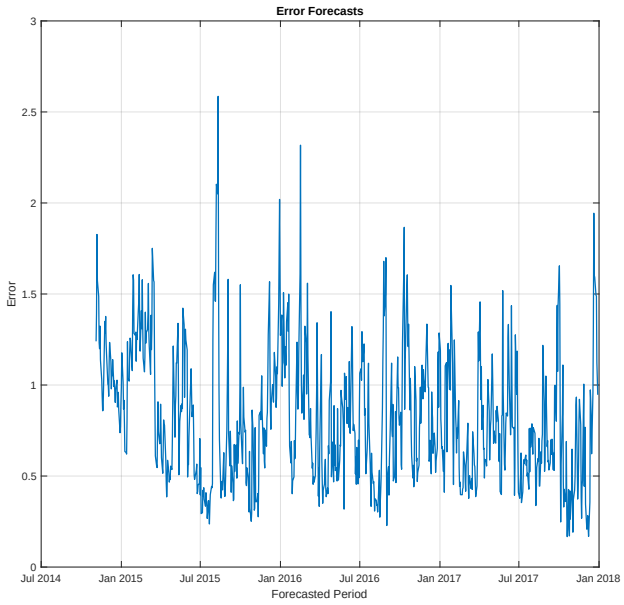

Figure 5.24: Forecast score function $\sqrt{e}$ during each time iteration in the training set represented as the *all at once* short-range log-volatility forecast error impact.

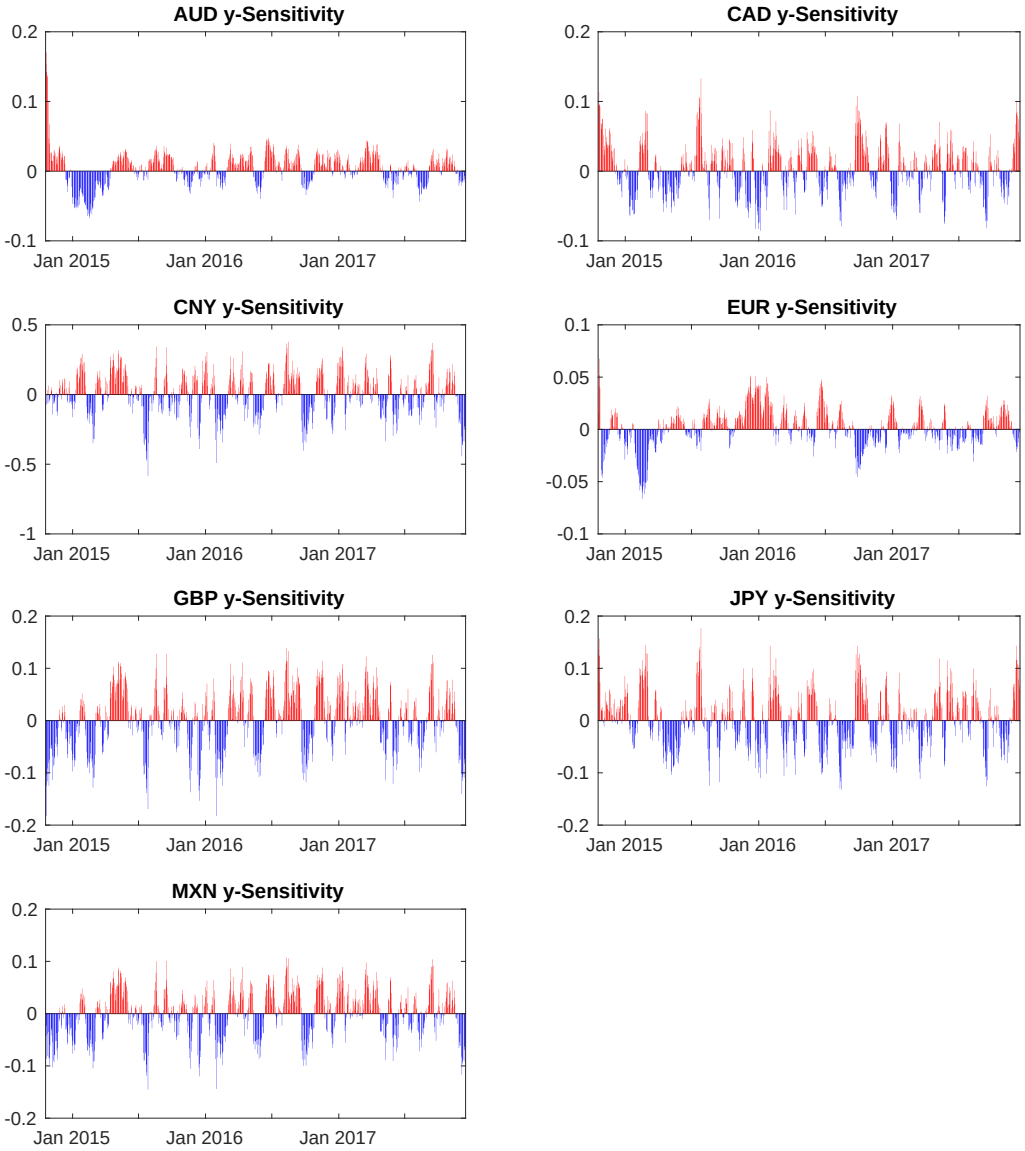

Figure 5.25: Log-volatility forecast error sensitivity to daily FX observations. These represent the log-volatility forecast error impact to daily changes of FX rates.

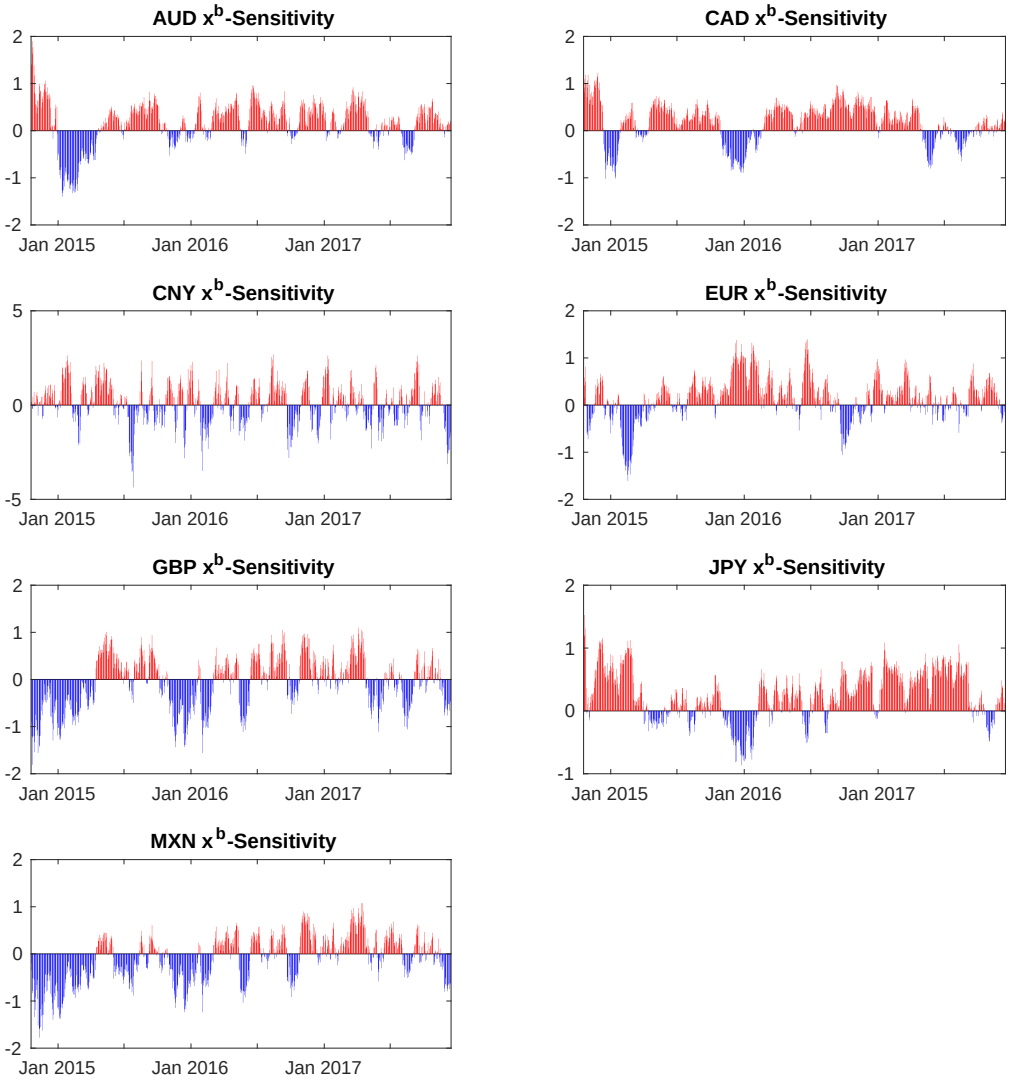

Figure 5.26: Log-volatility forecast error sensitivity to background estimates. These represent the log-volatility forecast error impact to prior log-volatility estimates of each FX rate.

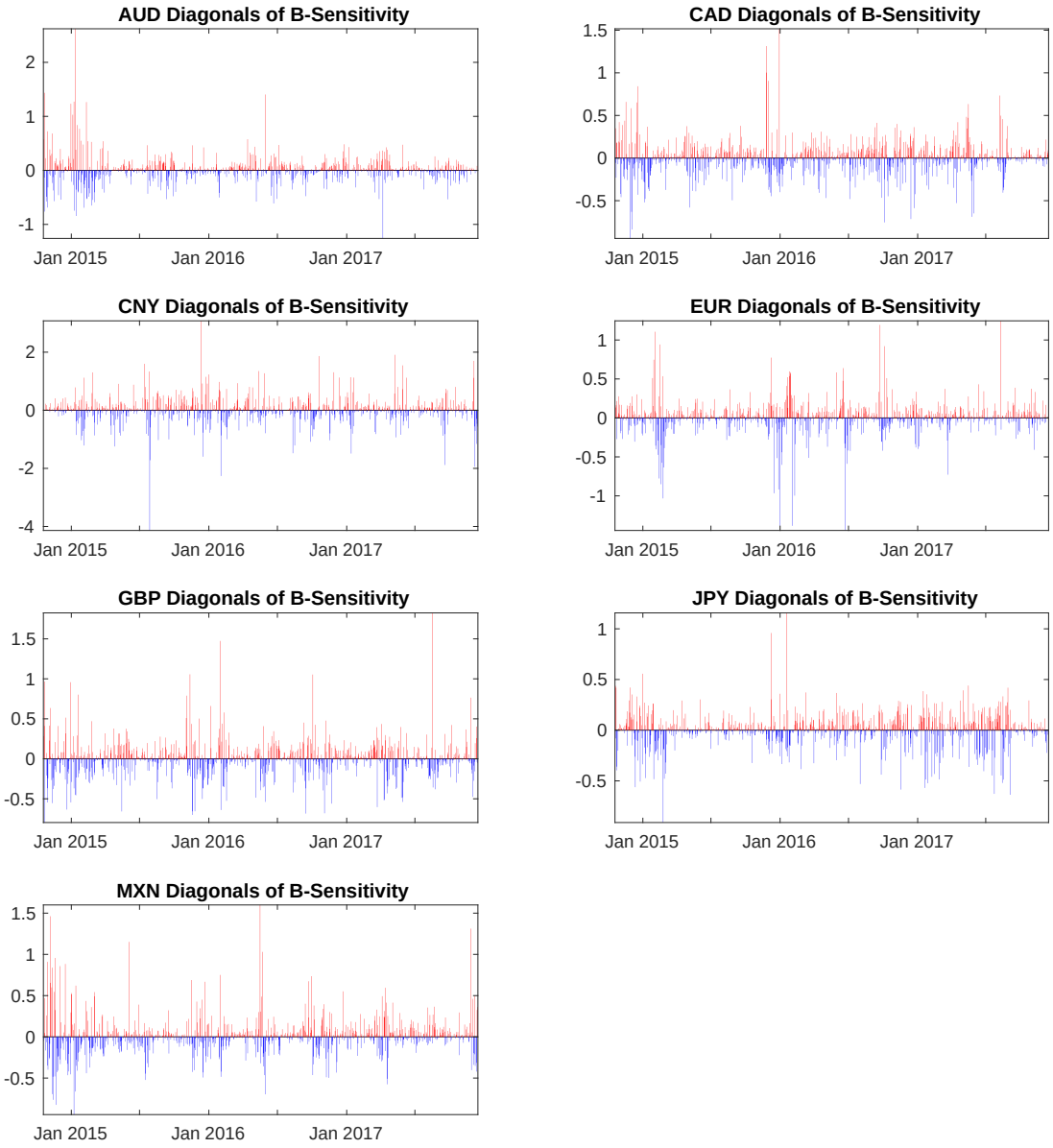

Figure 5.27: Log-volatility forecast error sensitivity to background covariance estimates. These represent the log-volatility forecast error impact to prior log-volatility covariance estimates taken as the diagonal entries of the covariance matrix.

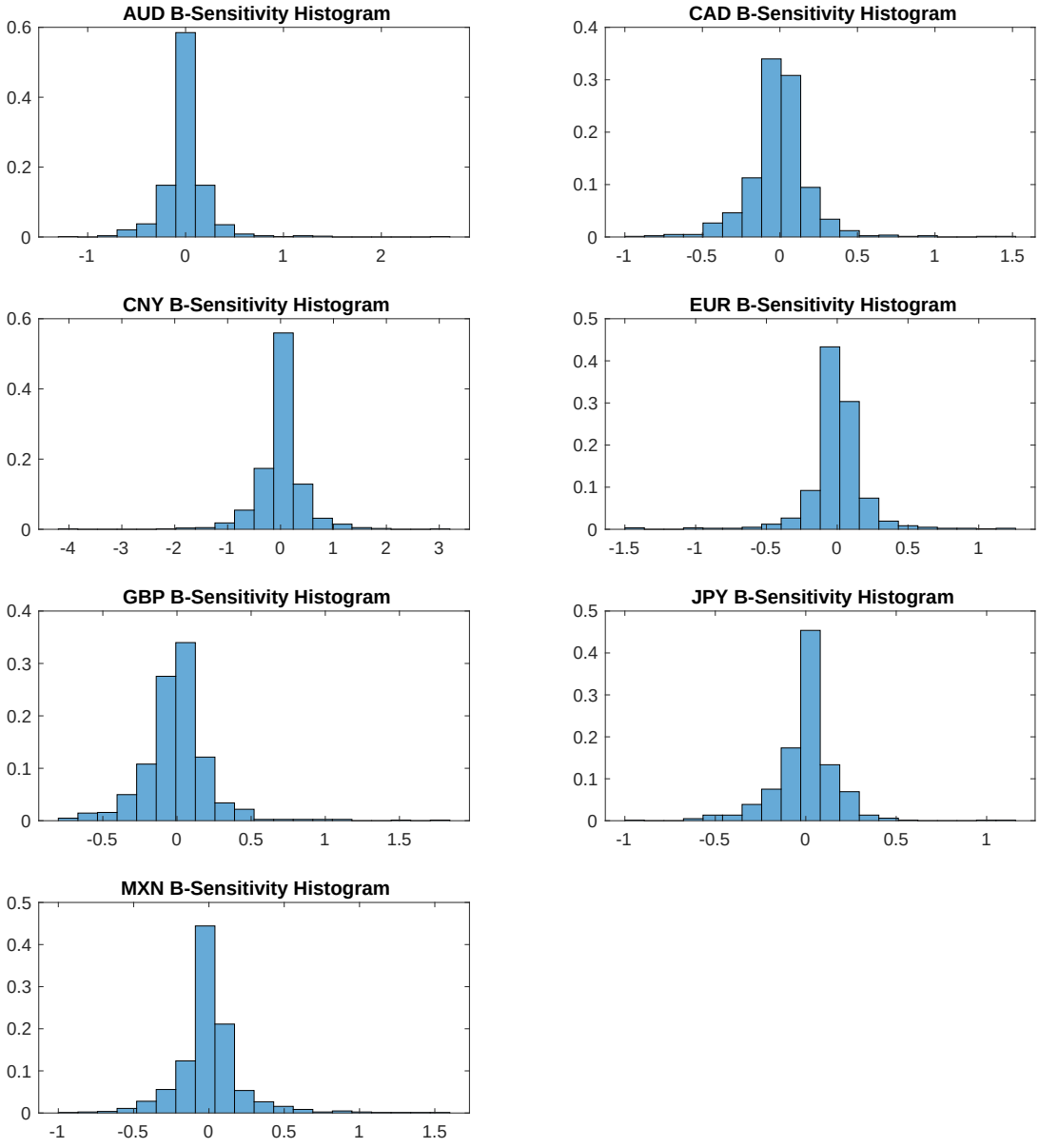

Figure 5.28: Distribution of the log-volatility forecast error sensitivity to background covariance estimates.

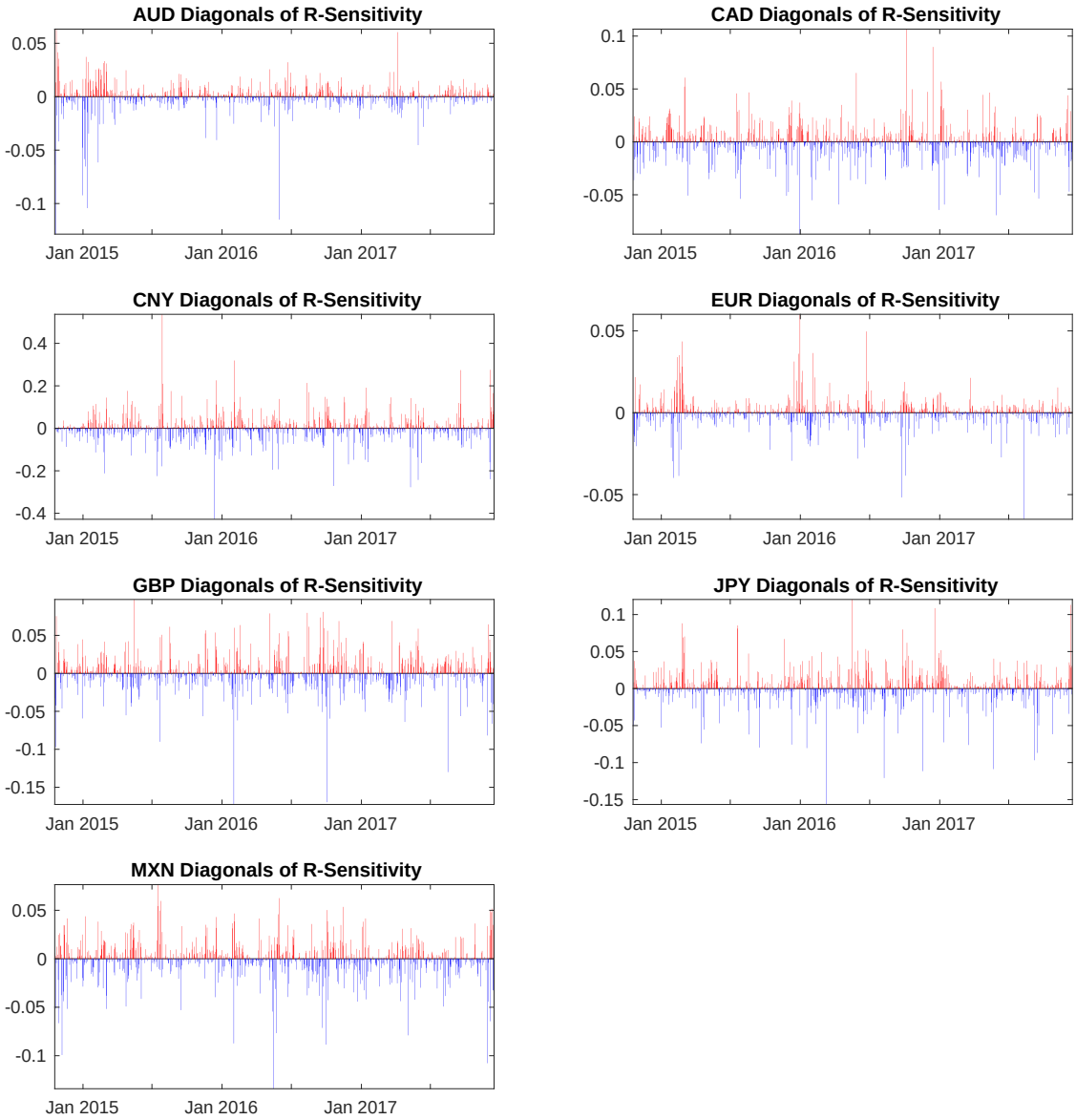

Figure 5.29: Log-volatility error sensitivity to the transformed observation error covariance matrix. These represent the log-volatility forecast error impact to the accuracy of the linearized observation model.

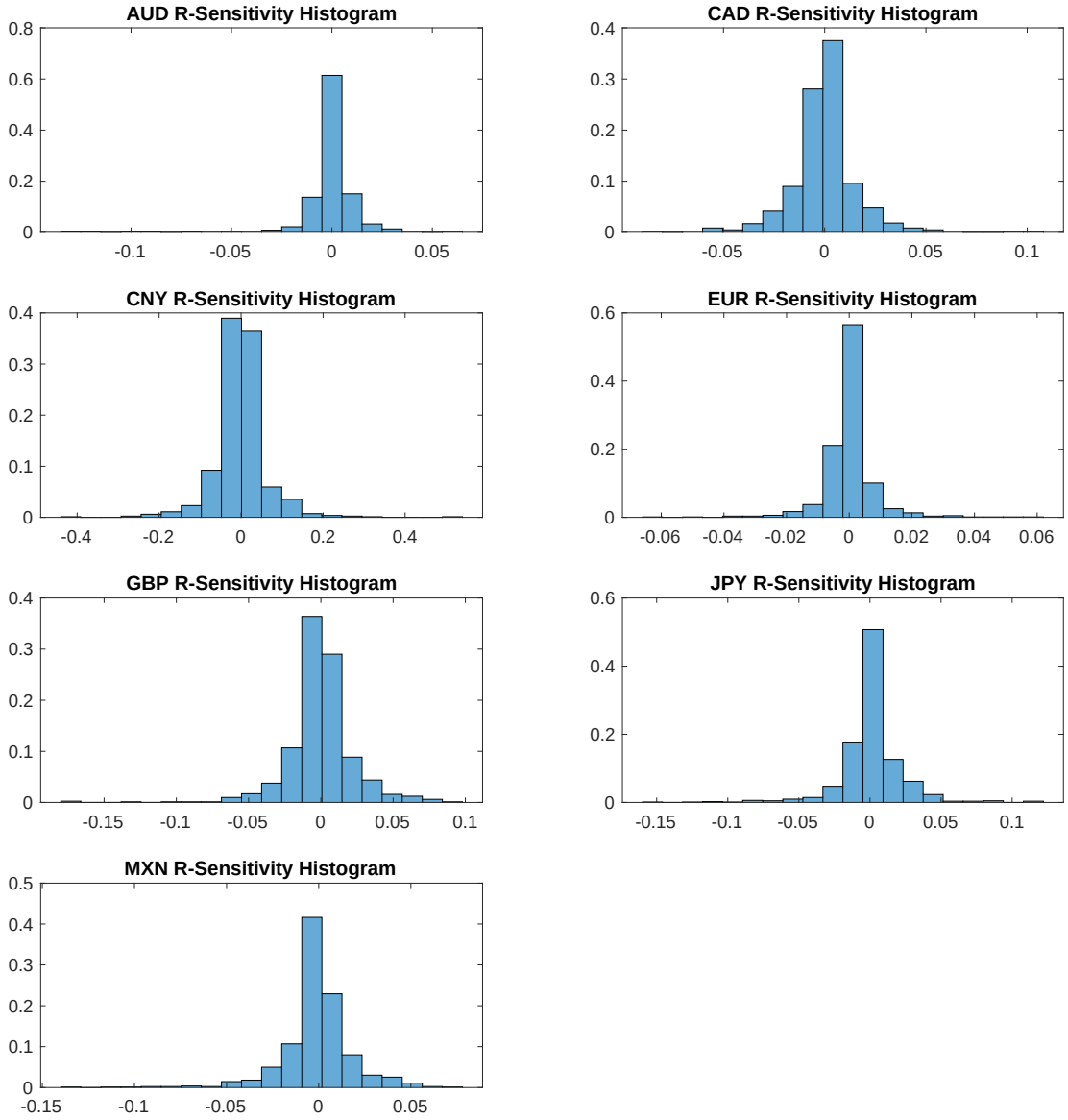

Figure 5.30: Distribution of the log-volatility error sensitivity to the transformed observation error covariance matrix.

Figures 5.24 - 5.29 provide diagnostic tools for the MSV's model performance during online estimation of the training set. In practice, as new observations $\mathbf{y}_t$ become available at time t, a Kalman filter is applied to obtain the volatility estimates of each FX series, while the diagnostic tools are implemented to assess the log-volatility

forecast error impact of this volatility estimate at time t due to each input parameter variations. These log-volatility forecast error sensitivities provides the analyst insight to model deficiency in an online manner, while the derivative information may provide guidance on parameter tuning for improved volatility estimation. As seen from Figure 5.26 and Figure 5.27, it is noted that the log-volatility forecast errors have the largest impact due to the prior estimates. This is to be expected since the initial priors were selected as $\mathbf{x}_0 \sim N(\mathbf{0}, \mathbf{P}_0)$ with $\mathbf{P}_0$ being a diagonal matrix whose entries are large ($10e^{10}$). The selection of this prior is common in the financial industry when very little information is known about the log-volatilities and selecting the large variances introduces the process as a diffusion process (see for example Chapter 11 of Tsay [53]). Nevertheless, the sensitivity analysis continues to provide performance diagnostics of the MSV model to the selection of these priors with sensitivity to background estimates results in line with expectations of the methodology.

On the other hand, the forecast sensitivity to the transformed observation error covariance matrix $\mathbf{R}$ of Figure 5.29 provides a diagnosis of each log-volatility estimate from the assumption that the linearized model with Gaussian distribution accurately approximates the true MSV model (5.25). Although these sensitivities show minimal impact to the log-volatility forecasts, their derivative information may be used in an adaptive procedure to produce updated log-volatilities based on an adaptive procedure of $\mathbf{R}$. If the q-MLE estimate of $\mathbf{R}$ provides reasonable accuracy of the Gaussian distribution to the true MSV model, then using the update procedure of $\mathbf{R}$ will have the added benefits to improve this approximation by providing updated log-volatility estimates during each time iteration. This procedure was implemented in Section 5.1 and their added benefit was shown during the exercise to demonstrate *proof-of-concept*. Thus, in the next section, this dissertation implements the update procedure of $\mathbf{R}^{1/2}$ in order to improve volatility estimates.

5.2.4 Adaptive Tuning

The adaptive procedure of Algorithm 4.1 is implemented to provide improved volatility estimates of the FX time series based on the derivative information of the forecast error sensitivities. Taking the search direction $\mathbf{p} = -\dfrac{\partial e}{\partial \mathbf{R}^{1/2}}$, an adaptive tuning step is added during each assimilation iteration of the Kalman filter with

$$\mathbf{R}_{\text{NEW}}^{1/2} = \mathbf{R}^{1/2} + \alpha \mathbf{p} \tag{5.34}$$

where $\mathbf{R}^{1/2}$ is the lower Cholesky covariance decomposition of the matrix $\mathbf{R}$. Recall that in the algorithm, if the log-volatility forecast error $e(\mathbf{x}_{t|t})$ is decreased under this new covariance matrix, the adaptive tuning step is accepted and a new analysis $\mathbf{x}_{t|t}$ is produce with a new forecast $\mathbf{x}_{t+1|t}$.

Proceeding to include the adaptive step during each time iteration of the Kalman filter using equation (5.34) and with the aid of Algorithm 4.1, updated volatility estimates are produced. Figure 5.31 presents the conditional volatility estimates of each FX rate during the training period, measured as the diagonal entries of $\mathbf{V}_t^{1/2}$, using the Kalman filter procedure (Original) and using the Kalman filter with adaptive $\mathbf{R}^{1/2}$ procedure (Updated). Figure 5.32 presents the step size α selected during each time iteration of the update procedure and serves as a guidance to acceptance/rejection of the adaptive step. It is observed from Figure 5.31 that the conditional volatilities of the AUD and EUR FX rates are similar under both procedures, however, the CAD, CNY, GBP, JPY and MXN FX rates conditional volatilities are different under the two procedure. For most of the FX rates, their conditional volatilities are different under the adaptive $\mathbf{R}^{1/2}$ procedure, while the plot of the step size α demonstrates that the adaptive procedure accepted many of these updates (825 steps accepted out of 834). Since the goal of the adaptive procedure is to reduce the value of the log-

volatility forecast score function ϵ, this dissertation accepts the updated procedure as an improvement to the estimation of FX rate conditional volatility.

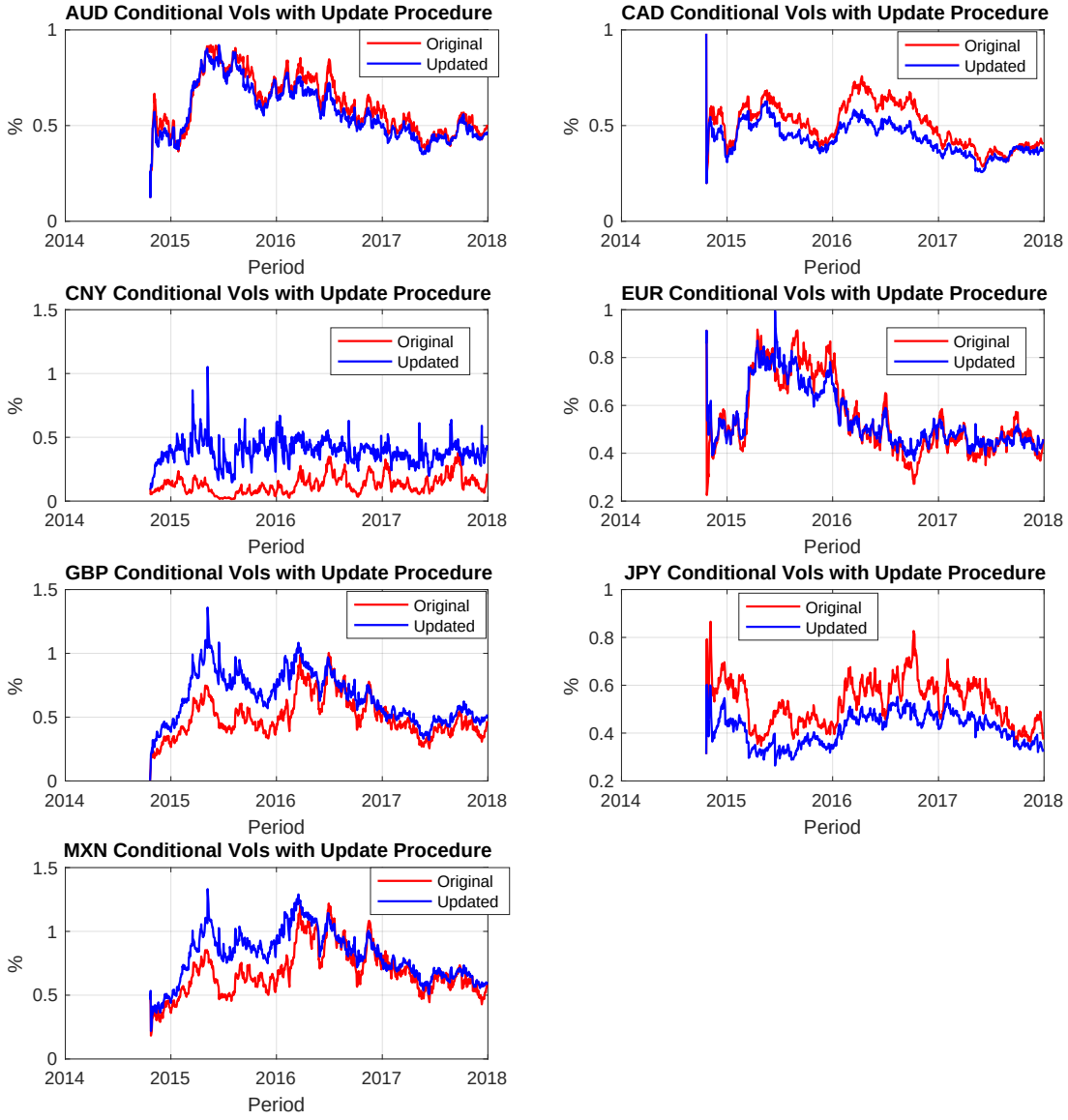

Figure 5.31: Conditional volatility of each FX rate measured as the diagonal entries of $\mathbf{V}_t^{1/2}$ during the training set. Volatilities using only the Kalman filter are labeled "Original" and volatilities using the Kalman filter with adaptive $\mathbf{R}^{1/2}$ are labeled "Updated".

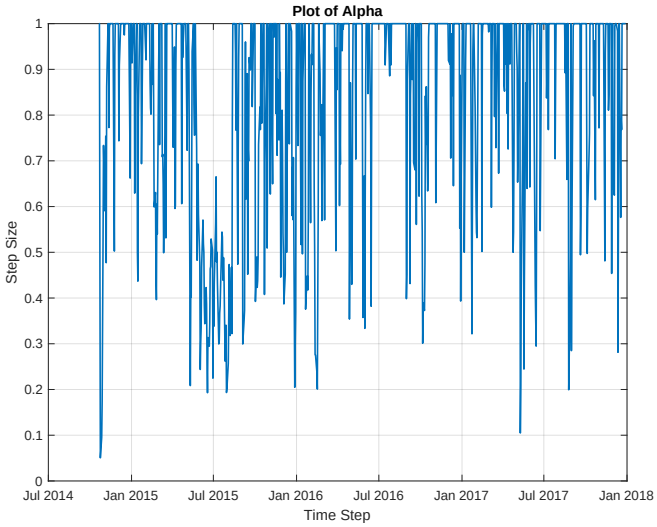

Figure 5.32: Step size α selected during each time iteration in the training set of the adaptive procedure.

The adaptive procedure of Algorithm 4.1 is also implemented during the out-of-sample test set to provide improved volatilities estimates. Figure 5.33 presents the conditional volatility estimates of each FX rate during the test set using the Kalman filter procedure (Original) and using the Kalman filter with adaptive $\mathbf{R}^{1/2}$ procedure (Updated). It is observed from Figure 5.33 that the conditional volatility estimates of each FX rate provide different results under each procedure. On the other hand, Figure 5.34 shows the step size α selected during each time iteration and it demonstrates the acceptance of these updates during the time iteration (461 steps accepted out of 469). Since the objective of the adaptive procedure is to provide improved conditional volatility estimates by minimizing the forecast error, this dissertation accepts these conditional volatilities under the adaptive $\mathbf{R}^{1/2}$ procedure as an improvement.

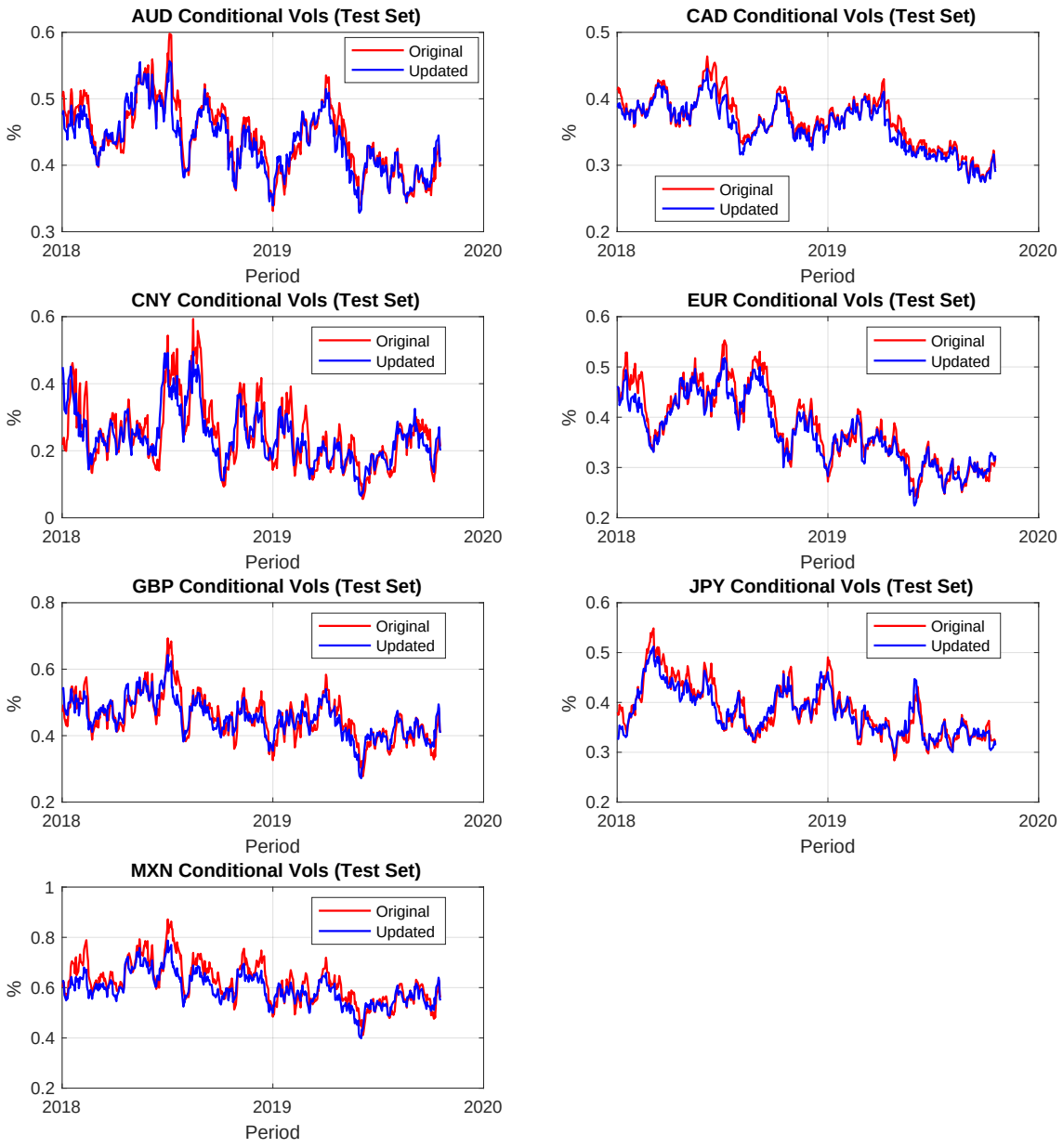

Figure 5.33: Conditional volatility of each FX rate measured as the diagonals of $\mathbf{V}_t^{1/2}$ during the test set. Volatilities using only the Kalman filter are labeled "Original" and volatilities using the Kalman filter with adaptive $\mathbf{R}^{1/2}$ are labeled "Updated".

Another added benefit to inclusion of the adaptive procedure is that the MSV model does not require to be re-estimated during the test set to provide reliable results. Recall that the main reason for model performance degradation in econometrics

is due to the fact that model parameters may become stale during the out-of-sample dataset and re-estimation (re-calibration of model parameters) of the MSV model may be required. On the other hand, the adaptive procedure overcomes this limitation as the covariance matrix $\mathbf{R}$ is being updated during each time iteration and, consequently, all volatility estimates will be updated.

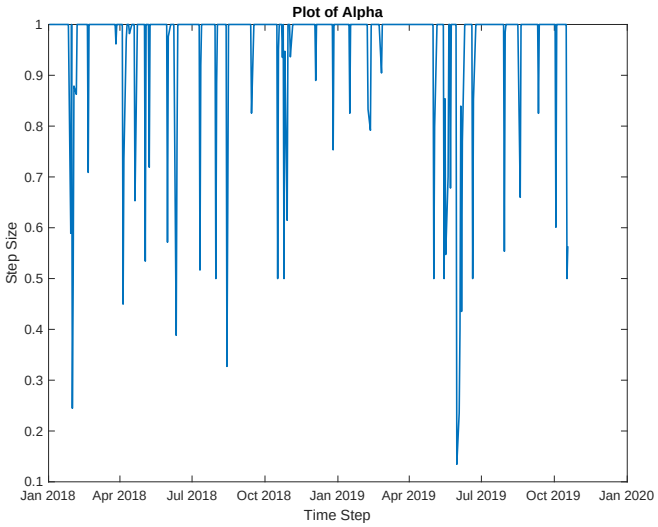

Figure 5.34: Step size α selected during each time iteration in the test set of the adaptive procedure.

Chapter 6

Conclusions and Future Directions

The Multivariate Stochastic Volatility (MSV) model requires an accurate specification of the unknown model parameters, in particular, the observation and background error covariance estimates. This dissertation presented a quasi-Maximum Likelihood Estimation (q-MLE) technique for estimation of these unknown parameters by first specifying the linearized MSV model, then calculating initial parameter estimates through a consistent methodology to ensure convergence of the q-MLE algorithm. Since these estimated parameters were obtained by fitting the MSV model to a dataset, the parameters inherently possess approximation errors as the estimation technique is only an approximation to the true model specification. Data assimilation methods were used in this dissertation as tools for obtaining conditional volatility estimates based on the fitted MSV model. This dissertation extended the current estimation literature on conditional multivariate volatility estimation by showing that the conditional volatility estimates obtained through a data assimilation method becomes dependent on the estimated model parameters. This dependance on model parameters was further extended to derive new model performance diagnostic tools in terms of the sensitivity of the log-volatility forecast error to the input parameters. The sensitivity information extracted from the data assimilation during each cycle provided an assessment on model performance due to model input variations and parameter misspecification. The sensitivity information also provided guidance on

parameter tuning for improved volatility estimation such that a new adaptive tuning procedure for the MSV was developed in order to demonstrate practical implementations in finance.

Future research directions stemming from this dissertation include the implementation of more advanced data assimilation techniques and its applications to the estimation of conditional volatility of swap rates. Chapter 2 presented many data assimilation methods that are available for estimation of multivariate conditional volatilities. In particular, the weak 4D-Var data assimilation methodology is of interest as it provides the flexibility to incorporate state innovation errors in the volatility estimation through the specification of the state error covariance matrix. As a consequence, future research is needed to understand the impact of the estimated state error covariance matrix to forecasted volatility estimates. Applications of the sensitivity analysis and adaptive tuning of the weak 4D-Var within the MSV model will be done to understand the behavior of interest rate swap volatility. The interest rate swap market is the largest over-the-counter derivatives market, with notional amounts in the trillions of dollars. This fixed-income market allows large financial institutions to trade Swaptions and Swaps, with Swaption expirations and Swap tenors ranging from one month to thirty years to accommodate the risk appetite of the investor. Swaptions and Swaps prices mainly depend on their corresponding forward swap rate, therefore, it is of interest in our research to model the volatility of the forward swap rate for the different Swaption expiries and Swap tenors as a multivariate stochastic process. The sensitivity analysis and adaptive tuning of the weak 4D-Var within the MSV model will become the main focus of our future research as these methodologies are capable of modeling the forward swap rates as a multivariate stochastic process, provide model performance diagnostics and be able to provide improved swap rate volatility estimates in an online manner.

References

[1] *Stochastic Volatility: Selected Readings*. Oxford University Press, Inc, 2004. 30

[2] T. G. Andersen, Richard A. Davis, Jr., J. P. Kreib, T. Mikosch, J. P. Kreia, and J. P. Kreiss. *Handbook of Financial Time Series*. Springer Publishing Company, Incorporated, 2016. 31, 38, 39

[3] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice-Hall, 1979. 14

[4] C. Andrieu and A. Doucet. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society*, 72(3):269–342, 2010. 26, 29

[5] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002. 26, 28, 29

[6] N. L. Baker and R Daley. Observation and background adjoint sensitivity in the adaptive observation-target problem. *Q J R Meteorol Soc*, 126:1431 – 1454, 2000. 45, 51

[7] L. Bauwens, S. Laurent, and J. V. K. Rombouts. Multivariate GARCH. a survey. *Journal of Applied Econometrics*, 21:346–354, 2006. 37

[8] A. F. Bennett and W. P. Budgell. The Kalman smoother for a linear quasi-geostrophic model of ocean circulation model. *Dynamics of Atmospheres and Oceans*, 13:219–267, 1989. 20

[9] N. Bergman. *Recursive Bayesian Estimation: Navigation and Tracking Applications*. PhD thesis, Linköping University, Linköping, Sweden, 1999. 27

[10] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973. 2, 30, 32, 135

[11] T. Bollerslev. Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 51:307–327, 1986. 36

[12] C. Broto and E. Ruiz. Estimation methods for stochastic volatility models: A survey. *Journal of Economic Survey*, 18(613-649), 2004. 30

[13] A. E Bryson and Y. C. Ho. *Applied Optimal Control: Optimization, Estimation, and Control*. Hemisphere Publishing Corporation, Washington, D.C., 1975. 20, 25

[14] D. Chan, R. Kohn, and C. Kirby. Multivariate stochastic volatility models with correlated errors. *Econometric Reviews*, 25:245–274, 2006. 31, 39

[15] P. Courtier, J. N. Thèpaut, and A. Hollingsworth. A strategy for operational implementation of 4d-var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 136(8):3050–3065, July 1994. 22

[16] D. N. Daescu. On the sensitivity equations of four-dimensional variational (4d-var) data assimilation. *Monthly Weather Review*, 136(8):3050–3–65, 2008. 45, 51

[17] D. N. Daescu and R. H. Langland. The adjoint sensitivity guidance to diagnosis and tuning of error covariance parameters. In Seon K. Park and Liang Xu, editors, *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, volume II, chapter 9, pages 205–232. Springer, 2013. 52, 54, 56, 59

[18] D. N. Daescu and R. H. Langland. Error covariance sensitivity and impact estimation with adjoint 4d-var: Theoretical aspects and first applications to NAVDAS-AR. *Q. J. Meteorol. Soc.*, (139):226–241, 2013. 45

[19] D. N. Daescu and R. Todling. Adjoint sensitivity of the model forecast to data assimilation system error covariance parameters. *Q. J. R. Meteorol. Soc.*, 136:1128–1145, 2010. 45

[20] J. Danielsson. Multivariate stochastic volatility models: Estimation and a comparison with vGARCH models. *Journal of Empirical Finance*, 5:155–173, 1998. 31

[21] P. de Jong. Smoothing and interpolation with the state space model. *Journal of the American Statistical Association*, 84:1085–1088, 1989. 13

[22] A. Doucet, J. F. G. de Freitas, and N. J. Gordon. *An Introduction to Sequential Monte Carlo Methods*. Springer-Verlag, 2001. 26, 27

[23] J. Durbin and S. J. Koopmans. *Time Series Analysis by State Space Methods*. Oxford University Press, Inc, Oxford, UK, 2001. 10, 133

[24] R. F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of the united kingdom inflation. *Econometrica*, 50:987–1007, 1982. 36

[25] R. F Engle and D. L. McFadden, editors. *Handbook of Econometrics*, volume 4. Elsevier Science B.V., 1994. 44

[26] E. F. Fama. The behavior of stock market prices. *Journal of Business*, 38(34-105), 1965. 30

[27] E. Ghysels, A. C. Harvey, and E. Renault. Stochastic volatility. *In: G. S. M. Rao, C. R. (Ed.): Statistical Models in Finance (Handbook of Statistics)*, pages 119–191, 1996. 30

[28] P. Hagan, D. Kumar, A. Lesniewski, and D. Woodward. Managing smile risk. *Willmot Magazine*, pages 84–108, September 2002. 35

[29] A. C. Harvey, E. Ruiz, and N. Shephard. Multivariate stochastic variance models. *Review of Economic Studies*, 61:247–264, 1994. 31, 42, 70

[30] Y. C. Ho. The method of least squares and optimal filtering theory. *The Rand Corporation, Memorandum*, 1962. 10

[31] J. Hull. *Options, Futures, and other Derivatives*. Pearson Education, Inc., ninth edition, 2015. 35

[32] J. Hull and A. White. The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42:281–300, 1987. 30

[33] K. Ito. Stochastic integral. *Proc. Imp. Acad.*, 20(8):519–524, 1944. 135

[34] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, NY, 1970. 4, 7, 9, 10, 14, 20, 23

[35] H. Johnson and D. Shanno. Option pricing when the variance is changing. *Journal of Financial and Quantitative Analysis*, 22:143–151, 1987. 30

[36] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 83 (Series D):35–45, 1960. 1, 4, 7, 25

[37] R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83(1):95–108, 1961. 4

[38] S. J. Koopmans, B. Jungbacker, and E. Hol. Forecasting daily variability of the SP 100 stock index using historical, realised, and implied volatility measurements. *Journal of Emperical Finance*, 12:445–475, 2005. 36

[39] Z. Li and M. Navon. Optimality of variational data assimilation and it relationship with the Kalman filter and smoother. *Q. J. R. Meteorol. Soc.*, 127:661–683, 2001. 17, 20, 24

[40] A. C. Lorenc. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, 112:1177–1194, 1986. 19, 20, 23

[41] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:394–419, 1963. 30

[42] R. Menard and R. Daley. The application of the Kalman smoother theory to the estimation of 4d-var error statistics. *Tellus A: Dynamic Meteorology and Oceanography*, 48(2):221–237, 1996. 20

[43] S. N. Neftci. *An Introduction to the Mathematics of Financial Derivatives*. Academic Press, San Diego, CA, 2000. 136

[44] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006. 66

[45] Y. Omori, S. Chib, N. Shephard, and J. Nakajima. Stochastic volatility with leverage: fast and efficient likelihood inference. *Journal of Econometrics*, 140:425–449, 2007. 31, 39

[46] E. Parzen. *Stochastic Processes*. Holden-Day, San Francisco, CA, 1962. 134

[47] F. Rabier and P. Courtier. Four-dimensional assimilation in the presence of baroclinic instability. *Q. J. R. Meteorol. Soc.*, 118:649–672, 1992. 19

[48] S. F. Schmidt. Applications of state space methods to navigation problems. *Advanced Control Systems*, 3:293–340, 1966. 10

[49] M. Smith and A. Pitts. Foreign exchange intervention by the bank of japan: Bayesian analysis using a bivariate stochastic volatility model. *Econometric Reviews*, 25:425–451, 2006. 31

[50] H. Song, I. Hoteit, B. D. Cornuelle, X. Luo, and A. C. Subramanian. An adjoint-based adaptive ensemble Kalman filter. *Monthly Weather Review*, 141(10):3343–3359, October 2013. 64

[51] S. J. Taylor. Financial returns modeled by the product of two stochastic processes - a study of daily sugar prices. *In: Anderson, O. D. (Ed.): Time Series Analysis: Theory and Practice*, 1:203–226, 1982. 30

[52] J. N. Thèpaut and P. Courtier. Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. *Q. J. R. Meteorol. Soc.*, 117:1125–1254, 1991. 19, 20

[53] R. S. Tsay. *Analysis of Financial Time Series*. John Wiley and Sons, Inc., Hoboken, NJ, 3rd edition, 2010. 12, 120

[54] P. Wilmott. *Derivatives: The theory and practice of financial engineering*. John Wiley and Sons, Inc., 1998. 31

[55] J. Yu. On leverage in a stochastic volatility model. *Journal of Econometrics*, 127:165–178, 2005. 31, 39

[56] Y. Zhu, R. Todling, and S. E. Cohn. Technical remarks on smoother algorithms'. DAO Office Note 99-2, NASA/GSFC Data Assimilation Office. 20

[57] E. Zivot. *Practical Aspects of GARCH-Modeling In: Andersen, T. G., Davis, R. A., Kreiss, J. P., and Mikosch, T. (Eds.): Handbook of Financial Time Series.* Springer, New York, NY, 2008. 36

[58] D. Zupanski. A general weak constraint applicable to operational 4d-var data assimilation systems. *Monthly Weather Review*, 125:2274–2292, 1997. 21

Appendix A

Probability Theory

We will present the basic definitions of stochastic processes that will aid in the implementation and representation of *Stochastic Differential Equations* that are used throughout this dissertation.

A.1 Stochastic Processes

A **Stochastic Process** is a family of random variables $\{x_t, t \in T\}$ that are indexed by a parameter set T. Now, the parameter set T can be a discrete set or a continuous set, depending on the application. If for each $t \in T$, the random variable x_t is continuous, we say that the process has a *continuous state space*. Otherwise, the stochastic process has a *discrete state space* if the random variable x_t has a discrete outcome.

Stochastic processes are characterized by specifying their *joint density function*. That is, all questions regarding the probabilistic properties of a stochastic process can be answered by specifying

$$p(x_{t_1}, ..., x_{t_n}) \tag{A.1}$$

for all finite sets $\{t_1, ..., t_n\} \subset T$. We say that the stochastic process $\{x_t, t \in T\}$ is a *Gaussian process* if their joint density function specified in (A.1) is **Gaussian** (**normal**). Since the normal distribution can be characterized by the first two moments (mean and variance), Gaussian processes can also be fully characterized by their first two moments. In general, one can specify any stochastic process $\{x_t, t \in T\}$ with any known distribution in order to accommodate the data. For example, Durbin [23] gives a treatment of **State-Space** modeling of financial time series by specifying non-Gaussian distributions such as Poisson, Binary, Binomial, Multinomial and Heavy-Tailed Distributions, to name a few.

Another stochastic process that is useful throughout the dissertation is the **Brownian Motion** (also called **Wiener**) process.

Definition A.1.1. *A stochastic process $\{x_t, t \geq 0\}$ is a Brownian motion if*

- *$\{x_t, t \geq 0\}$ has stationary independent increments. That is, the probability distribution of $x_{t+h} - x_{\tau+h}$, for any $h > 0$ is the same as $x_t - x_\tau$ (stationary)*

and
$x_{t_2} - x_{t_1}, ..., x_{t_n} - x_{t_{n-1}}$ *are independent for any finite set* $\{t_1, ..., t_n\}$

- *for all* $t \geq 0$, x_t *is normally distributed*
- *for all* $t \geq 0$, $\mathcal{E}\{x_t\} = 0$
- $x_0 = 0$ *with probability 1*

Clearly, the increments of Brownian motions have

$$\mathcal{E}\{x_t - x_\tau\} = 0, \forall t, \tau \geq 0$$

and, Parzen [46] showed that

$$\text{var}\{x_t - x_\tau\} = \sigma^2(t - \tau), \forall t, \tau \geq 0$$

This goes to show that Brownian motions are completely characterized by their first two moments.

The final ingredient needed to complete our review in stochastic processes is the notion of **Markov processes**. Markov processes are very important, useful in practice and applicabile in many sciences. Broadly speaking, a Markov process is a stochastic process whose probability about the future state of the process $x_{t_{n+1}}$ depends only on the previous information of the process x_{t_n} and no other previous information is needed. Formally, a stochastic process $\{x_t, t \in T\}$ is said to have the Markov property if

$$p(x_{t_n}|x_{t_1}, ..., x_{t_{n-1}}) \quad = \quad p(x_{t_n}|x_{t_{n-1}})$$

We can make use of the Markov property to have a very useful representation of the probability law. Indeed, for any finite set $\{t_1, ..., t_n\} \subset T$ we have

$$\begin{aligned} p(x_{t_n}, ..., x_{t_1}) &= p(x_{t_n}|x_{t_1}, ..., x_{t_{n-1}}) \cdot p(x_{t_{n-1}}|x_{t_1}, ..., x_{t_{n-2}}) \cdot ... \cdot p(x_{t_2}|x_{t_1}) \cdot p(x_{t_1}) \\ &= p(x_{t_n}|x_{t_{n-1}}) \cdot p(x_{t_{n-1}}|x_{t_{n-2}}) \cdot ... \cdot p(x_{t_2}|x_{t_1}) \cdot p(x_{t_1}) \end{aligned} \tag{A.2}$$

Where, in the first equation, we used the multiplication law of probabilities and, in the second equation, we used the Markov property. Therefore, in order to completely characterize a Markov stochastic process, one only needs the conditional densities $p(x_t|x_\tau)$ for all $t > \tau$, called the **transition probabilities.** Equation (A.2) is useful in calculating the probability distribution of a discrete State-Space model as shown in Chapter 2.

Appendix B

Option Pricing

B.1 Black-Scholes Model for Pricing

After the major break through of the work of Fischer Black and Myron Scholes, the use of stochastic differential equations became the backbone in financial mathematics. Black and Scholes [10] were able to price call and put options via stochastic differential equations. In their original work, they assumed that the evolution of the price of an asset S_t at time t followed a stochastic differential equation

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{B.1}$$

where μ is the instantaneous mean and σ is the instantaneous volatility of the process S_t and dW_t is a Brownian motion. Notice that the drift (μS_t) and diffusion (σS_t) components are allowed to not only change with respect to time, but to also change as the asset price of S_t changes. If we let $V(S_t, t)$ be the price of a call option that depends on the price of the underlying asset S_t, we can then use Ito's lemma [33] to derive the stochastic differential equation that governs the call option

$$dV = \left(\mu S \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + \frac{\partial V}{\partial t}\right) dt + \sigma S \frac{\partial V}{\partial S} dW \tag{B.2}$$

where we substituted equation (B.1) and have dropped the subscript time index to ease notation.

One can then put the price of the underlying asset S and the price of the call option V in a portfolio P whose evolution is given by $dP = \theta_1 dV + \theta_2 dS$. Since the market practitioner chooses the weights θ_1, θ_2, we can choose $\theta_1 = 1$ and $\theta_2 = -\frac{\partial V}{\partial S}$. This gives the evolution of the portfolio to be

$$dP = \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\right) dt$$

Notice that the choice of such weights gave the cancellation of the stochastic process S_t. This means that there are no random disturbances driving the portfolio. That is, the portfolio is *completely predictable* and thus "risk free". Since there is no risk in P

and in order to avoid arbitrage, the return of the portolio over a time dt is given by $dP_t = rP_t dt$, where r is the risk-free (constant) interest rate. This equation says that the riskless portolio should give returns equal to investing the portfolio and recieving a risk-free rate of return; otherwise, arbitrage opportunities in the portfolio will rise. We then have the relationship

$$rPdt = \left(\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\right) dt$$

Cancelling dt and substituting the value of the portfolio $P = V - \frac{\partial V}{\partial S} S$ with the given weights, we have

$$r\left(V - S\frac{\partial V}{\partial S}\right) = \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}$$

Rearranging the equation we have the much celebrated Black-Scholes partial differential equation for pricing call and put options

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0, \qquad S_t \geq 0, 0 \leq t \leq T \tag{B.3}$$

The solution is the price of the call option $V_t = V(S_t, t)$. In order to solve the partial differential equation, we need the following boundary condition. At time of maturity T, the call option will have value $V_T = \max(S_T - K, 0)$, where K is the strike price of the call. For a put option, the value at maturity is given by $V_T = \max(K - S_T, 0)$.

The Black-Scholes model became the framework for pricing other types of derivative securities. One can take this framework and assume much more realistic assumptions. For a further introduction of stochastic differential equations with financial applications, the book of Neftci [43] presents a comprehensive treatment of the models most currently used by market practitioners.