2010

ESTIMATION AND FORECASTING STOCHASTIC VOLATILITY MODELS WITH VOLATILITY OBSERVABLE

Linlan Xiao
Western University

ESTIMATION AND FORECASTING STOCHASTIC VOLATILITY MODELS WITH VOLATILITY OBSERVABLE

(Spine Title: Stochastic Volatility Models, Volatility Proxies)

(Thesis Format: Integrated Article)

by

Linlan Xiao
Department of Economics

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES
THE UNIVERSITY OF WESTERN ONTARIO
LONDON, CANADA
2010

THE UNIVERSITY OF WESTERN ONTARIO
SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Supervisor	Examiners
______________________ Dr. John Knight	______________________ Dr. Youngki Shin
Supervisory Committee	______________________ Dr. Stephen Sapp
______________________ Dr. Stephen Sapp	______________________ Dr. Reginald Kulperger
______________________ Dr. Martijn van Hasselt	______________________ Dr. Tony Wirjanto

The thesis by

Linlan Xiao

entitled

Estimation and Forecasting Stochastic Volatility Models with Volatility Observable

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date ______________________ ______________________
Chair of the Thesis Examining Board

To my parents, my husband, and my son

Table of Contents

List of Tables

List of Figures

Abstract

My dissertation consists of four essays focusing on the estimation and forecasting of the discrete and continuous-time stochastic volatility (SV) models with volatility observable.

The first essay examines the estimation of discrete-time SV models via a Monte Carlo study with both lagged inter-temporal and contemporaneous dependencies when volatility is observed. The statistical properties of both models are studied. Treating volatility as an observable variable, we apply traditional estimation methods including both full information maximum likelihood (FIML) and three-stage least squares (3SLS) methods. The estimation is straightforward and easy to implement. The Monte Carlo results suggest that both methods do a reasonable job at recovering the true parameters when the underlying volatility is observed. When the underlying volatility is unobserved, we should be careful in choosing an appropriate proxy such that the proxy error does not spread too much, and in this case, both FIML and 3SLS are able to provide good estimates.

The second essay, which is closely related to the first essay, focuses on estimating and forecasting the discrete-time SV models with lagged inter-temporal and contemporaneous dependencies using realized volatility. This essay contributes to the literature in three aspects. First, we examine the estimation of the discrete-time SV models with lagged inter-temporal and contemporaneous dependencies using realized volatility. Second, we investigate forecasting performance of discrete-time SV model with contemporaneous dependence. Third, we use realized volatility not only to evaluate the out-of-sample forecasting performance, but also in the in-sample estimation. The empirical results show that both FIML and 3SLS estimators produce good finite sample properties. The forecasting performances of four competing models, including SV models with lagged inter-temporal and contemporaneous dependencies, the simple regression model, and the heterogeneous autoregressive (HAR) model, are compared.

In the third essay, we extend our study to examine the estimation via a Monte Carlo study of the affine continuous-time SV model when volatility is observed. Specifically, we apply the consistent approximate maximum likelihood method (C-AMLE). We simulate asset returns and volatilities at both daily and monthly frequencies. The Monte Carlo results suggest that the C-AMLE approach does a good job at recovering the true parameters.

The fourth essay focuses on investigating the estimation of the affine continuous-time SV model using volatility proxies. Both realized volatility and model-free implied volatility are employed. We apply the C-AMLE approach as well as the quasi-maximum likelihood (QML) method. Our empirical analysis is based on both daily and monthly data of S & P 500 index and Dow Jones Industrial Average indexes. In general, the C-AMLE approach ourperforms the QML when the model-free implied volatility is used.

Keywords: Discrete-Time Stochastic Volatility Model, Full Information Maximum Likelihood, Three-Stage Least Squares, Affine Continuous-Time Stochastic Volatility Model, Consistent Approximate Maximum Likelihood, Quasi-Maximum Likelihood, Realized Volatility, Model-Free Implied Volatility.

Acknowledgements

I am deeply indebted to my supervisor, Professor John Knight, for his effort, patience, and advice, all of which made this work possible. I would also like to thank other members of my Ph. D. committee, Professor Stephen Sapp and Professor Martijn van Hasselt, for their valuable comments and constant encouragement.

I would like to thank Professor Audra Bowlus, Professor Hiroyuki Kasahara, Professor Youngki Shin for useful suggestions and comments. I also wish to express my deep gratitude to Professor Paul Natke for reading and providing comments on earlier version of my thesis. My sincere thanks are due to Professor Audra Bowlus, Professor Todd Stinebrickner, Professor Lance Lochner, Yvonne Adams, and Sharon Phillips for their helps when I was on the job market.

Finally, I would like to thank all the faculty, staff, and fellow students in the Economics Department at the University of Western Ontario, all of whom made my experience at Western unforgettable.

London, Canada — Linlan Xiao

July, 2010

Chapter 1

Introduction

My doctoral thesis consists of four essays focusing on estimation and forecasting of stochastic volatility (SV) models when volatility is observed. The first essay investigates the estimation of discrete-time SV models with both lagged inter-temporal and contemporaneous dependencies when volatility is observed via a Monte Carlo study. The second essay concentrates on the estimation and forecasting of those models using realized volatility. The third essay examines the estimation performance of the consistent approximate maximum likelihood (C-AMLE) method for the affine continuous-time SV model when volatility is observed via a Monte Carlo study. The fourth essay investigates the estimation of the affine continuous-time SV model using both realized and model-free implied volatilities.

Financial time series exhibit certain stylized facts, such as insignificant autocorrelation of asset return while profound serial correlation of squared and absolute returns, heavy tails of asset returns, leverage effect, volatility clustering, etc. Various models have been developed to capture these stylized facts. Among them, the SV model has been well recognized to be able to pick up the observed asymmetric behavior of time series and capture the heavy tails of asset returns hence placed at the center of volatility research of realistic pricing of options, efficient asset allocation and accurate risk assessment. However, in SV framework, the underlying volatility is latent, explicit maximum likelihood estimation and inference are not feasible. Researchers proposed numerous estimation methods. In general, these methods suffer from some drawbacks, such as failure of convergence, inefficiency, or the implementation is rather complicated and demands time consuming computation. The drawbacks of these methods treating volatility as latent motivate our interest in investigating estimation of SV models treating volatility as observed in the first essay. We investigate estimation of two alternative discrete-time SV models, SV model with lagged inter-temporal dependence, and SV model with contemporaneous dependence, with volatility observable via a Monte Carlo study. Jiang, Knight and Wang (2005) examined the theoretical differences between these two discrete-time SV models and showed the moments and cross moments of returns were different. We extend their study by deriving both conditional and unconditional correlations between volatility and past/future asset returns and show that the explicit expressions are different for two models. Treating volatility as an observable variable, we apply both full information maximum likelihood (FIML) and three-stage least squares (3SLS) procedures and undertake Monte Carlo experiments to examine their performance. We first undertake a Monte Carlo assuming that the true volatility is observed, the evidence shows that both methods do a reasonable job at recovering the true parameters. In realistic situations, volatility is not observable but often suitable proxies are available. Naturally there exists a measurement error associated with the volatility proxy. So we undertake a Monte Carlo experiment taking into account the volatility proxy error, and examine the estimation performance of both FIML and 3SSL. The evidence suggests that both methods do a reasonable job if an efficient volatility proxy is selected.

The second essay examines the estimation and forecasting performance of discrete-time SV models using realized volatility. Our study contributes to the literature in three aspects. First we examine the estimation of discrete-time SV models with both lagged inter-temporal and contemporaneous dependencies using realized volatility. Second, we first examine forecasting performance of discrete-time SV model with contemporaneous dependence. And third, we use daily realized volatility not only to evaluate the out-of-sample forecasting performance, but also in the in-sample estimation. Our empirical analysis is based on both low frequency daily and high frequency intra-day observations for S & P 500 index and three exchange rates, including CAD/USD, DEM/USD, and USD/GBP. When we construct daily realized volatility, we employ the "volatility signature plot" to select the optimal sampling frequency, we also consider different approaches to deal with the "closed effect" and the first order autocorrelation of high frequency asset returns. Using realized volatility, both the FIML and 3SLS estimators produce good finite sample properties, suggesting that both methods are appropriate when volatility is treated as an observable variable, and the statistical inference is reliable by using daily realized volatility as a proxy. We then examine the forecasting performance of four different models, including the above two models, a simple regression model, and heterogeneous autoregressive (HAR) model, using realized volatility. We first apply the famous Diebold and Mariano (1995) tests, the results indicate that the competing models provide unequally accurate forecasts. We then use four criteria, namely the root mean squared error (RMSE), the mean absolute error (MAE), the Theil's-U, and QLIKE, to evaluate the point forecasts of each candidate model. The result shows that the HAR model provides the most accurate point forecasts.

The third essay extends our study of the SV model by examining the estimation of the affine continuous-time SV model. As the continuous-time SV model can explain some empirical features of the joint time-series behavior of stock and option prices, such as time varying volatility, fat tails of asset return distribution, etc., it has dominated the option pricing literature since the mid-1980s. However, traditional inference for the continuous-time SV model has been viewed as difficult for some time. The continuous sample of observations is unavailable and thus often requiring the model to be discretized and introducing a significant discretization error. The estimation is even more difficult as the volatility process can not be directly observed. In addition, except for a few cases, the transition density does not have a closed form expression hence maximum likelihood method is not directly available. In practice, the estimation is computationally demanding or involves discretization error or is based on simulation methods. It is noticed that for the affine diffusion and affine jump diffusion processes, although the transition density functions are unknown, the corresponding conditional characteristic functions (CCF) can be derived explicitly, hence the empirical characteristic function (ECF) method can be applied. Jiang and Knight (2010) advanced the ECF approach by proposing an analytical approximation of the optimal weight function via an Edgeworth/Gram-Charlier expansion of the logarithmic transition density function. Their approach is similar to the approximate maximum likelihood (AMLE) method, but ensures the consistency of the estimation hence is named the consistent AMLE (C-AMLE). For the affine continuous-time SV model, when both return and volatility processes are observed, the volatility state variable does not have to be integrated out of the joint CCF hence the implementation of the C-AMLE is straightforward and computationally easy. In the third essay we examine the estimation performance of the C-AMLE method via a Monte Carlo study. We simulate the volatility process from its unconditional density function and generate return process using the almost exact simulation method, at both daily and monthly frequencies. The moments calculated from our simulations are very close to the true moments, suggesting that our simulations are accurate. We then apply the C-AMLE procedure for the affine continuous-time SV model, the result suggests that the C-AMLE method does a good job at recovering the true parameters.

The last essay investigates the estimation of the affine continuous-time SV model via an empirical study. Volatility measures have been widely employed in modern academic and financial market practitioner literatures over the last two decades. One popular measure is realized volatility constructed from high frequency data, and the other is model-free implied volatility inferred from option prices. Motivated by the accuracy of these volatility proxies and computational convenience by using them in the estimation, in the fourth essay, we employ both realized volatility and model-free implied volatility, and apply the C-AMLE as well as the quasi-maximum likelihood (QML) procedures for the affine continuous-time SV model. Our empirical analysis is based on returns and volatilities of both the S & P 500 and Dow Jones Industrial Average indexes. The evidence shows that using daily realized volatilities, neither the C-AMLE or the QML procedure fit the data. However, when using monthly data, the estimation improves. Especially when we employ the model-free implied volatilities, the estimates are stable, and the moments of returns and model-free implied volatilities calculated from the estimates are close to the moments of real time series. In general, we find that the C-AMLE outperforms the QML approach.

1.1 Bibliography

DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–63.

JIANG, G., AND J. KNIGHT (2010): "ECF Estimation of Markov Models Where the Transition Density is Unknown," *Econometrics Journal*, 13, 245–70.

JIANG, G., J. KNIGHT, AND G. WANG (2005): "Alternative Specifications of Stochastic Volatility Models-Theoretical and Empirical Comparison," Working paper.

Chapter 2

Estimation of Stochastic Volatility Models with Volatility Observable

2.1 Introduction

Volatility plays an important role in both financial theory and financial markets. The widely recognized positive trade-off relationship between risk and return demands for accurate estimates and forecasts of return volatility as volatility is a common definition of risk. The large volatility in the equity and foreign exchange markets has a significant impact on the whole economy hence can raise a policy issue. In financial markets, the pricing of derivatives often requires the volatility of the underlying asset to be estimated. In some derivatives, the underlying asset is volatility itself. Further, for hedging against risk and portfolio management, accurate volatility estimates and forecasts are crucial.

As discussed in McNeil, Frey, and Embrechts (2005), financial time series exhibit certain stylized facts:

(1) The autocorrelation in returns is often small and insignificant, however, the series of absolute or squared returns show profound serial correlation.

(2) Asset return series are leptokurtic or heavy-tailed. As Mandelbrot (1963), and further Fama (1963) observed, the distribution of asset returns exhibits excess kurtosis or fatter tails than those of a normal distribution.

(3) Volatility appears to vary over time.

(4) Volatility clustering. Financial time series observations always reveal high volatility followed by high volatility and similarly for low volatility, indicating notable clustering but also persistency of volatility.

(5) Leverage effects. As Black (1976) first suggested, the past stock returns are negatively associated with future volatility. Nelson (1991), Gallant, Rossi and Tauchen (1992), Campbell and Kyle (1993) reported empirical evidence on leverage effects.

These observations about asset return and volatility series led researchers to focus on the study of these stylized facts. Various approaches have been proposed to capture these stylized facts. A benchmark model is the so called Autoregressive Conditional Heteroscedasticity (ARCH) model, proposed by Engle (1982). In the ARCH framework, the conditional variance is a time-varying, positive, linear function of past squared errors, and thereby able to capture volatility clustering in financial data. Bollerslev (1986) extended the ARCH model by allowing the conditional variance to be a linear function of both past squared errors and lagged conditional variances, and developed the new popular Generalized ARCH (GARCH) model. On the other hand, Taylor (1986) formulated a discrete-time stochastic volatility (SV) model as an alternative to the ARCH/GARCH model. In the SV framework, unlike ARCH/GARCH, volatility is allowed to follow a stochastic process with its own error source.

The ARCH/GARCH model has become very popular in the modeling financial time series due to its ability to capture many of the stylized facts but perhaps, more importantly, due to its ease of estimation via maximum likelihood. However, it is important to notice that for a standard ARCH/GARCH model, positive and negative past values have a symmetric effect on the conditional variance, while it is very likely financial series are strongly asymmetric and subject to the "leverage effect" between asset return and volatility. In the SV framework, correlation between error terms in the asset return process and the conditional volatility process enables the SV model to pick up the often observed asymmetric behavior. In particular, a negative correlation between the two error terms induces a leverage effect. Moreover, in the SV model, the asset return is a mixture of distributions, thus the excess kurtosis or fat tails, a significant stylized fact, is able to be captured. In fact, as Kim, Shephard and Chib (1998) showed, the standard SV models perform better in-sample than the GARCH models.

In the discrete-time SV model framework, there are different specifications of the dependence between the conditional volatility and the asset return. There are models with lagged inter-temporal dependence and models with contemporaneous dependence. The former assumes that the disturbance term in the conditional volatility process is correlated to that in the lagged asset return process, while the latter assumes those two error terms are correlated contemporaneously. While most studies focus on the SV model with lagged intertemporal dependence due to its tractability, the SV model with contemporaneous dependence has received little attention in the literature. In their recent study, Jiang, Knight and Wang (2005) investigated the properties of SV models with both lagged inter-temporal and contemporaneous dependence. They theoretically showed that an SV model with contemporaneous dependence was more flexible in its ability to fit into the high kurtosis of asset returns than an SV model with lagged inter-temporal dependence.

Although SV models have been recognized as able to overcome some of the drawbacks of the ARCH/GARCH models and placed at the center of volatility research for the realistic pricing of options, efficient asset allocation and accurate risk assessment, the estimation is always viewed as difficult since volatility itself is a random process and not observable. Explicit maximum likelihood estimation and inference are not feasible, and thus most researchers focus on estimating the SV models using moment conditions. An incomplete list of alternative estimation methods includes the method of moments (MM) proposed by Taylor (1986), the efficient method of moments (EMM) introduced by Bansal, Gallant, Hussey and Tauchen (1993), Engle (1994), Gallant and Tauchen (1996), and Jiang and van der Sluis (2000), the simulated method of moments (SMM) by Duffie and Singleton (1993), the generalized method of moments (GMM) introduced by Melino and Turnbull (1990), Andersen (1994) and Andersen and Sorensen (1996), the Quasi-Maximum likelihood (QML) by Harvey, Ruiz and Shephard (1994), the simulated maximum likelihood (SML) estimation by Danielsson and Richard (1993), the Bayesian Markov Chain Monte Carlo (MCMC) method proposed by Shephard (1993), Jacquier, Polson and Rossi (1994, 2004), and Chib, Elerian and Shephard (2001), and the empirical characteristic function (ECF) by Singleton (2001), Knight and Yu (2002) and Jiang and Knight (2002, 2010), etc.[1]

Overall, these estimation methods have advantages in some respects, on the other hand, they suffer from some drawbacks, such as failure of convergence, inefficiency, or the implementation is rather complicated and demands

[1]Broto and Ruiz (2004) provide an excellent survey of estimation methods for SV models.

time consuming computation. The main difficulty is that in the SV framework, volatility is latent, the estimation is solely based on the observations of asset returns. The drawbacks of these methods in the estimation of the discrete-time SV models when volatility is latent motivate our interest in investigating estimation of SV models treating volatility as observed. In this chapter, we investigate estimation of two alternative discrete-time SV models, SV model with lagged inter-temporal dependence, and SV model with contemporaneous dependence when volatility is observed. In Jiang, Knight, and Wang (2005), the theoretical differences between these two discrete-time SV models were examined via an analysis of various moments and cross moments of asset returns. In a small empirical study, they showed that the SV model with contemporaneous dependence fitted the asset return data the best. We extend their study by deriving both conditional and unconditional correlations between volatility and past/future asset returns, and show that the explicit expressions are significantly different for two models. In Jiang, Knight and Wang (2005), they treated volatility as unobserved and considered GMM in estimating both models. The estimated correlation coefficient parameter ρ in general is not very desirable. In our study, we treat volatility as an observable variable, consequently we can easily apply some traditional estimation methods, specifically, we are able to apply full information maximum likelihood (FIML), or its asymptotic equivalent, three stage least squares (3SLS) method. The estimation is simple and easy to implement. In order to examine the estimation performance of these traditional methods, we undertake a Monte Carlo experiment. We first undertake a Monte Carlo assuming that the true volatility is observable. This assumption is not realistic, but it sets a benchmark for comparison. In realistic situations, volatility is not observable but often suitable proxies are available. Naturally there exists a measurement error associated with the volatility proxy. So we undertake a Monte Carlo experiment taking into account the volatility proxy error, and examine the estimation performance of both FIML and 3SLS. We compare the means, medians, and standard errors of the estimators and note that the evidence supports our approaches.

The outline of the rest of this chapter is follows. In Section 2.2, we briefly review the literature on discrete-time stochastic volatility models as well the estimation methods existing in the literature. In Section 2.3, the statistical properties of alternative SV model specification are discussed. We report the exact moment results from Jiang, Knight and Wang (2005) and derive lag-lead correlations between asset returns and volatility. Section 2.4 concentrates on the estimation procedures of FIML and 3SLS for the two models. Section 2.5 reports the result of Monte Carlo experiments. A brief conclusion is contained in Section 2.6.

2.2 Literature Review

2.2.1 Discrete-time SV Model

Volatility is widely believed to be time varying, persistent, and clustered. An explanation for changing volatility, as Shephard (1996) mentioned, would be to assume that price changes occur due to a random number of intra-daily price movements responding to information arrivals following the work of Clark (1973) and Tauchen and Pitts (1983). Over the last two decades, numerous models have been developed to study changing volatility. Following Cox (1981), these models could be divided into two types, namely, observation-driven and parameter-driven models, respectively. The simplest examples of the former are ARCH and GARCH models introduced by Engle (1982) and Bollerslev (1986), respectively. Alternatively, the parameter-driven SV models allow the variance to depend not on past observations, but on its own error term in a autoregressive specification.

Typically, if we let S_t denote the asset price at time t, then $x_t = ln(S_t/S_{t-1}) - \mu$ refers to the excess asset return. The GARCH (1,1) model may be written as:

$$x_t = \sigma_t \varepsilon_t \quad (2.1a)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (2.1b)$$

where $\varepsilon_t \sim i.i.d.N(0,1)$. The parameters must satisfy $\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0$, and $\alpha_1 + \beta_1 < 1$ to ensure that the conditional variance is positive and that the excess asset return series $\{x_t\}$ is covariance stationary. Obviously in the GARCH (1,1) process, the variance is a deterministic function of past variance and squared returns.

The SV model is similar to the GARCH (1,1) process except it introduces another error source into the conditional variance equation:

$$x_t = \sigma_t \varepsilon_t \quad (2.2a)$$

$$\sigma_t^2 = exp(h_t) \quad (2.2b)$$

$$h_t = \alpha + \beta h_{t-1} + \sigma \eta_t \quad (2.2c)$$

where $\varepsilon_t \sim i.i.d.N(0,1)$, $\eta_t \sim i.i.d.N(0,1)$. The two error terms may be correlated with each other.

Shephard (1996) provides an excellent comparison of ARCH/GARCH and SV models' statistical properties. In the ARCH/GARCH models, the method of maximum likelihood makes the estimation statistically efficient and computationally easy. While the mixture of distributions in the SV model specification creates difficulties in estimation, it enables the model to capture both excess kurtosis as well as the asymmetric behavior of asset returns. (See Ghysel, Harvey and Renault (1996)).

There are two leading explanations for the asymmetry in the relationship between asset returns and volatility, namely the feedback effect and the leverage effect. The feedback effect is associated with the positive relationship between asset returns and volatility. As French, Schwert and Stambaugh (1987) and Campbell and Hentschel (1992) discussed, the higher volatility demands a higher asset return. The leverage effect, on the other hand, as Black (1976) and Christie (1982) first discussed, indicates that volatility increases when the stock price falls. Little agreement has been achieved about the magnitude of the two effects.

In the studies of the relationship between the asset return and volatility, most researchers focus on the lagged inter-temporal dependence due to its tractability. Yet, within the SV model framework, dependence between asset returns and conditional volatility, namely contemporaneous dependence has also been specified in the literature. However, as the statistical properties for this model are unknown, it has held little attention of academics and practitioners. Jiang, Knight and Wang (2005) investigated and compared the properties of the SV models under the two alternative specifications. They derived closed form expressions for the moments and cross moments of asset returns and showed that the statistical properties of asset returns under these two specifications were clearly different. The model with contemporaneous dependence allows for negative skewness of the asset return distribution.

2.2.2 Estimation of SV Models

As Broto and Ruiz (2004) discussed, since volatility is latent, estimation approaches of the SV models are mainly based on the statistical properties of the return process. The most popular approaches are based on the method of moments (MM), first introduced to estimate SV models by Taylor (1986). Melino and Turnbull (1990), and later Andersen (1994), Andersen and Sorensen (1996), Jiang, Knight and Wang (2005), Bollerslev and Zhou (2006) also proposed the generalized method of moments (GMM) to estimate SV models, and showed that the estimator was consistent. GMM is easy to implement, however, the finite sample properties are very poor, and the estimator is not as efficient as maximum likelihood.

Some studies focus on the maximum likelihood via the Monte Carlo Markov Chain (MCMC) procedures. Shephard (1993), and Jacquier, Polson, and Rossi (1994) were the first to propose the MCMC procedure for the estimation of SV models. Later, Chib, Elerian and Shephard (2001) also employed the MCMC to estimate SV models. The advantage of MCMC procedures is that as the simulation size becomes very large, they have asymptotically the same distribution as the maximum likelihood estimator. Moreover, as the inference is based on finite sample distributions, the asymptotic approximation is not needed. However, the empirical implementation of MCMC is complicated and computationally demanding.

Gourieroux, Monfort, and Renault (1993) proposed the indirect inference method. Specifically, they proposed the quasi-likelihood function as an auxiliary model to estimate the continuous-time SV model. On the other hand, Bansal, Gallant, Hussey and Tauchen (1993), Gallant and Tauchen (1996), proposed the efficient method of moments (EMM). The EMM procedure is computationally easier than the indirect inference procedure, and has been extensively used in the estimation of both continuous and discrete-time SV models, such as Engle (1994), Gallant, Hsieh, and Tauchen (1997), Jiang and van der Sluis (2000), etc. As Gourieroux, Monfort, and Renault (1993) showed, the indirect inference and EMM methods are asymptotically equivalent. The former approach performs better in finite samples, the latter approach is computationally easier. However, both approaches are computationally demanding.

Nelson (1988), and later Harvey, Ruiz, and Shephard (1994), linearized the SV models by taking logarithms of the nonlinear equations in SV models hence applied the Kalman filter to obtain the quasi-maximum likelihood function of logarithm of squared asset returns. The QML procedure is flexible, easy to implement, and the estimator is consistent and asymptotically normal. However, this procedure is inefficient since the estimation is based on an approximated likelihood function. As Ruiz (1994) showed, the finite sample bias could be significant.

In SV framework, volatility follows a random process with its own disturbance, hence the asset return series has a mixture distribution. Theoretically this property makes SV models attractive as it is more realistic, however, the empirical application is difficult as explicit likelihood function is unavailable. It has been noticed that although a closed form of the transition density function is unknown, the associated conditional characteristic function (CCF) of the state variables can be derived explicitly. Building on this observation, Singleton (2001), Jiang and Knight (2002, 2010), proposed the empirical characteristic function (ECF) method to estimate both discrete and continuous-time SV models. The basic idea is to minimize the integrated distance between the empirical characteristic function (ECF) and their theoretical counterpart, the CCF. As Jiang and Knight (2002, 2010) pointed out, due to the one to one correspondence between the transition density function and the CCF, the estimator is asymptotically consistent, and efficient. However, as volatility, one of the state variables, is unobserved, it has to be integrated out of the joint CCF, the estimation is complicated in practice.

In the past two decades, some volatility proxies have been developed in the literature. The most popularly used proxies include realized volatility (RV) constructed from high frequency intra day transaction prices, and model-free implied volatility (MFIV) which is ex-ante risk neutral expectations of future market volatilities computed from option prices. These volatility proxies have been proved to be consistent, accurate estimators for the unknown true volatility, and widely used in modeling and forecasting volatility ever since. The development of volatility proxies motivates our interest in examining estimation of SV models when volatility is an observable variable. In our study, we treat volatility as observed, consequently, we are able to apply some traditional estimation methods, including FIML as well 3SLS, given the joint distribution of asset return and volatility processes. Clearly the estimation will be straightforward, and easy to implement.

2.3 Statistical Properties of Alternative SV Model Specifications

Jiang, Knight and Wang (2005) showed that the statistical properties were different between the two specifications. In some sense, SV model with contemporaneous dependence fitted better to the asset returns. In this section, we further analyze some of the statistical properties of these models stating the results from Jiang, Knight and Wang (2005) and extending them to derive conditional and unconditional lag-lead correlations between returns and volatility.

2.3.1 SV Model with Lagged Inter-temporal Dependence (Model 1)

Let $\{x_t\}$, $t = 1, \ldots, T$ denote the asset return at time t, $\{e^{h_t}\}$ represent the volatility (variance) of the asset return, and $\{h_t\}$ is the natural logarithm of the volatility, the SV model with lagged inter-temporal dependence specifies the correlation between past asset returns and current volatility:

$$x_t = \lambda e^{h_t} + e^{h_t/2}\varepsilon_t \tag{2.3a}$$

$$h_t = \alpha + \beta h_{t-1} + \sigma v_t \tag{2.3b}$$

$$\begin{pmatrix} \varepsilon_{t-1} \\ v_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \tag{2.3c}$$

In this type of model, only lagged ε and current v, that is, lagged asset return and current volatility are correlated with each other.

An alternative way to write this model is to replace the second equation and distribution with the following:

$$h_{t+1} = \alpha + \beta h_t + \sigma v_t$$

$$\begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$$

There are five unknown parameters $\lambda, \alpha, \beta, \sigma$, and ρ in this model. In the asset return equation, the parameter λ refers to the risk-return trade-off, intuitively it should take a positive value. In the second equation, the logarithmic volatility follows an AR (1) process with a constant α. The parameter β represents the persistence in volatility and is usually assumed to be positive. The larger β, the more persistent the volatility. The parameter σ is the volatility of volatility which is positive. And ρ refers to the correlation coefficient which lies within a natural band [-1,1].

It is noted that in the standard SV model which we discussed in Section 2, the asset return x_t is equal to $e^{\frac{1}{2}h_t}\varepsilon_t$, hence the conditional first moment of the asset return does not involve h_t, whereas in Model 1, an extra term λe^{h_t} is introduced in the asset return equation to take into account the risk-return trade-off (or feedback effect). The conditional first moment of the asset return is:

$$E_{t-1}(x_t) = E_{t-1}(\lambda e^{h_t} + e^{h_t/2}\varepsilon_t) = E_{t-1}(\lambda e^{\alpha+\beta h_{t-1}+\sigma v_t}) = \lambda e^{\alpha+\beta h_{t-1}+\frac{1}{2}\sigma^2}$$

With λ being positive, the larger h_{t-1}, the larger expected asset return is.

2.3.2 SV Model with Contemporaneous Dependence(Model 2)

The SV model with contemporaneous dependence specifies the correlation between current asset return and volatility:

$$x_t = \lambda e^{h_t} + e^{h_t/2}\varepsilon_t \tag{2.5a}$$

$$h_t = \alpha + \beta h_{t-1} + \sigma v_t \tag{2.5b}$$

$$\begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}) \tag{2.5c}$$

From the model above, obviously current period asset return and volatility are correlated with each other. The mixture of distributions leads to excess kurtosis for asset returns.

Same as in Model 1, λe^{h_t} is included in the asset return equation in Model 2. However, as Jiang, Knight and Wang (2005) proved, the conditional expected asset return for Model 2 is different from that for Model 1.

For Model 2, the conditional expected return is[2]

$$E_{t-1}(x_t) = e^{\alpha+\beta h_{t-1}+\frac{1}{2}\sigma^2}\lambda + \tfrac{1}{2}e^{\frac{1}{2}\alpha+\frac{1}{2}\beta h_{t-1}+\frac{1}{8}\sigma^2}\rho\sigma$$

Both models have a common term in their conditional expectation which is positive if $\lambda > 0$. However, Model 2 contains an extra term which may be either positive or negative depending on the sign of correlation coefficient ρ. As Jiang, Knight and Wang (2005) demonstrated, the extra term in Model 2 is induced by the asymmetry between the return and volatility. If ρ is negative, a higher risk premium of volatility would be required in Model 2 in order to have the same level of expected return as in Model 1.

[2]see Jiang, Knight and Wang (2005) and Appendix 1 for proof.

2.3.3 Cross Moments for Model 1 and Model 2

Jiang, Knight and Wang (2005) derived the closed form expressions for the moments for each of the models above. We reproduce their results in Proposition 1.1 and Proposition 1.2.[3]

Proposition 1: (Conditional Return Moments) The conditional cross moments of $x_t$and x_{t+i} for Model 1 and Model 2 are:

$$
\begin{aligned}
E_{t-1}(x_t^l x_{t+i}^p) = \quad & \sum_{k=0}^{l}\sum_{s=0}^{p}\binom{l}{k}\binom{p}{s}\lambda^{l+p-k-s} \\
& *exp(\frac{\alpha((l+p-(k+s)/2)-\beta(l-k/2+(p-s/2)\beta^i))}{1-\beta}) \\
& *exp(\beta(l-k/2+(p-s/2)\beta^i)h_{t-1}) \\
& *exp(\frac{(1-\beta^2)(l-k/2+(p-s/2)\beta^i)^2+(p-s/2)^2(1-\beta^{2i})}{2(1-\beta^2)}\sigma^2) \\
& *C(l,k,p,s)
\end{aligned}
$$

where

$$C(l,k,p,s) = W_k((p-s/2)\rho\sigma\beta^{i-1})W_s(0)$$

for Model 1 and

$$C(l,k,p,s) = W_k(\rho\sigma((p-s/2)\beta^i+(l-k/2)))W_s((p-s/2)\rho\sigma)$$

for Model 2.

Proposition 2: (Unconditional Return Moments) The unconditional cross

[3]See Jiang, Knight and Wang (2005) page 9-10 and Appendix 2 for details and proof.

moments of $x_t$and x_{t+i} for Model 1 and Model 2 are:

$$
\begin{aligned}
E(x_t^l x_{t+i}^p) = \; & \sum_{k=0}^{l}\sum_{s=0}^{p}\binom{l}{k}\binom{p}{s}\lambda^{l+p-k-s} \\
& *exp(\frac{(l-k/2)^2+(p-s/2)^2+2\beta^i(l-k/2)(p-s/2)}{2(1-\beta^2)}\sigma^2) \\
& *exp(\frac{\alpha(l+p-\frac{k+s}{2})}{1-\beta})C(l,k,p,s)
\end{aligned}
$$

where

$$C(l,k,p,s) = W_k((p-s/2)\rho\sigma\beta^{i-1})W_s(0)$$

for Model 1 and

$$C(l,k,p,s) = W_k(\rho\sigma((p-s/2)\beta^i + (l-k/2)))W_s((p-s/2)\rho\sigma)$$

for Model 2.

2.3.4 Conditional and Unconditional Correlations between Volatility and Past and Future Returns

The asymmetric relationship between asset return and volatility processes has been studied over the past two decades. There are two leading explanations for this volatility asymmetry, namely, the leverage effect and volatility feedback effect, respectively. The so called leverage effect, referring to a negative correlation between asset return and volatility, was first discussed by Black (1976), and later Christie (1982). Intuitively, the lower past return causes an increase in the debt-to-equity ratio, hence results in a higher future volatility. In contrast, French, Schwert and Stambaugh (1987) find a positive correlation between asset return and volatility. The volatility feedback effect implies that an increase in volatility would demand a higher future return.

Based on different volatility models, theoretical analysis of correlations between volatility and returns has been implemented. For example, building on Heston (1993) one-factor continuous SV model, Bollerslev and Zhou (2006) provide theoretical analysis of relationship between return and realized volatility and implied volatility. Shi (2005) derives the closed-form expression for both conditional and unconditional correlations between volatility and past and future returns (lag-lead correlation) for a discrete-time SV in mean model.

In our study, we derive both conditional and unconditional lag-lead correlations between asset return and volatility processes for SV models with both lagged inter-temporal and contemporaneous dependence. The explicit expressions are reported below.

Proposition 3: The conditional lag-lead correlations between return and volatility for Model 1 and Model 2 are as follows; the proofs are given in Appendix 3:

1. The conditional lag correlations for Model 1 and Model 2 are:

$$\rho_{x_t, e^{h_{t-i}} | I_{t-i-1}} = 0$$

for $i \geq 0$

for Model 1 and

$$\rho_{x_t, e^{h_{t-i}} | I_{t-i-1}} = \frac{1}{\sqrt{V_1(e^{\sigma^2} - 1)}} (\lambda e^{\alpha \frac{1-\beta^{i+1}}{1-\beta} + \beta^{i+1} h_{t-i-1} + \frac{1-\beta^{2(i+1)}}{2(1-\beta^2)} \sigma^2} (e^{\beta^i \sigma^2} - 1) + \frac{1}{2} \rho \sigma e^{\alpha \frac{1-\beta^{i+1}}{2(1-\beta)} + \frac{1}{2} \beta^{i+1} h_{t-i-1} + \frac{1-\beta^{2(i+1)}}{8(1-\beta^2)}} (e^{\frac{1}{2} \beta^i \sigma^2} - 1))$$

for $i \geq 0$

for Model 2.

2. The conditional lead correlations for Model 1 and Model 2 are:

$$\rho_{x_t, e^{h_{t+i}}|I_{t-1}} = \frac{\rho\sigma\beta^{i-1}}{\sqrt{e^{\sigma^2\frac{1-\beta^{2i}}{1-\beta^2}}-1}}$$

for $i \geq 1$

for Model 1 and

$$\rho_{x_t, e^{h_{t+i}}|I_{t-1}} = \frac{1}{\sqrt{V_2(e^{\sigma^2\frac{1-\beta^{2(i+1)}}{1-\beta^2}}-1)}}(\lambda e^{\alpha+\beta h_{t-1}+\frac{1}{2}\sigma^2}(e^{\beta^i\sigma^2}-1)+ \frac{1}{2}\rho\sigma e^{\frac{1}{2}\alpha+\frac{1}{2}\beta h_{t-1}+\frac{1}{8}\sigma^2}((1+2\beta^i)e^{\frac{1}{2}\beta^i\sigma^2}-1))$$

for $i \geq 1$

for Model 2.

Proposition 4: The unconditional lag-lead correlations between return and volatility for Model 1 and Model 2 are as follows:

1. The unconditional lag correlations for Model 1 and Model 2 are as follows:

$$\rho_{x_t, e^{h_{t-i}}} = \frac{1}{\sqrt{V_3(e^{\sigma^2\frac{1}{1-\beta^2}}-1)}}\lambda e^{\alpha\frac{1}{1-\beta}+\frac{1}{2(1-\beta^2)}\sigma^2}(e^{\frac{1}{1-\beta^2}\beta^i\sigma^2}-1)$$

for $i \geq 0$

for Model 1 and

$$\rho_{x_t, e^{h_{t-i}}} = \frac{1}{\sqrt{V_4(e^{\sigma^2\frac{1}{1-\beta^2}}-1)}}(\lambda e^{\alpha\frac{1}{1-\beta}+\frac{\sigma^2}{2(1-\beta^2)}}(e^{\frac{1}{1-\beta^2}\beta^i\sigma^2}-1)+\frac{1}{2}\rho\sigma e^{\alpha\frac{1}{2(1-\beta)}+\frac{\sigma^2}{8(1-\beta^2)}}(e^{\frac{1}{2(1-\beta^2)}\beta^i\sigma^2}-1))$$

for $i \geq 0$

for Model 2.

2. The unconditional lead correlations for Model 1 and Model 2 are:

$$\rho_{x_t,e^{h_{t+i}}} = \frac{1}{\sqrt{V_3(e^{\sigma^2\frac{1}{1-\beta^2}}-1)}}(\lambda e^{\alpha\frac{1}{1-\beta}+\frac{1}{2(1-\beta^2)}\sigma^2}(e^{\frac{1}{1-\beta^2}\beta^i\sigma^2}-1)+\rho\sigma\beta^{i-1}e^{\frac{\alpha}{2(1-\beta)}+\frac{1+4\beta^i}{8(1-\beta^2)}\sigma^2})$$

for $i \geq 1$

for Model 1 and

$$\rho_{x_t,e^{h_{t+i}}} = \frac{1}{\sqrt{V_4(e^{\sigma^2\frac{1}{1-\beta^2}}-1)}}(\lambda e^{\alpha\frac{1}{1-\beta}+\frac{1}{2(1-\beta^2)}\sigma^2}(e^{\frac{1}{1-\beta^2}\beta^i\sigma^2}-1)+$$
$$\frac{1}{2}\rho\sigma e^{\frac{1}{2(1-\beta^2)}\alpha+\frac{1}{8(1-\beta^2)}\sigma^2}((1+2\beta^i)e^{\frac{1}{2(1-\beta^2)}\beta^i\sigma^2}-1))$$

for $i \geq 1$

for Model 2.

V_1, V_2 are conditional variance of asset return process in Proposition 3, V_3, V_4 represent unconditional variance of return process in Proposition 4. The explicit expression for these different V_s is provided in Appendix 3.

First we compare the conditional lag correlations for Model 1 and Model 2. Clearly they are different. For Model 1, the conditional correlation between return and current/past volatility is zero. This is obvious given the assumption that the error term in return process is only correlated with the future error term in the volatility process. Alternatively the conditional lag correlation for Model 2 is nonzero. Moreover, the persistence parameter β is naturally assumed to be positive to capture "volatility clustering" feature, and the local volatility parameter σ can not be negative, the sign of conditional lag correlations for Model 2 is controlled by signs of risk trade-off parameter λ and correlation coefficient parameter ρ. If signs of both parameters are negative, then the conditional lag correlations for Model 2 is negative. If the signs of the two parameters are opposite, then it is ambiguous. By setting $i = 1$, we get the conditional leverage effect, while when letting $i = 0$, the conditional feedback effect is obtained.

The closed-form expressions for the conditional lead correlations for the two models are both nonzero but different. For Model 1, the sign is determined by the sign of correlation coefficient parameter ρ and independent of trade off parameter λ indicating that letting $\rho \neq 0$ induces the conditional asymmetric response of volatility to the positive or negative lagged return. If $\rho < 0$, then the current return is negatively associated with future volatility, i.e. leverage effect. However, for Model 2 the sign of conditional lead correlation is controlled by ρ along with λ. In particular, setting both ρ and λ negative, the conditional leverage effect is captured.

Proposition 4 provides the closed-form expressions for the unconditional lag-lead correlations for both models. Unlike the conditional lag correlation for Model 1, the unconditional lag correlation for this model is independent of correlation coefficient parameter ρ. This result is consistent with the assumption of Model 1 that error term in return process is only correlated with future error term in volatility process. While the sign of the unconditional lag correlation for Model 1 is solely determined by parameter λ, it is controlled by both λ and ρ for Model 2.

The sign of the unconditional lead correlations depends on λ and ρ while the magnitude is determined by parameters $\lambda, \alpha, \beta, \rho$ and σ for both models. In particular, if λ and ρ are both negative, the current return is negatively associated with future volatility process for both models.

Overall, the expressions of both cross moments of the asset returns and lag-lead correlations between returns and volatilities are significantly different between SV model with inter-temporal dependence and SV model with contemporaneous dependence.

2.4 Estimation of Alternative SV Models Treating Volatility as Observable

Jiang, Knight and Wang (2005) studied the SV model with both lagged inter-temporal and contemporaneous dependences. They showed that the SV model with contemporaneous dependence fitted the asset returns better than the SV model with lagged inter-temporal dependence. As a consequence, they claimed that the contemporaneous dependent SV model deserved attention. From the various moment results they derived, they suggested GMM as a suitable estimation method. Unfortunately, results from both the Monte Carlo experiment and an empirical study were not very desirable. These results, further highlight the estimation difficulties in SV models when volatility is latent.

In this study, we treat volatility as observed in both models. Therefore we employ traditional estimation methods such as full information maximum likelihood (FIML)(or equivalently feasible generalized least square (FGLS)) and three-stage least squares (3SLS) approaches in the estimation.

2.4.1 Full Information Maximum Likelihood (FIML) or Feasible Generalized Least Squares (FGLS) Estimation for Model 1

When volatility (or variance) e^{h_t} is an observable variable, given a time series of observations $\{x_t, e^{h_t}\}, t = 1, \ldots, T$, traditional estimation methods can result in consistent and efficient estimators.

For Model 1, since the disturbance terms ε_{t-1} and v_t are correlated with each other, given the observations $\{x_t, e^{h_t}\}, t = 2, \dots, T$, the full information maximum likelihood (FIML) or equivalently feasible generalized least squares (FGLS) approach can be applied.

We transform the asset return equation in Model 1 by multiplying $e^{-\frac{1}{2}h_t}$ to both sides, resulting in:

$$y_t = x_t e^{-\frac{1}{2}h_t} = \lambda e^{\frac{1}{2}h_t} + \varepsilon_t$$

Let the transformed asset return equation be equation 1, the volatility equation be equation 2. The system of equations can be written as:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \lambda \\ \pi \end{pmatrix} + \begin{pmatrix} \varepsilon \\ \eta \end{pmatrix}$$

where

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1T-1} \\ y_{22} \\ \vdots \\ y_{2T} \end{pmatrix} = \begin{pmatrix} (x_1)e^{-\frac{1}{2}h_1} \\ \vdots \\ (x_{T-1})e^{-\frac{1}{2}h_{T-1}} \\ h_2 \\ \vdots \\ h_T \end{pmatrix}$$

$$\begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} = \begin{pmatrix} x_{11} & 0 \\ \vdots & \vdots \\ x_{1T-1} & 0 \\ 0 & x_{21} \\ \vdots & \vdots \\ 0 & x_{2T-1} \end{pmatrix} = \begin{pmatrix} e^{\frac{1}{2}h_1} & 0 & 0 \\ \vdots & \vdots & \vdots \\ e^{\frac{1}{2}h_{T-1}} & 0 & 0 \\ 0 & 1 & h_1 \\ \vdots & \vdots & \vdots \\ 0 & 1 & h_{T-1} \end{pmatrix}$$

$$\pi = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

$$\begin{pmatrix} \varepsilon \\ \eta \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{T-1} \\ \eta_2 \\ \vdots \\ \eta_T \end{pmatrix} \sim N(0, \Sigma \otimes I)$$

since

$$\begin{pmatrix} \varepsilon_{t-1} \\ \eta_t \end{pmatrix} \sim N(0, \Sigma)$$

where $\Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}$, $\eta_t = \sigma v_t$.

Model 1 is a seemingly unrelated regression (SUR) system. The asset return equation and the volatility equation seem unrelated, however they are related through the correlation in the error terms. For a SUR system, the generalized least squares (GLS) procedure is commonly applied and results in efficient estimators.

The GLS estimators are:

$$\begin{pmatrix} \hat{\lambda} \\ \hat{\pi} \end{pmatrix}$$

$$= \left(\begin{pmatrix} X_1' & 0 \\ 0 & X_2' \end{pmatrix} \begin{pmatrix} I & \rho\sigma I \\ \rho\sigma I & \sigma^2 I \end{pmatrix}^{-1} \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \right)^{-1} \begin{pmatrix} X_1' & 0 \\ 0 & X_2' \end{pmatrix} \begin{pmatrix} I & \rho\sigma I \\ \rho\sigma I & \sigma^2 I \end{pmatrix}^{-1} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

Simplifying above expression, the vector of parameters can be estimated by:

$$\hat{\lambda} = (X_1'(I - \rho^2 P_{X_2})X_1)^{-1}X_1'(I - \rho^2 P_{X_2})Y_1 - \frac{\rho}{\sigma}(X_1'(I - \rho^2 P_{X_2})X_1)^{-1}X_1'(I - P_{X_2})Y_2$$

and

$$\hat{\pi} = (X_2'X_2)^{-1}X_2'Y_2 - \rho\sigma(X_2'X_2)^{-1}X_2'(Y_1 - X_1\hat{\lambda})^4$$

where $\hat{\pi} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}$.

The estimators $\hat{\lambda}$, $\hat{\pi}$ are functions of coefficient parameter ρ and variance parameter σ which are unknown, so we need to estimate these two parameters by running OLS to each equation independently. The procedure is the so called feasible generalized least squares (FGLS).

The procedure is as follows:

Step 1: get initial residuals.

We run OLS to equation 1 and equation 2 independently, hence get initial estimates $\hat{\lambda}_0$, $\hat{\pi}_0$. Consequently, initial residuals $\hat{\varepsilon}_0$ and $\hat{\eta}_0$ can be obtained.

Step 2: construct estimated variance-covariance matrix

Given initial residuals, the estimated variance-covariance matrix of two error terms can be computed. In particular, the estimated variance parameter σ is calculated as:

$$\hat{\sigma_0} = \sqrt{var(\hat{\eta_0})}$$

The estimated correlation coefficient parameter ρ is:

$$\hat{\rho_0} = \frac{cov(\hat{\varepsilon_0}, \hat{\eta_0})}{\hat{\sigma_0}}$$

[4]Detail is provided in Appendix 4.

Step 3: substitute $\hat{\sigma_0}, \hat{\rho_0}$ into the formula for $\hat{\lambda}$ and $\hat{\pi}$.

Step 4: compute the new residuals, hence obtain new estimates of variance parameter and correlation coefficient parameter.

We then go back to step 3 and step 4, repeat until the vector of parameters converge.

2.4.2 Full Information Maximum Likelihood (FIML) and Three Stage Least Square (3SLS) for Model 2

The system of equations for Model 2 is:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \lambda \\ \pi \end{pmatrix} + \begin{pmatrix} \varepsilon \\ \eta \end{pmatrix}$$

where

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \\ y_{22} \\ \vdots \\ y_{2T} \end{pmatrix} = \begin{pmatrix} (x_2)e^{-\frac{1}{2}h_2} \\ \vdots \\ (x_T)e^{-\frac{1}{2}h_T} \\ h_2 \\ \vdots \\ h_T \end{pmatrix}$$

$$\begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} = \begin{pmatrix} x_{11} & 0 \\ \vdots & \vdots \\ x_{1T-1} & 0 \\ 0 & x_{21} \\ \vdots & \vdots \\ 0 & x_{2T-1} \end{pmatrix} = \begin{pmatrix} e^{\frac{1}{2}h_2} & 0 & 0 \\ \vdots & \vdots & \vdots \\ e^{\frac{1}{2}h_T} & 0 & 0 \\ 0 & 1 & h_1 \\ \vdots & \vdots & \vdots \\ 0 & 1 & h_{T-1} \end{pmatrix}$$

$$\pi = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

$$\begin{pmatrix} \varepsilon \\ \eta \end{pmatrix} = \begin{pmatrix} \varepsilon_2 \\ \vdots \\ \varepsilon_T \\ \eta_2 \\ \vdots \\ \eta_T \end{pmatrix} \sim N(0, \Sigma \otimes I)$$

since

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N(0, \Sigma)$$

where Σ is as defined earlier. $\eta_t = \sigma \upsilon_t$.

FIML for Model 2

In Model 2, because dependent and independent variables both take current values in equation 1, the variable Y_2 is an independent variable in equation 1, while it appears as a dependent variable in equation 2. Under the assumption that $cov(\varepsilon_t, \eta_t) = \rho\sigma$, the explanatory variable is correlated to the error term in equation 1. As a result, if we apply OLS to both equations at the first step, the OLS estimator for the first equation is not only biased but also inconsistent. Therefore, we need to introduce an instrumental variable to remove this correlation. A valid instrumental variable needs to be uncorrelated with the error term while highly related to the explanatory variable. The difficulty is, given the fact that Model 2 is a non-linear simultaneous equation model, we are not very clear what instrument would be the optimal choice. We employ three different instruments: the first instrument we use is the lagged volatility, $exp(h_{t-1}/2)$, we call it $IV1$; we then use polynomials of h_t, which we call $IV2$; the third instrument we use is the special case of polynomials of h_t, that is a constant and h_{t-1}, we denote it as $IV3$. In Monte Carlo experiments, we use these three different instruments in the estimation, and compare the performance. We find the estimation results are quite similar from these three instruments.

The procedure of FIML for Model 2 is very similar to that for Model 1, except that before we run OLS, we pre-multiply the instrumental variable to both sides of equation 1. Then we just follow the same procedure as we did for Model 1 until the vector of parameters converge.

3SLS for Model 2

Sargan (1964), Court (1974) and Amemiya (1977) discussed the three stage least squares (3SLS) approach to estimate simultaneous econometric models. As Sargan (1964) showed, 3SLS approach in the case of a nonsingular disturbance covariance matrix is asymptotically equivalent to FIML. The 3SLS estimator can be derived by first multiplying the set of predetermined variables to the system of equations then applying generalized least squares formula to the resulting transformed system. The set of parameters thus can be estimated by:

$$\hat{\lambda} = (X_1'X_2(X_2'X_2)^{-1}X_2'X_1)^{-1}X_1'X_2(X_2'X_2)^{-1}X_2'Y_1$$

and

$$\hat{\pi} = (X_2'X_2)^{-1}X_2'Y_2 - \rho\sigma(X_2'X_2)^{-1}X_2'(Y_1 - X_1\hat{\lambda})^5$$

[5] Appendix 4 provides detail.

The 3SLS does not involve iterating. If iterating, we will, in essence, get the FIML estimators.

The procedure of estimation is as follows:

Step 1: 2SLS on the first equation and get residual $\hat{\varepsilon}$

We use the set of predetermined variable, X_2, as instrument variable. The 2SLS gives $\hat{\lambda} = (X_1'X_2(X_2'X_2)^{-1}X_2'X_1)^{-1}X_1'X_2(X_2'X_2)^{-1}X_2'Y_1$. The residual then can be computed.

Step 2: OLS $\hat{\varepsilon}$ on X_2 to get estimate of ψ.

$$\hat{\varepsilon} = X_2\psi + \omega$$

$$\hat{\psi} = (X_2'X_2)^{-1}X_2'\hat{\varepsilon}$$

Step 3: OLS to the second equation and get residual $\hat{\eta}$.

Step 4: form the covariance of two residuals.

Once we get residual from two equations, the covariance is formed as $\hat{cov}(\hat{\varepsilon}, \hat{\eta}) = \frac{1}{T}\hat{\varepsilon}'\hat{\eta}$.

Step 5: obtain the 3SLS estimate of π.

The three-stage least square estimate of π has the formula

$$\hat{\pi_{3SLS}} = \hat{\pi_{OLS}} - cov(\hat{\varepsilon}, \hat{\eta})\hat{\psi}$$

2.5 Monte Carlo Evidence

In previous sections, we showed that treating volatility as an observable variable in the SV framework, FIML as well 3SLS can be applied. The estimation is straightforward and easy to implement. In this section, the objective is to examine estimation performance of the SV models via Monte Carlo experiments when volatility is observed.

We set the values of five parameters as $\lambda = 0.1, \alpha = -1, \beta = 0.8, \rho = -0.5, \sigma = 0.5$[6] for both models. The sample size is set as 3000, and the simulation is repeated 1000, 5000, and 10000 times, respectively.

We implement Monte Carlo experiments considering two different assumptions.

In the first Monte Carlo experiment, we concentrate on examining estimation performance of FIML and 3SLS methods when volatility is observed, hence we assume that the true volatility is observed. Given the assumption that the vector of error terms follow a bivariate normal distribution, we are able to simulate daily asset return series as well as the volatility series from model 1 and model 2, respectively. We then use these simulated series in the estimation of Model 1 and Model 2, respectively. The simulation and estimation are repeated, and the means, medians and standard deviations of the estimates are computed to compare with the true parameters.

In reality volatility is not observed, however, studies have shown that some suitable proxies are available. Unfortunately, all proxies have errors and none are perfect. Consequently, in our second Monte Carlo experiment, we examine the estimation when the measurement error is of various levels of seriousness.

[6]These parameter values are consistent with those usually obtained in empirical studies. For example, Jiang, Knight and Wang (2005), Shi (2005), Xiao (2007),etc. We also set the parameters as different values, the means of estimates always converge to the true values when the simulation is repeated 10000 times.

2.5.1 Monte Carlo Experiment under the Assumption of No Measurement Error

Our first Monte Carlo experiment is conducted under the assumption that underlying volatility itself is observed, hence there is no measurement error. The purpose of this experiment is to examine the performance of FIML and 3SLS estimation methods.

We simulate daily volatilities and daily asset returns, then use them both in the estimation. For Model 1, FIML (or equivalently iterative FGLS) is applied. For Model 2, we employ both FIML and 3SLS in the estimation.

Table 2.1 reports the results for Model 1. The top panel displays estimation results when the simulation is repeated 1000 times, the middle panel shows results with 5000 simulations, and the bottom panel reports results with 10000 times replication. The third column is the true values of parameters, the fourth to sixth column report the mean, median, and standard deviation of estimates, respectively.

Overall the means, and medians of estimated parameters are very close to true values, the standard deviations for most estimates are relatively small, suggesting that FIML method does a reasonable job for Model 1.

Comparing estimates in all three panels, we find both means and medians of estimates are very close to the true values from all three panels indicating that the estimates converge to the true parameter very quickly. The standard deviations of these estimates are relatively small, stable as replication increases from 1000 to 5000, 10000. The only exception is the estimated risk-return tradeoff parameter λ, we notice the standard deviation is always large, even when the replication increases to 10000 times, suggesting that this estimate is not very stable. In fact, when the simulation is repeated 1000 times, the mean of estimated λ is approximately thirteen percent higher than the true value, and the median is almost fourteen percent higher comparing to the true parameter. As the replication increases, say to 5000 times, the mean along with the median of estimated λ get closer to the true value, being five percent higher than the true value. When the simulation is repeated 10000 times, the difference between mean/median and the true values decreases to four/three percent. This finding suggests that the estimate of λ is not stable, and converges to the true value much slower than the others.

The procedure of estimating Model 2 is not as straightforward as that of estimating Model 1. As we discussed before, the explanatory variable in the asset return equation is correlated with the error term, hence unlike Model 1 which is a non-linear system of equations (SUR), Model 2 in fact is a non-linear simultaneous equation model (SEM). Thus if we apply OLS to the asset return equation as we did for Model 1, the estimator will be not only inconsistent, but also biased. In order to remove the correlation between the explanatory variable and error term, we need to introduce an instrumental variable.

We consider three different choices for the instrument. The natural choice is to use the lagged value of the explanatory variable, i.e. $exp(h_{t-1}/2)$. We denote it as $IV1$. Obviously this is a valid instrument, it is closely correlated to the explanatory variable given the fact that volatility is highly persistent and clustering, while it is not related to the error term. However, as the asset return equation is nonlinear, we are not clear whether this is the optimal instrument. We then consider other instruments. The second choice is to apply the polynomials of h_{t-1}. We call it $IV2$. As Kelejian (1974) showed, when the dependent variable is a nonlinear function of endogenous explanatory variables in a regression model, the instruments could be obtained by regressing the endogenous variable on the elements of a polynomial in the predetermined variable, the approximation would improve as the degree increases. In our study, we regress the endogenous explanatory variable, $exp(h_t/2)$, on a constant, lagged values of h_t and powers of them. The regression model is:

$$y_t = \gamma + \sum_{d=1}^{D} \theta_{1d} h_{t-1}^d + \sum_{d=1}^{D} \theta_{2d} h_{t-2}^d + \sum_{d=1}^{D} \theta_{3d} h_{t-3}^d + \epsilon_t$$

where $y_t = exp(h_t/2)$.

The predicted y_t, $\hat{y}_t = \hat{\gamma} + \sum_{d=1}^{D} \hat{\theta}_{1d} h_{t-1}^d + \sum_{d=1}^{D} \hat{\theta}_{2d} h_{t-2}^d + \sum_{d=1}^{D} \hat{\theta}_{3d} h_{t-3}^d$ is then computed and used as an instrument.

In our experiment, we first set D as 2 then $3, 4, 5, \ldots$, then compare the results. We find there is no significant improvement on the approximation when we increase the power from 5 to 6, 7, etc. We then choose the degree of power up to $D = 5$.

The third choice is a special case of $IV2$. The regression model is:

$$y_t = \gamma + \theta h_{t-1} + \epsilon_t$$

with $y_t = exp(h_t/2)$.

Again $\hat{y}_t$ is obtained and used as an instrument. We call it $IV3$. The procedure to obtain $IV3$ is much simpler than that to obtain $IV2$, later we compare their performance to find whether the simpler procedure is worthy of implementing.

We apply two estimation methods for Model 2. That is, we not only apply FIML but also employ 3SLS method. These two approaches are asymptotically equivalent to each other, while the finite sample properties may not be the same, hence we can compare estimation performance by applying these two different procedures.

Table 2.2-2.4 report the estimation results when the replication M is 1000, 5000, 10000 times, respectively. In all these three tables, the upper panels report FIML results using three different instruments, and the bottom panel reports 3SLS results. The third column displays the true values of the parameters, and the fourth to sixth column report the mean, median, and standard deviation of these estimates.

We find the estimates by applying FIML is very similar to those using 3SLS. Given the fact that FIML and 3SLS are asymptotically equal, we fix the sample size as 3000, and repeat the simulation and estimation at least 1000 times, this finding is not surprising. In FIML estimation, when the simulation parameter M is fixed, the values of means, medians and standard deviations are very similar by using three different kinds of instruments, moreover, they are similar to results using 3SLS approach. This finding suggests that all three kinds of instruments are valid, they work well in removing the relation between endogenous explanatory variable and error term, both estimation methods are reasonable for our models. In general, the standard deviations of estimated parameters are quite small except that of λ indicating that the estimated parameters are quite stable.

We also find as the replication increases, the means, medians of estimates get closer to the true values. For example, the mean of estimated λ is eight percent higher than the true value when $M = 1000$, it is approximately two percent higher when replications increases as 5000 times, and it is only 0.3 percent higher when $M = 10000$. We also find when $M = 1000$, the mean and median of estimated λ are significantly different in FIML, more specifically, the mean value is five percent higher than the median suggesting that there exist some extreme values of estimates by applying FIML. As we increase the replication, the difference starts to decrease. When $M = 10000$, the mean value is only one percent higher than the median. The estimates of other parameters, especially, those for α, β, are close to the true value when $M = 1000$, and do not improve dramatically as M increases because they converge to the true parameters very quickly.

Overall, Monte Carlo experiments suggest that our estimation methods work well for both models when true volatility is observed.

2.5.2 Monte Carlo Experiment Taking into Account Measurement Error

The first Monte Carlo experiment was implemented under perhaps a too strong assumption that volatility is observed and with no error. The main characteristic of the SV models is that volatility is a latent variable. While we can not observe volatility we can assume that we have an available proxy. We now examine the use of a proxy which is subject to varying degrees of measurement error.

In fact, some volatility proxies have been developed in literature. The most popular volatility proxy is realized volatility. Realized volatility is constructed from high frequency intra day squared returns, it is not only model-free, but also a unbiased, consistent estimator of unknown true volatility, and widely used in modeling and forecasting of volatility. In the second Monte Carlo experiment, we make a more realistic assumption, that is, we treat true volatility as an unobservable variable, and use a volatility proxy in estimating of two discrete-time SV models.

Using a volatility proxy, we have to take into account the measurement error. We define measurement error as the difference between the log of the volatility proxy and log of integrated volatility, and assume it follows a normal distribution with zero mean.

$$log(\hat{V_t}) = log(V_t) + u_t$$

where $\hat{V_t}$ represents volatility proxy, $V_t = exp(h_t)$ refers to the true volatility, $u_t \sim N(0, \sigma^{*2})$[7].

The normal distribution assumption on the error term is consistent with empirical evidence of most studies. For example, Xiao (2007) constructed daily realized volatility for the S & P 500 index and three exchange rates, including CAD/USD, USD/GBP, and DEM/USD. Both JB-test and qq-plot showed that the log of realized volatility is approximately normally distributed.

Given the assumption that the vector of error terms in SV models follows a bivariate normal distribution, we first simulate the daily asset return series and true volatility series. Then we take logarithm of true volatility. Next we simulate the measurement error series, then add this series to the logged values of true volatility series to obtain logged values of volatility proxies. Note the measurement error is normally distributed with mean zero, and variance σ^{*2}, which is unknown. We consider setting different values for this parameter,

[7]We use σ^* to distinguish the variance of measurement error to the parameter σ which is a variance of error term η in SV models

in particular, we set $\sigma^{*2} = 1, 0.1, 0.01$, respectively, and simulate daily volatility proxy series based on these different values.

We use both daily asset return and daily volatility proxy observations generated from Model 1 (or Model 2) in the estimation of Model 1 (or Model 2). Both FIML and 3SLS procedures are applied. The estimation results from these two procedures are quite similar. We report the results by using FIML.

Table 2.5 reports the estimation results for Model 1. The simulation and estimation are repeated 10000 times. The top panel shows the estimation results under the assumption that $\sigma^{*2} = 1$. The middle panel reports the estimates with the assumption that the variance of measurement error is 0.1, and the bottom panel displays the results assuming measurement error is normally distributed with mean zero and variance 0.01. The second column displays all five parameter with the third column providing their true values. The fourth to sixth column report mean, median, and standard deviation of the estimated parameters.

We find if the variance of the measurement error is large, say 1, the means, medians of estimated parameters are far different from the true values. For example, the mean of estimated α is -3.3710, more than three times of the true parameter which is -1 in absolute value. The means of other estimates are more than two times of true parameters in absolute values. The values of medians show similar patterns. This finding suggests that in this case using volatility proxy in the estimation of Model 1 is not a reasonable choice. When the size of the measurement error is smaller, that is, $\sigma^{*2} = 0.1$, as the middle panel shows, the estimates improve dramatically, but still significantly different from the true values. For example, the values of mean/median of estimated

α decrease in absolute value and are about fifteen percent higher than the absolute true value. When the variance of measurement error is 0.01, the means together with the medians of the estimates are quite close to the true values. In general, the values of mean/median of estimates are about three percent higher (or lower) than the true parameters in absolute value.

Table 2.6 reports the estimation results for Model 2. We use all three different instruments we introduced before in the estimation, the results are very similar. We report the results using $exp(h_{t-1}/2)$ as an instrument. Also we apply both FIML and 3SSL methods, the results are similar. We report those using FIML. The structure of Table 2.6 is the same as Table 2.5. The simulation and estimation are repeated 10000 times, then the mean, median, and standard deviation are reported.

The results in Table 2.6 show similar patterns as those in Table 2.5. Under the assumption that measurement error is large, i.e. $\sigma^{*2} = 1$, using a volatility proxy and applying a traditional method result in very poor estimates even when we set the sample size as a very large number, and repeat the simulation and estimation 10000 times. However, as the variance of the measurement error decreases, the estimates improve dramatically and when the variance is small enough, say 0.01, the estimates are quite close to the true values.

Clearly the assumption about size or severity of the measurement error, is crucial in determining whether it is appropriate to use a volatility proxy in the estimation of the discrete-time SV models. If the measurement error induced by a volatility proxy is small, then applying traditional estimation methods along with a volatility proxy are able to provide good estimates. The Monte Carlo experiment results suggest that if the variance of measurement error lies approximately between $(0, 0.07)$, FIML and 3SLS work well for both SV models using a volatility proxy.

2.6 Conclusion

In this chapter, we investigated the estimation of discrete-time SV models when volatility is observed. We studied the statistical properties of alternative discrete-time SV model specifications, SV model with lagged inter-temporal dependence and SV model with contemporaneous dependence. We derived the conditional and unconditional lag-lead correlations between returns and volatilities, and showed that the explicit expressions were different between two models. We then considered treating volatility as an observable variable, hence applied traditional estimation methods for both models. We showed that our estimation was straightforward, computationally easy. In order to examine the estimation performance of the traditional methods, including FIML and 3SLS approaches, we implemented two Monte Carlo experiments. The results suggested that if underlying volatility was observed, both FIML and 3SLS were reasonable methods in estimating Model 1 and Model 2. The two models did a good job in recovering data. On the other hand, if the underlying volatility was unobserved, consequently a volatility proxy was used in the estimation, we should be very careful in choosing an appropriate volatility proxy such that the measurement error did not spread too much. Specifically, when the variance of measurement error lies approximately between $(0, 0.07)$, using volatility proxy then applying FIML or 3SLS in SV models were able to provide good estimation performance.

In next chapter, we examine both estimation and forecasting performance of the discrete-time SV models. We employ a popularly used volatility proxy, realized volatility, in both in-sample estimation and out-of-sample forecasting. Our empirical analysis is based on observations of the S & P 500 index and three exchange rates, namely CAD/USD, USD/GBP, and DEM/USD. Both daily and high frequency intra-day data are used. Specifically, we use high frequency intra-day transaction prices to construct daily realized volatility series and use them in the estimation and forecasting. The estimates of Model 1 and Model 2 are reported. The volatility forecasting performance is examined. In particular, four candidate models, including the simple regression model, the SV model with lagged inter-temporal dependence, the SV model with contemporaneous dependence, and a heterogeneous autoregressive (HAR) model are considered. The well known Diebold and Mariano (1995)'s tests are applied for the null hypothesis that two competing models provide equally accurate forecasts. Further the forecasting performance is evaluated using differen criteria.

2.7 Appendix

Appendix 1:

For Model 2, under the assumption that $cov(\varepsilon_t, \upsilon_t) = \rho$, with joint moment generating function of ε_t and υ_t, we have

$$E(exp(w\varepsilon_t + u\upsilon_t)) = M(w,u)$$

where$M(w,u) = exp(\frac{1}{2}(w^2 + u^2 + 2\rho wu))$. Hence

$$E(exp(u\upsilon_t)\varepsilon_t) = \frac{\partial}{\partial\omega}M(w,u)|_{w=0} = \rho u e^{\frac{u^2}{2}}$$

Let $u = \frac{1}{2}\sigma$, we get

$$E(exp(\tfrac{1}{2}\sigma\upsilon_t)\varepsilon_t) = \tfrac{1}{2}\rho\sigma e^{\frac{\sigma^2}{8}}$$

Therefore, given the result above

$$\begin{aligned} E_{t-1}(x_t - \mu) &= E_{t-1}(\lambda e^{h_t} + e^{\frac{1}{2}h_t}\varepsilon_t) \\ &= E_{t-1}(\lambda e^{\alpha+\beta h_{t-1}+\sigma\upsilon_t} + e^{\frac{1}{2}(\alpha+\beta h_{t-1}+\sigma\upsilon_t)}\varepsilon_t) \\ &= \lambda e^{\alpha+\beta h_{t-1}+\frac{1}{2}\sigma^2} + e^{\frac{1}{2}(\alpha+\beta h_{t-1})}E_{t-1}(e^{\frac{1}{2}\sigma\upsilon_t}\varepsilon_t) \\ &= \lambda e^{\alpha+\beta h_{t-1}+\frac{1}{2}\sigma^2} + \frac{1}{2}\rho\sigma e^{\frac{1}{2}\alpha+\frac{1}{2}\beta h_{t-1}+\frac{1}{8}\sigma^2} \end{aligned}$$

Appendix 2:

Proof of **Proposition 1** and **Proposition 2**

For $l > 0, p > 0, i > 0$, we have

$$\begin{aligned} x_t^l x_{t+i}^p &= (\lambda e^{h_t} + e^{\frac{1}{2}h_t}\varepsilon_t)^l (\lambda e^{h_{t+i}} + e^{\frac{1}{2}h_{t+i}}\varepsilon_{t+i})^p \\ &= \sum_{k=0}^{l} \binom{l}{k} (\lambda e^{h_t})^{l-k} (e^{\frac{1}{2}h_t}\varepsilon_t)^k \sum_{s=0}^{p} \binom{p}{s} (\lambda e^{h_{t+i}})^{p-s} (e^{\frac{1}{2}h_{t+i}}\varepsilon_{t+i})^s \\ &= \sum_{k=0}^{l}\sum_{s=0}^{p} \binom{l}{k}\binom{p}{s} \lambda^{l+p-k-s} e^{h_t(l-\frac{1}{2}k)} e^{h_{t+i}(p-\frac{1}{2}s)} \varepsilon_t^k \varepsilon_{t+i}^s \end{aligned}$$

Consider:

$$E_{t-1}(e^{h_t(l-\frac{1}{2}k)} e^{h_{t+i}(p-\frac{1}{2}s)} \varepsilon_t^k \varepsilon_{t+i}^s)$$

Let $a = l - \frac{1}{2}k, b = p - \frac{1}{2}s$

$$E_{t-1}(e^{h_t(l-\frac{1}{2}k)} e^{h_{t+i}(p-\frac{1}{2}s)} \varepsilon_t^k \varepsilon_{t+i}^s)$$

$$\begin{aligned} &= E_{t-1}(e^{ah_t} e^{bh_{t+i}} \varepsilon_t^k \varepsilon_{t+i}^s) \\ &= E_{t-1}(e^{a(\alpha+\beta h_{t-1}+\sigma v_t)} e^{b(\frac{\alpha(1-\beta^{i+1})}{1-\beta}+\beta^{i+1}h_{t-1}+\sigma\sum_{j=0}^{i}\beta^j v_{t+i-j})} \varepsilon_t^k \varepsilon_{t+i}^s) \end{aligned}$$

Following Jiang, Knight and Wang (2005), the iterative functions are as:

$$W_0(u) = 1$$

$$W_1(u) = u$$

$$W_2(u) = u^2 + 1$$

$$W_3(u) = u^3 + 3u$$

$$\vdots$$

$$W_k(u) = (\frac{\partial^k e^{\frac{1}{2}u^2}}{\partial u^k})e^{-\frac{1}{2}u^2}$$

For Model 1, given the assumption $cov(\varepsilon_t, \upsilon_{t+1}) = \rho$, we have the joint moment generating function of ε_t and υ_{t+1}

$$M(w,u) = E(exp(w\varepsilon_t + u\upsilon_{t+1}))$$

where

$$M(w,u) = exp(\tfrac{1}{2}(w^2 + u^2 + 2\rho wu))$$

and

$$E(exp(u\upsilon_{t+1})\varepsilon_t) = \frac{\partial}{\partial w}M(w,u) \mid_{w=0} = \rho u e^{\frac{1}{2}u^2}$$

$$E(exp(u\upsilon_{t+1})\varepsilon_t^2) = \frac{\partial^2}{\partial w^2}M(w,u) \mid_{w=0} = (\rho^2 u^2 + 1)e^{\frac{1}{2}u^2}$$

$$\vdots$$

$$E(exp(u\upsilon_{t+1})\varepsilon_t^k) = \frac{\partial^k}{\partial w^k}M(w,u) \mid_{w=0} = W_k(\rho u)e^{\frac{1}{2}u^2}$$

Hence

$$E_{t-1}(e^{a(\alpha+\beta h_{t-1}+\sigma\upsilon_t)}e^{b(\frac{\alpha(1-\beta^{i+1})}{1-\beta}+\beta^{i+1}h_{t-1}+\sigma\sum_{j=0}^{i}\beta^j\upsilon_{t+i-j})}\varepsilon_t^k\varepsilon_{t+i}^s)$$

$$\begin{aligned}
= \quad & e^{a(\alpha+\beta h_{t-1})+b(\frac{\alpha(1-\beta^{i+1})}{1-\beta}+\beta^{i+1}h_{t-1})}E_{t-1}(e^{(a+b\beta^i)\sigma\upsilon_t+b\sigma\sum_{j=0}^{i-2}\beta^j\upsilon_{t+i-j}}) \\
& \times E_{t-1}(e^{b\sigma\beta^{i-1}}\upsilon_{t+1}\varepsilon_t^k)E_{t-1}(\varepsilon_{t+i}^s) \\
= \quad & e^{a(\alpha+\beta h_{t-1})+b(\frac{\alpha(1-\beta^{i+1})}{1-\beta}+\beta^{i+1}h_{t-1})}e^{\frac{1}{2}\sigma^2(a+b\beta^i)^2+\frac{1}{2}b^2\sigma^2\frac{1-\beta^{2(i-1)}}{1-\beta^2}} \\
& \times W_k(b\rho\sigma\beta^{i-1})e^{\frac{1}{2}b^2\sigma^2\beta^{2(i-1)}}W_s(0) \\
= \quad & e^{a(\alpha+\beta h_{t-1})+b(\frac{\alpha(1-\beta^{i+1})}{1-\beta}+\beta^{i+1}h_{t-1})}e^{\frac{1}{2}\sigma^2(a+b\beta^i)^2+\frac{1}{2}b^2\sigma^2\frac{1-\beta^{2i}}{1-\beta^2}} \\
& \times W_k(b\rho\sigma\beta^{i-1})W_s(0)
\end{aligned}$$

Plug $a = l - \frac{k}{2}, b = p - \frac{s}{2}$ in, we get the result for Model 1.

For Model 2, given the assumptions that $cov(\varepsilon_t, v_t) = \rho, cov(\varepsilon_{t+i}, v_{t+i}) = \rho$,

$$E_{t-1}(e^{a(\alpha+\beta h_{t-1}+\sigma v_t)} e^{b(\frac{\alpha(1-\beta^{i+1})}{1-\beta}+\beta^{i+1}h_{t-1}+\sigma\sum_{j=0}^{i}\beta^j v_{t+i-j})}\varepsilon_t^k \varepsilon_{t+i}^s)$$

$$\begin{aligned}
&= e^{a(\alpha+\beta h_{t-1})+b(\frac{\alpha(1-\beta^{i+1})}{1-\beta}+\beta^{i+1}h_{t-1})} E_{t-1}(e^{b\sigma\sum_{j=1}^{i-1}\beta^j v_{t+i-j}}) E_{t-1}(e^{(a+b\beta^i)\sigma v_t}\varepsilon_t^k) E_{t-1}(e^{b\sigma v_{t+i}}\varepsilon_{t+i}^s) \\
&= e^{a(\alpha+\beta h_{t-1})+b(\frac{\alpha(1-\beta^{i+1})}{1-\beta}+\beta^{i+1}h_{t-1})} e^{\frac{1}{2}b^2\sigma^2\beta^2\frac{1-\beta^{2(i-1)}}{1-\beta^2}} W_k(\rho\sigma(a+b\beta^i)) e^{\frac{1}{2}(a+b\beta^i)^2\sigma^2} W_s(b\rho\sigma) e^{\frac{1}{2}b^2\sigma^2}
\end{aligned}$$

Plug $a = l - \frac{k}{2}, b = p - \frac{s}{2}$ into the equations and add to the previous parts, we get the conditional cross moments for Model 1 and Model 2 in Proposition 1.

For the unconditional cross moments, h_{t-1} would not be treated as observed variable, hence we plug $h_{t-1} = \frac{\alpha}{1-\beta} + \sigma\sum_{j=0}^{\infty}\beta^j v_{t-1-j}$into the formula.

$$E(e^{ah_t+bh_{t+i}}\epsilon_t^k \epsilon_{t+i}^s)$$

$$\begin{aligned}
&= E(e^{a(\alpha+\beta h_{t-1}+\sigma v_t)+b(\alpha\frac{1-\beta^{i+1}}{1-\beta}+\beta^{i+1}h_{t-1}+\sigma\sum_{j=0}^{i}\beta^j v_{t+i-j})}\varepsilon_t^k \varepsilon_{t+i}^s) \\
&= e^{\alpha(a+b\frac{1-\beta^{i+1}}{1-\beta})} E(e^{(a\beta+b\beta^{i+1})h_{t-1}}) E(e^{(a+b\beta^i)\sigma v_t}) E(e^{b\sigma\sum_{j=0}^{i-2}\beta^j v_{t+i-j}}) E(e^{b\sigma\beta^{i-1}v_{t+1}}\varepsilon_t^k) E(\varepsilon_{t+i}^s) \\
&= e^{\alpha\frac{a+b}{1-\beta}} e^{\frac{a^2+2ab\beta^i+b^2}{2(1-\beta^2)}\sigma^2} W_k(b\rho\sigma\beta^{i-1}) W_s(0)
\end{aligned}$$

Plug $a = l - \frac{k}{2}, b = p - \frac{s}{2}$ into the formula, we get the result for Model 1. For Model 2, given $cov(\varepsilon_{t+i}, v_{t+i}) = \rho$, we use similar procedure and get the result.

Appendix 3:

Proof of **Proposition 3**

1. For Model 1:

$$cov(x_t, e^{h_{t-i}} \mid I_{t-i-1}) = 0$$

Since h_{t-i} is observed in this case.

And for Model 2:

$$cov(x_t, e^{h_{t-i}} \mid I_{t-i-1})$$

$$\begin{aligned}
= \quad & \lambda E_{t-i-1}(e^{h_t+h_{t-i}}) + E_{t-i-1}(e^{\frac{1}{2}h_t+h_{t-i}}\varepsilon_t) - \lambda E_{t-i-1}(e^{h_t})E_{t-i-1}(e^{h_{t-i}}) \\
& -E_{t-i-1}(e^{\frac{1}{2}h_t}\varepsilon_t)E_{t-i-1}(e^{h_{t-i}}) \\
= \quad & \lambda E_{t-i-1}(e^{\alpha(1+\frac{1-\beta^{i+1}}{1-\beta})+\beta(1+\beta^i)h_{t-i-1}+\sigma(1+\beta^i)v_{t-i}+\sigma\sum_{j=0}^{i-1}\beta^j v_{t-j}}) \\
& +E_{t-i-1}(e^{\alpha(1+\frac{1-\beta^{i+1}}{2(1-\beta)})+\beta(1+\frac{1}{2}\beta^i)h_{t-i-1}+\sigma(1+\frac{1}{2}\beta^i)v_{t-i}+\frac{1}{2}\sigma\sum_{j=0}^{i-1}\beta^j v_{t-j}+\frac{1}{2}\sigma v_t}\varepsilon_t) \\
& -\lambda E_{t-i-1}(e^{\alpha\frac{1-\beta^{i+1}}{1-\beta}+\beta^{i+1}h_{t-i-1}+\sigma\sum_{j=0}^{i}\beta^j v_{t-j}})E_{t-i-1}(e^{\alpha+\beta h_{t-i-1}+\sigma v_{t-i}}) \\
& -E_{t-i-1}(e^{\frac{1}{2}\alpha\frac{1-\beta^{i+1}}{1-\beta}+\frac{1}{2}\beta^{i+1}h_{t-i-1}+\frac{1}{2}\sigma\sum_{j=0}^{i}\beta^j v_{t-j}}\varepsilon_t)E_{t-i-1}(e^{\alpha+\beta h_{t-i-1}+\sigma v_{t-i}}) \\
= \quad & \lambda e^{\alpha(1+\frac{1-\beta^{i+1}}{1-\beta})+\beta(1+\beta^i)h_{t-i-1}+\frac{2-\beta^2-\beta^{2(i+1)}}{2(1-\beta^2)}\sigma^2}(e^{\beta^i\sigma^2}-1) \\
& +\frac{1}{2}\rho\sigma e^{\alpha(1+\frac{1-\beta^{i+1}}{2(1-\beta)})+\beta(1+\frac{1}{2}\beta^i)h_{t-i-1}+\frac{5-4\beta^2-\beta^{2(i+1)}}{8(1-\beta^2}\sigma^2)}(e^{\frac{1}{2}\beta^i\sigma^2}-1)
\end{aligned}$$

$$var(x_t \mid I_{t-i-1})$$

$$
\begin{aligned}
&= \lambda^2 E_{t-i-1}(e^{2h_t}) + E_{t-i-1}(e^{h_t}\varepsilon_t^2) + 2\lambda E_{t-i-1}(e^{\frac{3}{2}h_t}\varepsilon_t) \\
&\quad -\lambda^2 E_{t-i-1}(e^{h_t})^2 - E_{t-i-1}(e^{\frac{1}{2}h_t}\varepsilon_t)^2 - 2\lambda E_{t-i-1}(e^{h_t})E_{t-i-1}(e^{\frac{1}{2}h_t}\varepsilon_t) \\
&= \lambda^2 e^{2\alpha\frac{1-\beta^{i+1}}{1-\beta}+2\beta^{i+1}h_{t-i-1}+\sigma^2\frac{1-\beta^{2(i+1)}}{1-\beta^2}}(e^{\frac{1-\beta^{2(i+1)}}{1-\beta^2}\sigma^2}-1) \\
&\quad +e^{\alpha\frac{1-\beta^{i+1}}{1-\beta}+\beta^{i+1}h_{t-i-1}+\sigma^2\frac{1-\beta^{2(i+1)}}{4(1-\beta^2)}}(e^{\frac{1-\beta^{2(i+1)}}{4(1-\beta^2)}\sigma^2}(\rho^2\sigma^2+1)-\frac{1}{4}\rho^2\sigma^2) \\
&\quad +\lambda\rho\sigma e^{\alpha\frac{3(1-\beta^{i+1})}{2(1-\beta)}+\frac{3}{2}\beta^{i+1}h_{t-i-1}+\sigma^2\frac{5-5\beta^{2(i+1)}}{8(1-\beta^2)}}(3e^{\frac{1-\beta^{2(i+1)}}{2(1-\beta^2)}\sigma^2}-1) \\
&= V_1
\end{aligned}
$$

$$var(e^{h_{t-i}} \mid I_{t-i-1})$$

$$
\begin{aligned}
&= E_{t-i-1}(e^{2h_{t-i}}) - E_{t-i-1}(e^{h_{t-i}})^2 \\
&= E_{t-i-1}(e^{2\alpha+2\beta h_{t-i-1}+2\sigma^2}) - E_{t-i-1}(e^{\alpha+\beta h_{t-i-1}+\sigma v_{t-i}})^2 \\
&= e^{2\alpha+2\beta h_{t-i-1}+\sigma^2}(e^{\sigma^2}-1)
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\rho_{x_t,e^{h_{t-i}}} \mid I_{t-i-1} = & \frac{1}{\sqrt{V_1(e^{\sigma^2}-1)}}(\lambda e^{\alpha\frac{1-\beta^{i+1}}{1-\beta}+\beta^{i+1}h_{t-i-1}+\frac{1-\beta^{2(i+1)}}{2(1-\beta^2)}\sigma^2}(e^{\beta^i\sigma^2}-1) \\
& +\frac{1}{2}\rho\sigma e^{\alpha\frac{1-\beta^{i+1}}{2(1-\beta)}+\frac{1}{2}\beta^{i+1}h_{t-i-1}+\frac{1-\beta^{2(i+1)}}{8(1-\beta^2)}\sigma^2}(e^{\frac{1}{2}\beta^i\sigma^2}-1))
\end{aligned}
$$

2. For Model 1:

$$cov(x_t, e^{h_{t+i}} \mid I_{t-1})$$

$$
\begin{aligned}
&= E_{t-1}((x_t - E_{t-1}x_t)(e^{h_{t+i}} - E_{t-1}(e^{h_{t+i}}))) \\
&= E_{t-1}(e^{\frac{1}{2}h_t+h_{t+i}}\varepsilon_t - e^{\frac{1}{2}h_t}\varepsilon_t E_{t-1}(e^{h_{t+i}})) \\
&= E_{t-1}(e^{\frac{1}{2}h_t+h_{t+i}}\varepsilon_t) \\
&= E_{t-1}(e^{\alpha\frac{1-\beta^i}{1-\beta}+(1/2+\beta^i)h_t+\sigma\sum_{j=0}^{i-2}\beta^j v_{t+i-j}}e^{\sigma\beta^{i-1}v_{t+1}}\varepsilon_t) \\
&= \rho\sigma\beta^{i-1}e^{\alpha\frac{1-\beta^i}{1-\beta}+(1/2+\beta^i)h_t+\frac{1}{2}\sigma^2\frac{1-\beta^{2i}}{1-\beta^2}}
\end{aligned}
$$

$$var(x_t \mid I_{t-1})$$

$$\begin{aligned} &= E_{t-1}((x_t - E_{t-1}(x_t))^2) \\ &= E_{t-1}(e^{h_t}\varepsilon_t^2) \\ &= e^{h_t} \end{aligned}$$

$$var(e^{h_{t+i}} \mid I_{t-1})$$

$$\begin{aligned} &= E_{t-1}(e^{2h_{t+i}}) - E_{t-1}(e^{h_{t+i}})^2 \\ &= e^{2\alpha\frac{1-\beta^i}{1-\beta}+2\beta^i h_t+\sigma^2\frac{1-\beta^{2i}}{1-\beta^2}}(e^{\sigma^2\frac{1-\beta^{2i}}{1-\beta^2}} - 1) \end{aligned}$$

Therefore

$$\rho_{x_t,e^{h_{t+i}}} \mid I_{t-1} = \frac{\rho\sigma\beta^{i-1}}{\sqrt{(e^{\sigma^2\frac{1-\beta^{2i}}{1-\beta^2}} - 1)}}$$

And for Model 2:

$$cov(x_t, e^{h_{t+i}} \mid I_{t-1})$$

$$\begin{aligned} &= \lambda E_{t-1}(e^{h_t+h_{t+i}}) + E_{t-1}(e^{\frac{1}{2}h_t+h_{t+i}}\varepsilon_t) - \lambda E_{t-1}(e^{h_t})E_{t-1}(e^{h_{t+i}}) \\ &\quad - E_{t-1}(e^{\frac{1}{2}h_t}\varepsilon_t)E_{t-1}(e^{h_{t+i}}) \\ &= \lambda E_{t-1}(e^{\alpha(1+\frac{1-\beta^{i+1}}{1-\beta})+\beta(1+\beta^i)h_{t-1}+\sigma(1+\beta^i)\upsilon_t+\sigma\sum_{j=0}^{i-1}\beta^j\upsilon_{t+i-j}}) \\ &\quad + E_{t-1}(e^{\alpha(\frac{1}{2}+\frac{1-\beta^{i+1}}{1-\beta})+\beta(\frac{1}{2}+\beta^i)h_{t-1}+\sigma(\frac{1}{2}+\beta^i)\upsilon_t+\sigma\sum_{j=0}^{i-1}\beta^j\upsilon_{t+i-j}}\varepsilon_t) \\ &\quad - \lambda E_{t-1}(e^{\alpha+\beta h_{t-1}+\sigma\upsilon_t})E_{t-1}(e^{\alpha\frac{1-\beta^{i+1}}{1-\beta}+\beta^{i+1}h_{t-1}+\sigma\sum_{j=0}^{i}\beta^j\upsilon_{t+i-j}}) \\ &\quad - E_{t-1}(e^{\alpha/2+\beta h_{t-1}/2+\sigma\upsilon_t/2}\varepsilon_t)E_{t-1}(e^{\alpha\frac{1-\beta^{i+1}}{1-\beta}+\beta^{i+1}h_{t-1}+\sigma\sum_{j=0}^{i}\beta^j\upsilon_{t+i-j}}) \\ &= \lambda e^{\alpha(1+\frac{1-\beta^{i+1}}{1-\beta})+\beta(1+\beta^i)h_{t-1}+\frac{2-\beta^2-\beta^{2(i+1)}}{2(1-\beta^2)}\sigma^2}(e^{\beta^i\sigma^2} - 1) \\ &\quad + \frac{1}{2}\rho\sigma e^{\alpha(\frac{1}{2}+\frac{1-\beta^{i+1}}{1-\beta})+\beta(\frac{1}{2}+\beta^i)h_{t-1}+\frac{5-\beta^2-4\beta^{2(i+1)}}{8(1-\beta^2)}\sigma^2}((1+2\beta^i)e^{\beta^i\sigma^2/2} - 1) \end{aligned}$$

$$var(x_t \mid I_{t-1})$$

$$\begin{aligned}
&= \lambda^2 E_{t-1}(e^{2h_t}) + E_{t-1}(e^{h_t}\varepsilon_t^2) + 2\lambda E_{t-1}(e^{3h_t/2}\varepsilon_t) \\
&\quad -\lambda^2 E_{t-1}(e^{h_t})^2 - E_{t-1}(e^{h_t/2}\varepsilon_t)^2 - 2\lambda E_{t-1}(e^{h_t})E_{t-1}(e^{h_t/2}\varepsilon_t) \\
&= \lambda^2 e^{2\alpha+2\beta h_{t-1}+\sigma^2}(e^{\sigma^2}-1) + e^{\alpha+\beta h_{t-1}+\sigma^2/4}((\rho^2\sigma^2+1)e^{\sigma^2/4} - \rho^2\sigma^2/4) \\
&\quad +\lambda\rho\sigma e^{3\alpha/2+3\beta h_{t-1}/2+5\sigma^2/8}(3e^{\sigma^2/2}-1) \\
&= V_2
\end{aligned}$$

$$var(e^{h_{t+i}} \mid I_{t-1})$$

$$\begin{aligned}
&= E_{t-1}(e^{2h_{t+i}}) - E_{t-1}(e^{h_{t+i}})^2 \\
&= e^{2\alpha\frac{1-\beta^{i+1}}{1-\beta}+2\beta^{i+1}h_{t-1}+\frac{1-\beta^{2(i+1)}}{1-\beta^2}\sigma^2}(e^{\frac{1-\beta^{2(i+1)}}{1-\beta^2}\sigma^2}-1)
\end{aligned}$$

Hence

$$\rho_{X_t,e^{h_{t+i}}} \mid I_{t-1} = \frac{\lambda e^{\alpha+\beta h_{t-1}+\sigma^2/2}(e^{\beta^i\sigma^2}-1) + \frac{1}{2}\rho\sigma e^{\alpha/2+\beta h_{t-1}/2+\sigma^2/8}((1+2\beta^i)e^{\frac{1}{2}\beta^i\sigma^2}-1)}{\sqrt{V_2(e^{\sigma^2\frac{1-\beta^{2(i+1)}}{1-\beta^2}}-1)}}$$

Proof of **Proposition 4**

1. **For Model 1:**

   for $i \geqslant 0$

$$cov(x_t, e^{h_{t-i}})$$

$$\begin{aligned}
&= \lambda E(e^{h_t+h_{t-i}}) + E(e^{\frac{1}{2}h_t+h_{t-i}}\varepsilon_t) - \lambda E(e^{h_t})E(e^{h_{t-i}}) - E(e^{\frac{1}{2}h_t}\varepsilon_t)E(e^{h_{t-i}}) \\
&= \lambda E(e^{\alpha(\frac{1-\beta^i}{1-\beta})+(1+\beta^i)h_{t-i}+\sigma\sum_{j=0}^{i-1}\beta^j v_{t-j}}) - \lambda E(e^{h_t})E(e^{h_{t-i}}) \\
&= \lambda e^{\alpha(\frac{1-\beta^i}{1-\beta})+(1+\beta^i)\frac{\alpha}{1-\beta}+\frac{1+\beta^i}{1-\beta^2}\sigma^2} - \lambda e^{\alpha(\frac{2}{1-\beta})+\frac{1}{1-\beta^2}\sigma^2} \\
&= \lambda e^{\frac{2\alpha}{1-\beta}+\frac{\sigma^2}{1-\beta^2}}(e^{\frac{\beta^i\sigma^2}{1-\beta^2}}-1)
\end{aligned}$$

$$var(x_t)$$

$$\begin{aligned}
&= E(x_t^2) - E(x_t)^2 \\
&= \lambda^2 E(e^{2h_t}) + E(e^{h_t}\varepsilon_t^2) - \lambda^2 E(e^{h_t})^2 \\
&= \lambda^2 e^{\frac{2\alpha}{1-\beta}+\frac{2\sigma^2}{1-\beta^2}} + e^{\frac{\alpha}{1-\beta}+\frac{\sigma^2}{2(1-\beta^2)}} - \lambda^2 e^{\frac{2\alpha}{1-\beta}+\frac{\sigma^2}{1-\beta^2}} \\
&= V_3
\end{aligned}$$

$$var(e^{h_{t-i}})$$

$$\begin{aligned}
&= E(e^{2h_{t-i}}) - E(e^{h_{t-i}})^2 \\
&= e^{\frac{2\alpha}{1-\beta}+\frac{\sigma^2}{1-\beta^2}}(e^{\frac{\sigma^2}{1-\beta^2}} - 1)
\end{aligned}$$

Therefore

$$\rho_{x_t, e^{h_{t-i}}} = \frac{\lambda e^{\frac{\alpha}{1-\beta}+\frac{\sigma^2}{2(1-\beta^2)}}(e^{\frac{\beta^i\sigma^2}{1-\beta^2}} - 1)}{\sqrt{V_3(e^{\frac{\sigma^2}{1-\beta^2}} - 1)}}$$

And for Model 2:

for $i \geqslant 0$

$$cov(x_t, e^{h_{t-i}})$$

$$\begin{aligned}
&= \lambda E(e^{h_t+h_{t-i}}) + E(e^{\frac{1}{2}h_t+h_{t-i}}\varepsilon_t) - \lambda E(e^{h_t})E(e^{h_{t-i}}) - E(e^{\frac{1}{2}h_t}\varepsilon_t)E(e^{h_{t-i}}) \\
&= \lambda E(e^{\alpha(\frac{1-\beta^i}{1-\beta})+(1+\beta^i)h_{t-i}+\sigma\sum_{j=0}^{i-1}\beta^j\upsilon_{t-j}}) + E(e^{\frac{\alpha(1-\beta^i)}{2(1-\beta)}+(1+\beta^i/2)h_{t-i}+\sigma\upsilon_t/2+\frac{1}{2}\sigma\sum_{j=1}^{i-1}\beta^j\upsilon_{t-j}}\varepsilon_t) \\
&\quad -\lambda E(e^{h_t})E(e^{h_{t-i}}) - E(e^{\alpha/2+\beta h_{t-1}/2+\sigma\upsilon_t/2}\varepsilon_t)E(e^{h_{t-i}}) \\
&= \lambda e^{(\frac{2\alpha}{1-\beta})+\frac{\sigma^2(1+\beta^i)}{1-\beta^2}} + \frac{1}{2}\rho\sigma e^{\frac{3\alpha}{2(1-\beta)}+\frac{5+4\beta^i}{8(1-\beta^2)}\sigma^2} - \lambda e^{\frac{2\alpha}{1-\beta}+\frac{1}{1-\beta^2}\sigma^2} - \frac{1}{2}\rho\sigma e^{\frac{3\alpha}{2(1-\beta)}+\frac{5}{8(1-\beta^2)}\sigma^2} \\
&= \lambda e^{\frac{2\alpha}{1-\beta}+\frac{\sigma^2}{1-\beta^2}}(e^{\frac{\beta^i\sigma^2}{1-\beta^2}} - 1) + \frac{1}{2}\rho\sigma e^{\frac{3\alpha}{2(1-\beta)}+\frac{5}{8(1-\beta^2)}\sigma^2}(e^{\frac{\beta^i\sigma^2}{2(1-\beta^2)}} - 1)
\end{aligned}$$

$$var(x_t)$$

$$\begin{aligned}
&= E(x_t^2) - E(x_t)^2 \\
&= \lambda^2 E(e^{2h_t}) + E(e^{h_t}\varepsilon_t^2) + 2\lambda E(e^{3h_t/2}\varepsilon_t) \\
&\quad -\lambda^2 E(e^{h_t})^2 - E(e^{h_t/2}\varepsilon_t)^2 - 2\lambda E(e^{h_t})E(e^{h_t/2}\varepsilon_t) \\
&= \lambda^2 e^{\frac{2\alpha}{1-\beta}+\frac{2\sigma^2}{1-\beta^2}} + e^{\frac{\alpha}{1-\beta}+\frac{\sigma^2\beta^2}{2(1-\beta^2)}}(\rho^2\sigma^2+1)e^{\sigma^2/2} \\
&\quad +3\lambda\rho\sigma e^{\frac{3\alpha}{2(1-\beta)}+\frac{9\beta^2\sigma^2}{8(1-\beta^2)}+\frac{9\sigma^2}{8}} - \lambda^2 e^{\frac{2\alpha}{1-\beta}+\frac{\sigma^2}{1-\beta^2}} \\
&\quad -e^{\frac{\alpha}{1-\beta}+\frac{\beta^2\sigma^2}{4(1-\beta^2)}}(\frac{1}{4}\rho^2\sigma^2)e^{\frac{1}{4}\sigma^2} - \lambda\rho\sigma e^{\frac{3\alpha}{1-\beta}+\frac{\sigma^2}{2(1-\beta^2)}+\frac{\beta^2\sigma^2}{8(1-\beta^2)}+\sigma^2/8} \\
&= V_4
\end{aligned}$$

$$var(e^{h_{t-i}})$$

$$\begin{aligned}
&= E(e^{2h_{t-i}}) - E(e^{h_{t-i}})^2 \\
&= e^{\frac{2\alpha}{1-\beta}+\frac{\sigma^2}{1-\beta^2}}(e^{\frac{\sigma^2}{1-\beta^2}} - 1)
\end{aligned}$$

Hence

$$\rho_{x_t,e^{h_{t-i}}} = \frac{\lambda e^{\frac{\alpha}{1-\beta}+\frac{\sigma^2}{2(1-\beta^2)}}(e^{\frac{\beta^i\sigma^2}{1-\beta^2}} - 1) + \frac{1}{2}\rho\sigma e^{\frac{\alpha}{2(1-\beta)}+\frac{\sigma^2}{8(1-\beta^2)}}(e^{\frac{\beta^i\sigma^2}{2(1-\beta^2)}} - 1)}{\sqrt{V_4(e^{\frac{\sigma^2}{1-\beta^2}} - 1)}}$$

2. For Model 1:

$$cov(x_t, e^{h_{t+i}})$$

$$
\begin{aligned}
&= \lambda E(e^{h_t+h_{t+i}}) + E(e^{\frac{1}{2}h_t+h_{t+i}}\varepsilon_t) - \lambda E(e^{h_t})E(e^{h_{t+i}}) - E(e^{\frac{1}{2}h_t}\varepsilon_t)E(e^{h_{t+i}}) \\
&= \lambda E(e^{\alpha(\frac{1-\beta^i}{1-\beta})+(1+\beta^i)h_t+\sigma\sum_{j=0}^{i-1}\beta^j v_{t+i-j}}) + E(e^{\alpha(\frac{1-\beta^i}{1-\beta})+(1/2+\beta^i)h_t+\sigma\sum_{j=0}^{i-1}\beta^j v_{t+i-j}}\varepsilon_t) \\
&\quad -\lambda E(e^{h_t})E(e^{h_{t+i}}) \\
&= \lambda e^{\frac{2\alpha}{1-\beta}+\frac{1+\beta^i}{1-\beta^2}\sigma^2} + \rho\sigma\beta^{i-1}e^{\frac{3\alpha}{2(1-\beta)}+\frac{5+4\beta^i}{8(1-\beta^2)}\sigma^2} - \lambda e^{\frac{2\alpha}{1-\beta}+\frac{1}{1-\beta^2}\sigma^2} \\
&= \lambda e^{\frac{2\alpha}{1-\beta}+\frac{1}{1-\beta^2}\sigma^2}(e^{\frac{\beta^i\sigma^2}{1-\beta^2}}-1) + \rho\sigma\beta^{i-1}e^{\frac{3\alpha}{2(1-\beta)}+\frac{5+4\beta^i}{8(1-\beta^2)}\sigma^2}
\end{aligned}
$$

$$var(x_t) = V_3$$

$$var(e^{h_{t+i}}) = var(e^{h_{t-i}}) = e^{\frac{2\alpha}{1-\beta}+\frac{1}{1-\beta^2}\sigma^2}(e^{\frac{\sigma^2}{1-\beta^2}}-1)$$

Hence

$$\rho_{x_t,e^{h_{t+i}}} = \frac{\lambda e^{\frac{\alpha}{1-\beta}+\frac{\sigma^2}{2(1-\beta^2)}}(e^{\frac{\beta^i\sigma^2}{1-\beta^2}}-1) + \rho\sigma\beta^{i-1}e^{\frac{\alpha}{2(1-\beta)}+\frac{\sigma^2(1+4\beta^i)}{8(1-\beta^2)}}}{\sqrt{V_3(e^{\frac{\sigma^2}{1-\beta^2}}-1)}}$$

For Model 2:

$$cov(x_t, e^{h_{t+i}})$$

$$
\begin{aligned}
&= \lambda E(e^{h_t+h_{t+i}}) + E(e^{\frac{1}{2}h_t+h_{t+i}}\varepsilon_t) - \lambda E(e^{h_t})E(e^{h_{t+i}}) - E(e^{\frac{1}{2}h_t}\varepsilon_t)E(e^{h_{t+i}}) \\
&= \lambda E(e^{\alpha(\frac{1-\beta^i}{1-\beta})+(1+\beta^i)h_t+\sigma\sum_{j=0}^{i-1}\beta^j v_{t+i-j}}) \\
&\quad +E(e^{\alpha(\frac{1-\beta^i}{1-\beta})+(1/2+\beta^i)(\alpha+\beta h_{t-1}+\sigma v_t)+\sigma\sum_{j=1}^{i-1}\beta^j v_{t+i-j}}\varepsilon_t) \\
&\quad -\lambda E(e^{h_t})E(e^{h_{t+i}}) - E(e^{\frac{1}{2}\alpha+\frac{1}{2}\beta h_{t-1}+\frac{1}{2}\sigma v_t}\varepsilon_t)E(e^{h_{t+i}}) \\
&= \lambda e^{\frac{2\alpha}{1-\beta}+\frac{1+\beta^i}{1-\beta^2}\sigma^2} + \frac{1}{2}\rho\sigma(1+2\beta^i)e^{\frac{3\alpha}{2(1-\beta)}+\frac{5+4\beta^i}{8(1-\beta^2)}\sigma^2} \\
&\quad -\lambda e^{\frac{2\alpha}{1-\beta}+\frac{1}{1-\beta^2}\sigma^2} - \frac{1}{2}\rho\sigma e^{\frac{3\alpha}{2(1-\beta)}+\frac{5}{8(1-\beta^2)}\sigma^2} \\
&= \lambda e^{\frac{2\alpha}{1-\beta}+\frac{1}{1-\beta^2}\sigma^2}(e^{\frac{\beta^i\sigma^2}{1-\beta^2}}-1) + \frac{1}{2}\rho\sigma e^{\frac{3\alpha}{2(1-\beta)}+\frac{5}{8(1-\beta^2)}\sigma^2}((1+2\beta^i)e^{\frac{\beta^i\sigma^2}{2(1-\beta^2)}}-1)
\end{aligned}
$$

$$var(x_t) = V_4$$

$$var(e^{h_{t+i}}) = var(e^{h_{t-i}}) = e^{\frac{2\alpha}{1-\beta}+\frac{1}{1-\beta^2}\sigma^2}(e^{\frac{\sigma^2}{1-\beta^2}} - 1)$$

Hence

$$\rho_{x_t,e^{h_{t+i}}} = \frac{\lambda e^{\frac{\alpha}{1-\beta}+\frac{\sigma^2}{2(1-\beta^2)}}(e^{\frac{\beta^i\sigma^2}{1-\beta^2}} - 1) + \frac{1}{2}\rho\sigma e^{\frac{\alpha}{2(1-\beta)}+\frac{\sigma^2}{8(1-\beta^2)}}((1+2\beta^i)e^{\frac{\beta^i\sigma^2}{2(1-\beta^2)}} - 1)}{\sqrt{V_4(e^{\frac{\sigma^2}{1-\beta^2}} - 1)}}$$

Appendix 4:

Full Information Maximum Likelihood as Iterated SUR:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \lambda \\ \pi \end{pmatrix} + \begin{pmatrix} \varepsilon \\ \eta \end{pmatrix}$$

where $\pi = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$.

GLS gives:

$$\begin{pmatrix} \hat{\lambda} \\ \hat{\pi} \end{pmatrix}$$

$$= (\begin{pmatrix} X_1' & 0 \\ 0 & X_2' \end{pmatrix} \begin{pmatrix} I & \rho\sigma I \\ \rho\sigma I & \sigma^2 I \end{pmatrix}^{-1} \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix})^{-1}$$

$$\begin{pmatrix} X_1' & 0 \\ 0 & X_2' \end{pmatrix} \begin{pmatrix} I & \rho\sigma I \\ \rho\sigma I & \sigma^2 I \end{pmatrix}^{-1} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

$$= (\begin{pmatrix} w^{11}X_1'X_1 & w^{12}X_1'X_2 \\ w^{21}X_2'X_1 & w^{22}X_2'X_2 \end{pmatrix})^{-1} \begin{pmatrix} w^{11}X_1'Y_1 + w^{12}X_1'Y_2 \\ w^{21}X_2'Y_1 + w^{22}X_2'Y_2 \end{pmatrix}$$

Let $A = (\begin{pmatrix} w^{11}X_1'X_1 & w^{12}X_1'X_2 \\ w^{21}X_2'X_1 & w^{22}X_2'X_2 \end{pmatrix})^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} w^{11}X_1'X_1 & w^{12}X_1'X_2 \\ w^{21}X_2'X_1 & w^{22}X_2'X_2 \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}$$

Equivalently,we have four equations

$$A_{11}w^{11}X_1'X_1 + A_{12}w^{21}X_2'X_1 = I$$

$$A_{11}w^{12}X_1'X_2 + A_{12}w^{22}X_2'X_2 = 0$$

$$A_{21}w^{11}X_1'X_1 + A_{22}w^{21}X_2'X_1 = 0$$

$$A_{21}w^{12}X_1'X_2 + A_{22}w^{22}X_2'X_2 = I$$

From the second equation,we derive

$A_{12} = -A_{11}w^{12}X_1'X_2(w^{22}X_2'X_2)^{-1}$.

Substitute into equation one,

$A_{11} = (w^{11}X_1'X_1 - \frac{w^{12}w^{21}}{w^{22}}X_1'X_2(X_2'X_2)^{-1}X_2'X_1)^{-1}$

And then

$A_{12} = -(w^{11}X_1'X_1 - \frac{w^{12}w^{21}}{w^{22}}X_1'X_2(X_2'X_2)^{-1}X_2'X_1)^{-1}\frac{w^{12}}{w^{22}}X_1'X_2(X_2'X_2)^{-1}$

Similarly,

$A_{21} = (w^{12}X_1'X_2 - \frac{w^{11}w^{22}}{w^{21}}X_1'X_1(X_2'X_1)^{-1}X_2'X_2)^{-1}$

$A_{22} = -(w^{12}X_1'X_2 - \frac{w^{11}w^{22}}{w^{21}}X_1'X_1(X_2'X_1)^{-1}X_2'X_2)^{-1}\frac{w^{11}}{w^{21}}X_1'X_1(X_2'X_1)^{-1}$

Obviously,

$w^{11} = \frac{1}{1-\rho^2}, w^{12} = w^{21} = -\frac{\rho}{\sigma(1-\rho^2)}, w^{22} = \frac{1}{\sigma^2(1-\rho^2)}$ and $\frac{w^{12}w^{21}}{w^{22}} = \frac{\rho^2}{1-\rho^2}$.

As a result:

$$\hat{\lambda}$$

$$\begin{aligned}
&= A_{11}(w^{11}X_1'Y_1 + w^{12}X_1'Y_2) + A_{12}(w^{21}X_2'Y_1 + w^{22}X_2'Y_2) \\
&= (w^{11}X_1'X_1 - \frac{w^{12}w^{21}}{w^{22}}X_1'X_2(X_2'X_2)^{-1}X_2'X_1)^{-1} \\
&*(w^{11}X_1'Y_1 + w^{12}X_1'Y_2 - \frac{w^{12}}{w^{22}}X_1'X_2(X_2'X_2)^{-1}(w^{21}X_2'Y_1 + w^{22}X_2'Y_2)) \\
&= (w^{11}X_1'X_1 - \frac{w^{12}w^{21}}{w^{22}}X_1'P_{X_2}X_1)^{-1} \\
&*(w^{11}X_1'Y_1 - \frac{w^{12}w^{21}}{w^{22}}X_1'P_{X_2}Y_1 + w^{12}(X_1'Y_2 - X_1'P_{X_2}Y_2)) \\
&= (X_1'(I-\rho^2P_{X_2})X_1)^{-1}X_1'(I-\rho^2P_{X_2})Y_1 - \frac{\rho}{\sigma}(X_1'(I-\rho^2P_{X_2})X_1)^{-1}X_1'(I-\rho^2P_{X_2})Y_2
\end{aligned}$$

and

$$\hat{\pi}$$

$$
\begin{aligned}
&= (w^{22}X_2'X_2)^{-1}w^{22}X_2'Y_2 + (w^{22}X_2'X_2)^{-1}w^{21}X_2'Y_1 - (w^{22}X_2'X_2)^{-1}w^{21}X_2'X_1\hat{\lambda} \\
&= (X_2'X_2)^{-1}X_2'Y_2 + \frac{w^{21}}{w^{22}}(X_2'X_2)^{-1}X_2'(Y_1 - X_1\hat{\lambda}) \\
&= (X_2'X_2)^{-1}X_2'Y_2 - \rho\sigma(X_2'X_2)^{-1}X_2'(Y_1 - X_1\hat{\lambda})
\end{aligned}
$$

Three-Stage Least Squares as Iterated SUR:

Multiply the set of predetermined variables to both equations:

$$\begin{pmatrix} X_2'Y_1 \\ X_2'Y_2 \end{pmatrix} = \begin{pmatrix} X_2'X_1 & 0 \\ 0 & X_2'X_2 \end{pmatrix} \begin{pmatrix} \lambda \\ \pi \end{pmatrix} + \begin{pmatrix} X_2'\varepsilon \\ X_2'\eta \end{pmatrix}$$

The joint distribution of error terms is:

$$\begin{pmatrix} X_2' & 0 \\ 0 & X_2' \end{pmatrix} \begin{pmatrix} \varepsilon \\ \eta \end{pmatrix} \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \bigotimes (X_2'X_2))$$

GLS gives:

$$\begin{pmatrix} \hat{\lambda} \\ \hat{\pi} \end{pmatrix}$$

$$= (\begin{pmatrix} X_1'X_2 & 0 \\ 0 & X_2'X_2 \end{pmatrix} \begin{pmatrix} w^{11}(X_2'X_2)^{-1} & w^{12}(X_2'X_2)^{-1} \\ w^{21}(X_2'X_2)^{-1} & w^{22}(X_2'X_2)^{-1} \end{pmatrix} \begin{pmatrix} X_2'X_1 & 0 \\ 0 & X_2'X_2 \end{pmatrix})^{-1}$$

$$* \begin{pmatrix} X_1'X_2 & 0 \\ 0 & X_2'X_2 \end{pmatrix} \begin{pmatrix} w^{11}(X_2'X_2)^{-1} & w^{12}(X_2'X_2)^{-1} \\ w^{21}(X_2'X_2)^{-1} & w^{22}(X_2'X_2)^{-1} \end{pmatrix} \begin{pmatrix} X_2'Y_1 \\ X_2Y_2 \end{pmatrix}$$

$$= (\begin{pmatrix} w^{11}X_1'X_2(X_2'X_2)^{-1}X_2'X_1 & w^{12}X_1'X_2 \\ w^{21}X_2'X_1 & w^{22}X_2'X_2 \end{pmatrix})^{-1}$$

$$* \begin{pmatrix} w^{11}X_1'X_2(X_2'X_2)^{-1}X_2'Y_1 + w^{12}X_1'X_2(X_2'X_2)^{-1}X_2'Y_2 \\ w^{21}X_2'Y_1 + w^{22}X_2'Y_2 \end{pmatrix}$$

Simplify the inverse matrix,

$$\hat{\lambda}_{2SLS}$$

$$= (w^{11} - \frac{w^{12}w^{21}}{w^{22}})^{-1}(X_1'P_{X_2}X_1)^{-1}(w^{11}X_1'P_{X_2}Y_1 + w^{12}X_1'P_{X_2}Y_2)$$
$$-(w^{11} - \frac{w^{12}w^{21}}{w^{22}})^{-1}(X_1'P_{X_2}X_1)^{-1}\frac{w^{12}}{w^{22}}X_1'X_2(X_2'X_2)^{-1}(w^{21}X_2'Y_1 + w^{22}X_2'Y_2)$$
$$= (w^{11} - \frac{w^{12}w^{21}}{w^{22}})^{-1}(X_1'P_{X_2}X_1)^{-1}$$
$$*(w^{11}X_1'P_{X_2}Y_1 + w^{12}X_1'P_{X_2}Y_2 - \frac{w^{12}w^{21}}{w^{22}}X_1'P_{X_2}Y_1 - w^{12}X_1'P_{X_2}Y_2)$$
$$= (X_1'X_2(X_2X_2)^{-1}X_2X_1)^{-1}X_1'X_2(X_2'X_2)^{-1}X_2'Y_1$$

and

$$\hat{\pi}_{2SLS} = (X_2'X_2)^{-1}X_2'Y_2 - \rho\sigma(X_2X_2)^{-1}X_2'(Y_1 - X_1\hat{\lambda})$$

Appendix 5:

Computation of GLS standard errors:

Computation of standard errors of $\hat{\lambda}, \hat{\alpha}$ and $\hat{\beta}$ are straightforward.

We apply Delta method to calculate standard errors of $\hat{\rho}$ and $\hat{\sigma}$.

$$\hat{\rho} = \frac{cov(\hat{\varepsilon}, \hat{\eta})}{var(\hat{\eta})} = \frac{E(\hat{\varepsilon}'\hat{\eta}) - E(\hat{\varepsilon})'E(\hat{\eta})}{\sqrt{E(\hat{\eta}'\hat{\eta}) - E(\hat{\eta})'E(\hat{\eta})}} = \frac{(Y_1 - X_1\hat{\lambda})'(Y_2 - X_2\hat{\pi})/T}{\sqrt{(Y_2 - X_2\hat{\pi})'(Y_2 - X_2\hat{\pi})/T}}$$

Let $\hat{\rho} = f(\hat{\lambda}, \hat{\pi}) = f(\hat{\delta})$.

Using Delta method

$$\hat{\rho} = f(\hat{\delta}) = f(\delta_0) + \frac{\partial f}{\partial \hat{\delta}'}(\hat{\delta} - \delta_0) = \rho_0 + \hat{F} \times (\hat{\delta} - \delta_0)$$

$$\hat{\rho} - \rho_0 = \hat{F} \times (\hat{\delta} - \delta_0) \tag{2.6}$$

$$E(\hat{\rho} - \rho_0)^2 = \hat{F}E(\hat{\delta} - \delta_0)(\hat{\delta} - \delta_0)'\hat{F}' = \hat{F}var(\hat{\delta})\hat{F}'$$

Similar procedure can be applied for the calculation of standard error of $\hat{\sigma}$.

Appendix 6:

Table 2.1: Monte Carlo Simulation: Estimation of Model 1

FIML Estimation of Model 1					
DGP	**Parameter**	**True Value**	**Mean**	**Median**	**St. Dev.**
M=1000	λ	0.10	0.1133	0.1138	0.1855
	α	−1.00	−1.0023	−1.0003	0.0488
	β	0.80	0.7997	0.8003	0.0094
	ρ	−0.50	−0.4998	−0.5002	0.0174
	σ	0.50	0.4999	0.4999	0.0066
M=5000	λ	0.10	0.1054	0.1047	0.1808
	α	−1.00	−1.0040	−1.0030	0.0498
	β	0.80	0.7992	0.7995	0.0096
	ρ	−0.50	−0.4998	−0.4997	0.0172
	σ	0.50	0.5000	0.5000	0.0065
M=10000	λ	0.10	0.1037	0.1025	0.1798
	α	−1.00	−1.0039	−1.0030	0.0497
	β	0.80	0.7992	0.7994	0.0096
	ρ	−0.50	−0.4997	−0.4998	0.0171
	σ	0.50	0.4999	0.4998	0.0065

Table 2.2: Monte Carlo Simulation: Estimation of Model 2(M=1000)

DGP	Parameter	True Value	Mean	Median	St. Dev.
FIML Estimation of Model 2					
IV1	λ	0.10	0.1080	0.1028	0.1939
	α	-1.00	-1.0053	-1.0053	0.0487
	β	0.80	0.7990	0.7990	0.0094
	ρ	-0.50	-0.5005	-0.5008	0.0174
	σ	0.50	0.4996	0.4995	0.0065
IV2	λ	0.10	0.1080	0.1028	0.1939
	α	-1.00	-1.0053	-1.0053	0.0487
	β	0.80	0.7990	0.7990	0.0094
	ρ	-0.50	-0.5005	-0.5008	0.0174
	σ	0.50	0.4996	0.4995	0.0067
IV3	λ	0.10	0.1080	0.1028	0.1939
	α	-1.00	-1.0053	-1.0053	0.0487
	β	0.80	0.7990	0.7990	0.0094
	ρ	-0.50	-0.5005	-0.5008	0.0174
	σ	0.50	0.4996	0.4995	0.0067
3SLS Estimation of Model 2					
3SLS	λ	0.10	0.1080	0.1056	0.1958
	α	-1.00	-1.0053	-1.0051	0.0488
	β	0.80	0.7990	0.7991	0.0094
	ρ	-0.50	-0.5005	-0.5008	0.0174
	σ	0.50	0.4996	0.4995	0.0067

Table 2.3: Monte Carlo Simulation: Estimation of Model 2(M=5000)

DGP	Parameter	True Value	Mean	Median	St. Dev.
FIML Estimation of Model 2					
IV1	λ	0.10	0.1024	0.1019	0.1900
	α	-1.00	-1.0054	-1.0043	0.0493
	β	0.80	0.7989	0.7990	0.0095
	ρ	-0.50	-0.5005	-0.5007	0.0176
	σ	0.50	0.4999	0.5000	0.0064
IV2	λ	0.10	0.1024	0.1019	0.1900
	α	-1.00	-1.0054	-1.0043	0.0493
	β	0.80	0.7989	0.7990	0.0095
	ρ	-0.50	-0.5005	-0.5007	0.0176
	σ	0.50	0.4999	0.5000	0.0064
IV3	λ	0.10	0.1024	0.1019	0.1900
	α	-1.00	-1.0054	-1.0043	0.0493
	β	0.80	0.7989	0.7990	0.0095
	ρ	-0.50	-0.5005	-0.5007	0.0176
	σ	0.50	0.4999	0.5000	0.0064
3SLS Estimation of Model 2					
3SLS	λ	0.10	0.1020	0.1014	0.1925
	α	-1.00	-1.0054	-1.0046	0.0494
	β	0.80	0.7989	0.7990	0.0095
	ρ	-0.50	-0.5005	-0.5007	0.0176
	σ	0.50	0.4999	0.5000	0.0064

Table 2.4: Monte Carlo Simulation: Estimation of Model 2(M=10000)

DGP	Parameter	True Value	Mean	Median	St. Dev.
FIML Estimation of Model 2					
IV1	λ	0.10	0.1003	0.0993	0.1879
	α	−1.00	−1.0049	−1.0041	0.0490
	β	0.80	0.7990	0.7992	0.0095
	ρ	−0.50	−0.5000	−0.5002	0.0177
	σ	0.50	0.4998	0.4999	0.0064
IV2	λ	0.10	0.1003	0.0993	0.1879
	α	−1.00	−1.0049	−1.0041	0.0490
	β	0.80	0.7990	0.7992	0.0095
	ρ	−0.50	−0.5000	−0.5002	0.0177
	σ	0.50	0.4998	0.4999	0.0064
IV3	λ	0.10	0.1003	0.0993	0.1879
	α	−1.00	−1.0049	−1.0041	0.0490
	β	0.80	0.7990	0.7992	0.0095
	ρ	−0.50	−0.5000	−0.5002	0.0177
	σ	0.50	0.4998	0.4999	0.0064
3SLS Estimation of Model 2					
3SLS	λ	0.10	0.1002	0.1001	0.1911
	α	−1.00	−1.0049	−1.0041	0.0491
	β	0.80	0.7990	0.7992	0.0095
	ρ	−0.50	−0.5000	−0.5003	0.0177
	σ	0.50	0.4998	0.4999	0.0064

Table 2.5: Monte Carlo Simulation: Estimation of Model 1 with measurement error (M=10000)

FIML Estimation of Model 1					
DGP	Parameter	True Value	Mean	Median	St. Dev.
$\varepsilon_t \sim N(0,1)$	λ	0.10	0.0625	0.0614	0.1132
	α	−1.00	−3.3710	−3.3702	0.1158
	β	0.80	0.3256	0.3255	0.0220
	ρ	−0.50	−0.2276	−0.2272	0.0242
	σ	0.50	1.2290	1.2291	0.0167
$\varepsilon_t \sim N(0,0.1)$	λ	0.10	0.0984	0.0969	0.1772
	α	−1.00	−1.5109	−1.5088	0.0732
	β	0.80	0.6978	0.6981	0.0142
	ρ	−0.50	−0.3963	−0.3964	0.0182
	σ	0.50	0.6369	0.6368	0.0084
$\varepsilon_t \sim N(0,0.01)$	λ	0.10	0.1031	0.1010	0.1854
	α	−1.00	−1.0618	−1.0603	0.0533
	β	0.80	0.7877	0.7879	0.0103
	ρ	−0.50	−0.4849	−0.4853	0.0172
	σ	0.50	0.5159	0.5159	0.0067

Table 2.6: Monte Carlo Simulation: Estimation of Model 2 with measurement error (M=10000)

FIML Estimation of Model 2					
DGP	Parameter	True Value	Mean	Median	St. Dev.
$\varepsilon_t \sim N(0,1)$	λ	0.10	0.3658	0.3625	0.2398
	α	−1.00	−3.3830	−3.3829	0.1123
	β	0.80	0.3251	0.3251	0.0214
	ρ	−0.50	−0.2527	−0.2523	0.0422
	σ	0.50	1.2290	1.2291	0.0161
$\varepsilon_t \sim N(0,0.1)$	λ	0.10	0.1381	0.1358	0.1945
	α	−1.00	−1.5121	−1.5097	0.0721
	β	0.80	0.6978	0.6981	0.0140
	ρ	−0.50	−0.3997	−0.3997	0.0199
	σ	0.50	0.6369	0.6368	0.0084
$\varepsilon_t \sim N(0,0.01)$	λ	0.10	0.1070	0.1053	0.1915
	α	−1.00	−1.0619	−1.0602	0.0527
	β	0.80	0.7877	0.7880	0.0102
	ρ	−0.50	−0.4853	−0.4855	0.0179
	σ	0.50	0.5159	0.5159	0.0067

2.8 Bibliography

AMEMIYA, T. (1977): "A Note on a Heteroscedasticity Model," *Journal of Econometrics*, 6(3), 365–70.

ANDERSEN, T. G. (1994): "Comment on E. Jacquier, N. Polson and E. Rossi," *Journal of Business and Economic Statistics*, 12, 389–92.

ANDERSEN, T. G., AND B. SORENSEN (1996): "GMM Estimation of A Stochastic Volatility Model: A Monte Carlo Study," *Journal of Business and Economic Statistics*, 14, 329–52.

BANSAL, R., A. R. GALLANT, R. HUSSEY, AND G. TAUCHEN (1993): "Computational Aspects of Nonparametric Simulation Estimation," *Computational Techniques for Econometrics and Economic Analysis*, pp. 3–22.

BLACK, F. (1976): "Studies of Stock Price Volatility Changes," *Proceedings of the American Statistical Association,Business and Economics Statistics Section*, pp. 177–81.

BOLLERSLEV, T. (1986): "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 51, 307–327.

BOLLERSLEV, T., AND H. ZHOU (2006): "Volatility Puzzles: A Simple Framework for Gauging Return-Volatility Regressions," *Journal of Econometrics*, 131(1-2), 123–50.

BROTO, C., AND E. RUIZ (2004): "Estimation Methods for Stochastic Volatility Models:A Survey," *Journal of Economic Surveys*, 18(5), 613–649.

CAMPBELL, J. Y., AND L. HENTSCHEL (1992): "No News is Good News:An

Asymmetric Model of Changing Volatility in Stock Returns," *Journal of Financial Economics*, 31, 281–318.

CAMPBELL, J. Y., AND A. S. KYLE (1993): "Smart Money, Noise Trading, and Stock Price Behavior," *The Review of Economic Study*, 60(1), 1–34.

CHIB, S., O. ELARIAN, AND N. SHEPHARD (2001): "Likelihood Inference for Discretely Observed Nonlinear Diffusions," *Econometrica*, 69(4), 959–93.

CHRISTIE, A. (1982): "The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects," *Journal of Financial Economics*, pp. 407–432.

CLARK, P. K. (1973): "A Subordinated Stochastic Process Model with Fixed Variance for Speculative Prices," *Econometrica*, 41, 135–156.

COURT, R. H. (1973): "Efficient Estimation of the Reduced Form from Incomplete Econometric Models," *The Review of Economic Studies*, 40(3), 411–417.

——— (1974): "Three Stage Least Squares and Some Extensions Where the Structural Disturbance Covariance Matrix May be Singular," *Econometrica*, 42(3), 547–558.

COX, D. R. (1981): "Statistical Analysis of Time Series: Some Recent Developments," *Scandanavian Journal of Statistics*, 8, 93–115.

DANIELSSON, J., AND J. F. RICHARD (1993): "Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models," *Journal of Applied Econometrics*, 8, 153–73.

DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–63.

DUFFIE, D., AND K. J. SINGLETON (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–52.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of the United Kingdom Inflation," *Econometrica*, 50, 987–1007.

——— (1994): "Indirect Inference on Volatility Diffusions and Stochastic Volatility Models," Manuscript, UC San Diego.

FAMA, E. F. (1963): "Mandelbrot and the Stable Paretian Hypothesis," *The Journal of Business*, 36(4), 420–429.

FRENCH, K. R., G. W. SCHWERT, AND R. F.STAMBAUGH (1987): "Expected Stock Returns and Volatility," *Journal of Financial Economics*, 15, 265–302.

GALLANT, A. R., D. HSIEH, AND G. TAUCHEN (1997): "Estimation of Stochastic Volatility Models with Diagnostic," *Journal of Econometrics*, 81, 159–192.

GALLANT, A. R., P. E. ROSSI, AND G. TAUCHEN (1992): "Stock Price and Volume," *The Review of Financial Studies*, 5(2), 199–242.

GALLANT, A. R., AND G. TAUCHEN (1996): "Which Moments to Match?," *Econometric Theory*, 12, 657–81.

GHYSEL, E., A. C. HARVEY, AND E. RENAULT (1996): "Stochastic Volatility," *Stochastic Methods in Finance*, Stochastic Methods in Finance, Amsterdam: Elsevier Science Publishers(G.S. Maddala and C.R. Rao (eds)), 119–192.

GOURIEROUX, C., A. MONFORT, AND E. RENAULT (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, 85–118.

HARVEY, A. C., E. RUIZ, AND N. G. SHEPHARD (1994): "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–64.

HESTON, S. L. (1993): "A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327–44.

JACQUIER, E., N. G. POLSON, AND P.ROSSI (1994): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371–417.

——— (2004): "Bayesian Analysis of Stochastic Volatility Models with Fat-tails and Correlated Errors," *Journal of Econometrics*, 122, 185–212.

JIANG, G., AND J. KNIGHT (2002): "Estimation of Continuous-Time Processes via the Empirical Characteristic Function," *Journal of Business and Economics Statistics*, 21(2), 198–212.

JIANG, G., AND J. KNIGHT (2010): "ECF Estimation of Markov Models Where the Transition Density is Unknown," *Econometrics Journal*, 13, 245–70.

JIANG, G., J. KNIGHT, AND G. WANG (2005): "Alternative Specifications of Stochastic Volatility Models-Theoretical and Empirical Comparison," Working paper.

JIANG, G., AND P. J. VAN DER SLUIS (2000): "Index Option Pricing Models with Stochastic Volatility and Stochastic Interest Rates," *European Finance Review*, 3, 273–310.

KELEJIAN, H. H. (1974): "Random Parameters in a Simultaneous Equation Framework: Identification and Estimation," *Econometrica*, 42(3), 517–28.

KIM, S., N. SHEPHARD, AND S. CHIB (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *The Review of Economic Studies*, 65(3), 361–393.

KNIGHT, J., AND J. YU (2002): "Empirical Characteristic Function in Time Series Estimation," *Econometric Theory*, 18, 691–721.

MANDELBROT, B. (1963): "The Variation of Certain Speculative Prices," *Journal of Business*, 36(4), 394.

MCNEIL, A., R. FREY, AND P. EMBRECHTS (2005): "Financial Time Series," *Ch. 4, Quantitative Risk Management*, Princeton Series in Finance.

MELINO, A., AND S. M. TURNBULL (1990): "Pricing Foreign Currency Options with Stochastic Volatility," *Journal of Econometrics*, 45, 239–65.

NELSON, D. B. (1988): "The Time Series Behavior of Stock Market Volatility and Retruns," Unpublished ph.d. dissertation, MIT, Cambridge, MA.

——— (1991): "Conditional Heteroscedasticity in Asset Returns: A New Approach," *Econometrica*, 59(2), 347–70.

RUIZ, E. (1994): "Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Econometrics*, 63, 289–306.

SARGAN, J. D. (1964): "Three-Stage Least-Squares and Full Maximum Likelihood Estimates," *Econometrica*, 32, 77–81.

SHEPHARD, N. G. (1993): "Fitting Non-Linear Time Series Models, with Application to Stochastic Variance Models," *Journal of Applied Econometrics*, 8, 135–52.

——— (1996): "Statistical Aspects of ARCH and Stochastic Volatility," Time series models in econometrics, finance and other fields, Cox, D.R. and Hinkley,D.V. and Barndorff-Nielsen,O.E.eds.

SHI, L. (2005): "Estimation of Discrete-Time Stochastic Volatility in Mean Model Using Realized Volatility," Chapter 2, unpublished thesis.

SINGLETON, K. J. (2001): "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics*, 102, 111–41.

TAUCHEN, G., AND M. PITTS (1983): "The Price Variability Volume Relationship on Speculative Markets," *Econometrica*, 51, 485–505.

TAYLOR, S. J. (1986): "Modeling Financial Time Series," John wiley, chichester.

XIAO, L. (2007): "Estimation of Stochastic Volatility Models Using Realized Volatility," Working paper.

Chapter 3

Estimating and Forecasting Stochastic Volatility Models Using Realized Volatility

3.1 Introduction

Volatility has long been a persistent interest for academics, policy makers, and practitioners. Policy makers use volatility of financial markets as a measurement of risk with large volatility producing a significant impact on the economy and hence policy responses. For practitioners, asset volatility is a key input to many investment decisions. It drives the creation of financial assets, and in some derivatives, volatility itself is viewed as the underlying asset. Academics interest is not only on modeling but also forecasting volatility, as they are crucial in determining model adequacy.

However, there is an issue in volatility modeling and forecasting. That is, volatility itself, in fact, is not directly observable. In financial markets, volatility typically displays certain stylized facts, such as high kurtosis or fat tails, volatility clustering, persistency, leverage effect between asset return and volatility, etc. An estimator which is able to capture these stylized facts is demanded in both theoretical studies and practical applications.

Over the last two decades, with the availability of high frequency intra-day data, realized volatility has been developed as a proxy for the latent volatility. The idea of realized volatility was initiated by Merton (1980), who gave rise to the notion that an estimate of asset return variance for a given period can be obtained by summing the intra-period squared returns from a Gaussian diffusion process. As discussed by Andersen, Bollerslev, Diebold and Labys (2001), as well as Barndorff-Nielsen and Shephard (2002), under suitable conditions, realized volatility can be a consistent, unbiased, and highly efficient estimator of true return volatility. By sampling intra-day asset returns with sufficient frequency, realized volatility can be not only model-free but also arbitrarily close to the underlying integrated volatility, hence one can treat volatility, when modeling and forecasting, as essentially observed.

Realized volatility has been widely used in estimation of both discrete-time and continuous-time stochastic volatility (SV) models, to replace the true latent volatility hence simplifying estimation procedures. For example, Bandorff-Nielsen and Shephard (2002) used the quasi-maximum likelihood (QML) estimation method based on the time series of realized volatility to estimate the parameters of continuous-time SV models. Bollerslev, Gibson and Zhou (2004) constructed monthly realized volatility from 5-minute transaction prices, and employed the generalized method of moments (GMM) to estimate the volatility risk premium or risk aversion. Bollerslev and Zhou (2006) proposed GMM estimation for stochastic volatility diffusions using realized volatility constructed from both equity market and exchange rates. Shi (2005) used realized volatility of S & P 100 index in estimating the modified SV in mean (SV-M) model.

The iterated generalized least square(IGLS) and maximum likelihood (MLE) methods were applied in her study. In general, the empirical evidence in these studies shows, with volatility being "observed", not only the estimation is straightforward, computationally easy, but also the statistical inference is reliable. Realized volatility has also been widely used to evaluate volatility forecasting performance. For example, Andersen and Bollerslev (1998), Hansen and Lunde (2005), and Patton (2006) used realized volatility to evaluate the out-of-sample forecasting performance of GARCH models. Andersen, Bollerslev, Diebold, and Labys (2003) applied a simple long-memory Gaussian vector autoregression for the logarithmic daily realized volatilities forecasting. Liu and Maheu (2008) proposed a Bayesian model average approach for realized volatility forecasting.

Motivated by the accuracy of realized volatility, the main purpose of this chapter is to examine the estimation and forecasting performance of discrete-time SV models using realized volatility. As we discussed in Chapter 2, both FIML and 3SLS estimation performance for SV models using a volatility proxy was affected by the severity of measurement error. When the volatility proxy error follows a normal distribution with a very small variance, the estimates are fairly close to the true values of the parameters. As theoretical studies have shown realized volatility is a highly efficient estimator for the latent volatility under suitable conditions, in this chapter we focus on examining both estimation and forecasting of discrete-time SV models using realized volatility.

Our study contributes to the literature in three aspects. First, we examine the estimation of discrete-time SV models with both lagged inter-temporal dependence and contemporaneous dependence using realized volatility. Jiang, Knight and Wang (2005), and Shi (2005) studied the estimation of discrete-time

SV models. In Jiang, Knight and Wang (2005), they provided an excellent theoretical analysis of discrete-time SV models with lagged inter-temporal dependence and contemporaneous dependence, and demonstrated that both models should receive further attention. However, in their study, the volatility process was treated as latent, GMM was applied. Shi (2005) constructed daily realized volatility of S & P 100 index from high frequency intra-day transaction prices, and employed IGLS and MLE methods, but her study only focused on estimation of discrete-time SV model with lagged inter-temporal dependence, while leaving SV model with contemporaneous dependence unexamined. This chapter extends the study of discrete-time SV model estimation by using daily realized volatility along with both FIML and 3SLS to estimate SV models with both lagged inter-temporal dependence and contemporaneous dependence. Second, we first examine forecasting performance of discrete-time SV model with contemporaneous dependence. Some studies have included the discrete-time SV model as a candidate model in volatility forecasting. For example, Lopez (2001) compared performance of GARCH, EWMA, and SV model in forecasting volatility of exchange rates. Yu (2002) included SV model as a candidate in forecasting volatility in New Zealand stock market. Hol and Koopman (2002) examined the forecasting performance of SV model for stock index volatility. However, these studies concentrated on examining the forecasting performance of SV model with two error terms independently distributed. Comparison of volatility forecasting performance of SV model with two error terms correlated, especially contemporaneously correlated, has not yet been made in the literature. Third, unlike other studies only using realized volatility to evaluate the out-of-sample forecasting performance of SV models, in this chapter we use daily realized volatility not only to evaluate the out-of-sample forecasting performance, but also in the in-sample estimation.

Our empirical analysis is based on both low frequency daily and high frequency intra-day observations for S & P 500 index and three foreign exchange rates, including CAD/USD, DEM/USD, and USD/GBP. The high frequency intra-day transaction prices are used to construct daily realized volatility. As Andersen, Bollerslev, Diebold and Labys (hence ABDL) (1999) demonstrated, although from the theory of quadratic asset return variation, realized volatility constructed on the finest time interval can be an extremely accurate estimator of latent volatility, in reality, the true price process is often contaminated by market microstructure effects, such as the bid-ask spread and asynchronous trading. Consequently, realized volatility constructed over extremely small time intervals fails to converge to the underlying quadratic variation of the log price process. As a result, the optimal sampling frequency to construct realized volatility should be some intermediate value that is high enough to produce a volatility estimate with negligible sampling variation, yet low enough to avoid microstructure bias. To select the optimal sampling frequency, we employ the "volatility signature plot", which was first proposed by ABDL (1999). In addition, we consider different approaches to deal with either distortions introduced by the market closing, or the autocorrelation of intra-day returns.

The rest of this chapter is organized as follows. Section 3.2 briefly reviews realized volatility estimation, then describes empirical data and different approaches to construct daily realized volatility. The statistical analysis of daily asset return and realized volatility series is also reported in this section. In Section 3.3, we concentrate on estimation of SV models with lagged intertemporal and contemporaneous dependence using different realized volatilities, and results are reported. We focus on the forecasting of realized volatility in Section 3.4. We list four competing models and different measures employed to evaluate the forecasting performance. Forecasting performance is reported.

A brief conclusion is contained in the last section.

3.2 Realized Volatility Estimator, Data and Construction

3.2.1 Realized Volatility Estimator and Its Distribution

The main idea behind realized volatility is to sum high-frequency intra-day squared returns in a trading day to approximate the daily quadratic variation of the log price process (or to sum daily squared returns to approximate the monthly variation of the log price process).

Consider a discrete-time model in which the daily asset return is expressed as:

$$r_t = \sigma_t v_t$$

where $v_t \sim iidN(0, 1)$.

In a typical trading day t, the prices, $P_{t,d}, d = 1, \ldots, D$, are observed tick-by-tick. D refers to the total number of observations at day t. The dth intra-period return at day t can be calculated by taking the difference between logarithmic price at d and that at $d - 1$:

$$r_{t,d} = log(P_{t,d}) - log(P_{t,d-1})$$

Assume $r_{t,d} = \sigma_{t,d} v_{t,d}$ where $v_{t,d} \sim iidN(0, \frac{1}{D})$

Then, the daily return is:

$$r_t = log(P_{t,D}) - log(P_{t,0}) = \sum_{d=0}^{D} r_{t,d}$$

and

$$\sigma_t = \frac{1}{D}\sum_{d=1}^{D} \sigma_{t,d}$$

The squared daily asset return is:

$$r_t^2 = \sum_{d=0}^{D} r_{t,d}^2 + crossproduct$$

The daily realized volatility (or variance) is computed as:

$$RV_t = \sum_{d=0}^{D} r_{t,d}^2$$

If the intra-day returns are uncorrelated, then

$$Var(r_t) = E(r_t^2) = E(\sum_{d=0}^{D} r_{t,d}^2) + E(crossproduct) = E(RV_t) = \sigma_t^2$$

Realized volatility as well as squared asset return are unbiased estimators for the variance of daily asset returns.

As McAleer and Medeiros (2008) showed,

$$Var(RV_t) = \frac{2}{D}\sum_{d=0}^{D} \frac{\sigma_{t,d}^4}{D} < \frac{2}{D}(\sum_{d=0}^{D} \frac{\sigma_{t,d}^2}{\sqrt{D}})^2 = Var(r_t^2)$$

Realized volatility is a more efficient or more accurate estimator compared to the squared asset return.

Andersen, Bollerslev, Diebold, and Labys (2003) showed that when there is no microstructure noise and squared returns are summed over very small time intervals, i.e. as the sampling frequency of returns $D \to \infty$, the constructed realized volatility converges uniformly in probability to the underlying integrated variance:

$$RV_t \rightarrow_p V_t$$

Realized volatility is a consistent estimator of the variance of daily asset returns.

The asymptotic distribution of realized volatility has been derived by Bandorff-Nielsen and Shephard (2002). As the computation of the asymptotic distribution is infeasible given that the integrated quarticity is unknown, the asymptotic distribution of realized volatility can be approximated by:

$$\frac{1}{\sqrt{\frac{2}{3}\sum_{d=0}^{D} r_{t,d}^4}}(RV_t - V_t) \rightarrow^d N(0,1)$$

However, in Bandorff-Nielsen and Shephard (2002), the Monte Carlo results suggested very large sampling frequency D, the finite sample performance was poor. On the other hand, the logarithmic transformation of realized volatility is well recognized to be able to provide better finite sample properties. The approximate asymptotic distribution of logged realized volatility is:

$$\frac{1}{\sqrt{\frac{2}{3}\frac{\sum_{d=0}^{D} r_{t,d}^4}{(\sum_{d=0}^{D} r_{t,d}^2)^2}}}(log(RV_t) - log(V_t)) \rightarrow^d N(0,1)$$

3.2.2 Data

Our empirical analysis is based on daily returns and realized volatilities of both S & P 500 index and three currency exchange rates, namely Canadian dollar (CAD/USD), British Pound (USD/GBP) and German Mark (DEM/USD). The daily return series are obtained from WRDS, the high-frequency intra-day data is provided by Disktrading. The high-frequency transaction prices for S & P 500 stock index are available from 9:30 to 16:00 for the period from November 12, 1997 through July 28th, 2006, and those for the three exchange rates are available twenty four hours from April 13th, 1998 through July 28th, 2006. To avoid complicating the inference by the decidedly slower trading activity during certain holiday periods, we exclude a number of inactive days (July 4th, December 24, 25, 26, and January 1, 2) from the sample. For the S & P 500 index, the sample period covers 2187 days, and for exchange rates, the sample size is 2126.

3.2.3 Constructing Daily Realized Volatility

We consider different approaches to construct daily realized volatility series. First, equity markets are open between 9:30 and 16:00 every trading day, whereas exchange rate markets are open twenty four hours each day. It is typically found that the changes of prices during the closed parts (overnight) are much larger than those during open parts, so we need to take into account the "closed effect" when constructing daily realized volatility of S & P 500 index. For exchange rate markets, as they open twenty four hours each day, there is no closed part distortion. However, it is well known that there is significant autocorrelation in intra-day returns, although the autocorrelation in daily asset returns is not obvious. We consider an approach to deal with the first order autocorrelation in each market.

Construction of Realized Volatility for S & P 500 Index

Let $P_{t,d}$ represent the d-th intra day transaction price on trading day t with $d = 1, \ldots, D$. Let f be the sampling frequency of returns. For example, $f = 5$ indicates five-minute returns, $f = 25$ means 25-minute returns. In each trading day, there are $D_f = D/f$ intra day returns.

For the open part, the high-frequency intra day return is computed as:

$$x_{f,t,d} = 100(\ln P_{t,fd} - \ln P_{t,f(d-1)}),\ t = 1, \ldots, T$$

For the closed part (overnight), the return is calculated as:

$$x_t^{ON} = 100(\ln P_{t,1} - \ln P_{t-1,D})$$

where $P_{t,1}$ is the first price of trading day t and $P_{t-1,D}$ represents the last price on day $t-1$.

The common definition of daily realized volatility for equity markets is to simply sum up the open part squared returns and the overnight squared returns. We denote it as $RV1$:

$$RV1 = \sum_{d=1}^{D_f} x_{f,t,d}^2 + (x_t^{ON})^2$$

However, it is observed that the changes in the stock index prices during the closed part are relatively much larger than those during open parts. The above computation may produce a significant amount of noise because a large value of $(x_t^{ON})^2$ will have a distorting effect. The simplest solution for this issue is to exclude the overnight squared returns. Only summing open part squared returns, on the other hand, may underestimate the true volatility. To solve this problem, Martens (2002) proposed scaling the sum of open part squared returns. We denote it as $RV2$.

$$RV2 = \omega \sum_{d=1}^{D_f} x_{f,t,d}^2$$

where

$$\omega = \frac{var(\ln P_{t,D} - \ln P_{t,1}) + var(\ln P_{t,1} - \ln P_{t-1,D})}{var(\ln P_{t,D} - \ln P_{t,1})} > 1$$

Even though there is little autocorrelation in the daily returns, there is strong autocorrelation in the intra day returns, caused by market structure effects. Standard measures of realized volatility may suffer from a bias due to autocorrelation in the intra day returns. As sampling frequency increases, the autocorrelation in intra day returns becomes more of an issue. In order to deal with the autocorrelation problem when constructing realized volatility, Zhou (1996) introduced a modified kernel-based estimator which includes the cross products of adjacent returns. Hansen and Lunde (2005a), and Cornish (2007) used this estimator in their studies. As Hansen and Lunde (2006) showed, the kernel-based estimator that utilizes higher order auto-covariances can eliminate the bias of realized volatility. Specifically, Hansen and Lunde (2005) defined the daily realized volatility as:

$$RV_{ACF} = \sum_{d=1}^{D_f} x_{f,t,d}^2 + 2\sum_{h=1}^{q} \frac{D_f}{D_f-h} \sum_{d=1}^{D_f-h} x_{f,t,d}x_{f,t,d+h}$$

where q refers to the order of the auto-covariances. The fraction $\frac{D_f}{D_f-1}$ is included to compensate the "missing" cross product which requires intra day returns outside of the interval.

A drawback of this kernel-based estimator is that it may produce a negative estimate of volatility, as Hansen and Lunde (2005) pointed out, especially when intra day high-frequency returns have significant negative autocorrelation. This is typically observed in foreign exchange rates intra day returns. To overcome this drawback, Hansen and Lunde (2005b) demonstrated that the Bartlett kernel should be used and defined the daily realized volatility as:

$$RV_{NW} = \sum_{d=1}^{D_f} x_{f,t,d}^2 + 2\sum_{h=1}^{q}(1 - \frac{h}{q+1}) \sum_{d=1}^{D_f-h} x_{f,t,d}x_{f,t,d+h}$$

Using this approach, non-negative realized volatility is guaranteed.

In our study, we consider constructing daily realized volatility with the first-order autocorrelation of intra-day asset returns taken into account, that is, we set $q = 1$ in RV_{NW}. Following Hansen and Lunde (2005b), we employ the Bartlett kernel estimator to ensure a positive estimate. $\frac{D_f}{D_f-1}$ is included in the second term to compensate the"missing" cross product.

$$RV3 = \sum_{d=1}^{D_f} x_{f,t,d}^2 + \frac{D_f}{D_f-1} \sum_{d=1}^{D_f-1} x_{f,t,d}x_{f,t,d+1}$$

Construction of Realized Volatility for Foreign Exchange Rates

Since foreign exchange markets open twenty four hours each day, there is no "closed part" distorting. Hence summing up intra day squared returns in each day, we get daily realized volatility. We denote it as $RV1$.

$$RV1 = \sum_{d=1}^{D_f} x_{f,t,d}^2$$

In the second approach, we take first-order autocorrelation into account, and denote it as $RV2$:

$$RV2 = \sum_{d=1}^{D_f} x_{f,t,d}^2 + \frac{D_f}{D_f-1} \sum_{d=1}^{D_f-1} x_{f,t,d}x_{f,t,d+1}$$

3.2.4 Optimal Sampling Frequency

Theoretically the sum of squared returns from finest time intervals can be not only a model-free but also an error free estimator of integrated volatility. However, in reality, when the time intervals are very small, microstructure effects, such as discreteness of prices, bid/ask bounce, and trading volume, etc., may distort the realized volatility. Hence, there is a trade-off between bias and variance when choosing the sampling frequency to construct realized volatility. As ABDL (1999) and others demonstrated, the optimal sampling frequency to construct daily realized volatility should be moderate such that it is fine enough to produce a volatility estimate with negligible sampling variation, yet low enough to avoid microstructure bias. ABDL (1999) proposed a simple graphical diagnostic called "volatility signature plot", a plot of average of realized volatilities constructed from different sampling frequencies against these time intervals, to provide some guide for selecting the optimal sampling frequency.

The basic procedure of constructing "volatility signature plot" is, considering a financial market which opens twenty-four hours at day t, setting sampling frequency f as $f = 5, 10, 15, \ldots$, we obtain $288, 144, 96, \ldots$ intra day asset returns, respectively. For each sampling frequency, the t-th daily realized volatility, RV_t, can be constructed. Compute the average of RV_t for $t = 1, \ldots, T$ for each value of sampling frequency f and plot them against f.

As sampling frequency increases, microstructure effects may manifest themselves in these "volatility signature plots" by distorting the average of realized volatility. By plotting the averages of realized volatility constructed from different sampling frequencies against these time intervals, allows one to choose the optimal sampling frequency for the realized volatility construction.

Figure 3.1-3.4 shows volatility signature plots for S & P 500 index and three foreign exchange rates, respectively. The daily realized volatilities are constructed from 5-minute to 90-minute frequencies.[1] Then for each set of realized volatilities, the average is computed. Overall we find the change of means of realized volatilities is more pronounced in equity market than those in foreign

[1]We report RV2 for S & P 500 index, and RV1 for exchange rates.

exchange rates. In S & P 500 index, volatility signature plot starts to display "saw tooth" pattern from 15-minute frequency suggesting that microstructure effects are significant during these time intervals. From 5-minute to 15-minute time frequency, the means of realized volatilities display relatively smooth pattern. This finding suggests that from 5-minute to 15-minute time interval, microstructure effects are small. Combining these findings with the theory that realized volatility constructed from fine enough time interval can be a very accurate estimator, we select 5-minute sample frequency to construct daily realized volatility for S & P 500 index. This choice is consistent with other equity market realized volatility studies, including Andersen and Bollerslev (1998), Shi (2005) etc.

On the other hand, foreign exchange rates volatility signature plots display a different pattern. We notice that at 5-minute time interval, the average of realized volatility is very high. As sampling frequency decreases from 5-minute to 25-minute, the average of realized volatility decreases and becomes relatively smooth at 25-minute. However, there is a significant change from 30-minute frequency for both CAD/USD and USD/GBP, and from 35-minute frequency for DEM/USD, suggesting that the optimal frequency to construct foreign exchange rate realized volatility may be approximately 25 minutes. This finding is similar to that in Andersen, Bollerslev, Diebold and Labys (2002). They suggested that 30-minute transaction prices was the best choice for construction of daily exchange rate realized volatility to achieve the accuracy as well as minimize market microstructure effect. We also notice that the average of realized volatility is relatively smooth around 45-minute time interval indicating that at that time interval, microstructure effects are not large, however, it is not an optimal sampling frequency because daily realized volatility constructed from low sampling frequency will suffer from a high sampling error.

Based on our findings from the volatility signature plots, we select a 5-minute time interval to construct daily realized volatility for the equity index and a 25-minute interval for the foreign exchange rates.

3.2.5 Statistical Analysis of Daily Return and Realized Volatility

The sample for the S & P 500 index consists of 2187 daily observations for return and RV series. Table 3.1 reports the summary statistics for daily return (x_t), square daily return (x_t^2), daily RV1, RV2, RV3, and their natural logs. The top panel displays moments. The middle panel reports quantiles, extreme values. In addition, we apply a Jarque-Bera (JB) test for the null hypothesis that the data are from a normal distribution. The test statistic JB is defined as: $JB = \frac{n}{6}(S^2 + \frac{1}{4}K^2)$, with $S = \frac{\frac{1}{n}\sum(x_i-\bar{x})^3}{(\frac{1}{n}\sum(x_i-\bar{x})^2)^{3/2}}$, and $K = \frac{\frac{1}{n}\sum(x_i-\bar{x})^4}{(\frac{1}{n}\sum(x_i-\bar{x})^2)^2} - 3$. Since samples from a normal distribution have an expected skewness of zero and expected excess kurtosis of zero, any deviation from this increases the JB statistic. We report the JB results along with p-values at the bottom of the middle panel. The bottom panel shows autocorrelation of these series.

The mean daily return is 0.0132, indicating that overall the daily return of S & P 500 index is positive, but very close to zero. The standard deviation is 1.1250, suggesting that the equity market is volatile. The negative value of skewness, -0.1, suggests that the return series is left-skewed. The daily return and squared return series both display high kurtosis with the values being 5.5988, and 85.3414, respectively, indicating fatter tail features which is a typical finding in financial markets.

The minimum daily return is -6.7953 while the maximum is 5.7522 making the range of this series around 12. A JB test is conducted to determine whether the series follows a normal distribution. The results strongly reject the null hypothesis.

The mean values for all three RV_s are above 1, and all the mean values for their natural logs are negative. The positive skewness along with large positive values of kurtosis indicate that RV is asymmetrically distributed with fatter tails. However, we notice that the skewness for natural log of RV_s are close to 0 and the fourth moment is around 3, indicating that the logged realized volatility is approximately Gaussian.

The sample for the three foreign exchange rates consists of 2126 daily observations. Table 3.2, 3.3, 3.4 reports the summary statistics for daily return, square return, RV1, RV2 and their natural logs for CAD/USD, USD/GBP, and DEM/USD, respectively.

The mean of returns for exchange rates are all very close to zero. The standard deviations for exchange rates are much smaller than those for the equity market, suggesting that exchange rate markets are not as volatile as the equity market. This result can also be confirmed by the values of RV. RV values for the equity market are above 1, while those for exchange rate markets are less than 0.5. The values of third and fourth moments again suggest that both return and volatility series are asymmetrically distributed with fatter tails, while the natural logs of RV_s are approximately Gaussian.

The range of returns for exchange rates are much smaller than that for the S & P 500 index. The JB results strongly reject the null that these series follow a normal distribution.

The bottom panel in each table reports sample autocorrelations of these series. The lagged value is up to 15. The daily asset return exhibits little serial correlation, whereas the autocorrelation increases significantly and is not negligible for the squared return series which is consistent with the GARCH theory. On the other hand, both RV series and logged RV series exhibit pronounced serial correlation and decay very slowly. This evidence is consistent with the fact that the volatility is clustering and persistent, suggesting that even though return is hardly predictable, the volatility process has a long memory hence is predictable.

Figures 3.5-3.12 display the plot of daily return, squared return, RV_s, and logged RV_s from the equity market. As all three exchange rates display similar patterns, we only report those plots for CAD/USD in figures 3.13-3.18. We observe some volatile periods from these plots. For example, the daily return series for S & P 500 is volatile at the end of 1997, the third quarter of 1998, the beginning of 2000, period after September 11th of 2001 and the Summer of 2002, while it is relatively smooth during the period from middle of 2003 through July 2006. The volatile periods are corresponding to the Asian finance crisis, 911 attack, and internet bubble, respectively. The plot of squared returns, RV_s and $ln(RV_s)$ confirm this observation. The plots for exchange rates display different patterns. For example, the daily return series for CAD/USD, and USD/GBP, is volatile at around May 1998, and the period from 2003 through 2006. For DEM/USD, we can identify the volatile period occurring from 1998 through 2000, which coincides with the Deutsche Mark (or German Mark) being gradually replaced by the euro. After a relatively smooth period, a volatile period occurs in 2004 again.

The autocorrelation function plot of these time series are reported from Figure 3.19 to Figure 3.25[2]. The lag value is up to 15. The straight lines parallel

[2]Again we only report ACF for CAD/USD since the ACF plots of USD/GBP and DEM/USD display similar patterns.

above and below zero indicate the approximate upper and lower 95 pencent confidence bounds. Note values of ACF that are effectively zero lie within these bounds. From these plots, clearly the ACF values of daily asset return series, no matter whether they are from the equity market or exchange rates, all lie within these bounds suggesting that this series is hardly serially correlated. However, when we square the asset return series, the ACF values increase significantly and are not negligible. On the other hand, from the graphs of RV_s and $ln(RV_s)$, we find all the ACF values lie above the upper bound, providing strong evidence of autocorrelation.

3.3 Estimation of Discrete-Time Stochastic Volatility Models Using Realized Volatility

In Chapter 2, we investigated the estimation of discrete-time SV models via a Monte Carlo study with both lagged inter-temporal dependence (M1) and SV model with contemporaneous dependence (M2) when volatility is an observable variable. In this section, we focus on an empirical investigation. Specifically, we use daily realized volatility constructed from high-frequency intra-day transaction prices as a proxy in both the M1 and M2 models and consequently examine the estimation performance of both FIML and 3SLS approaches. In the following parts, we first list our models, then report the empirical results.

3.3.1 Model Specification

The discrete-time SV model with lagged inter-temporal dependence (M1) is:

$$x_t = \lambda e^{h_t} + e^{h_t/2}\varepsilon_t \tag{3.1a}$$

$$h_t = \alpha + \beta h_{t-1} + \sigma v_t \tag{3.1b}$$

$$\begin{pmatrix} \varepsilon_{t-1} \\ v_t \end{pmatrix} \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}) \tag{3.1c}$$

The discrete-time SV model with contemporaneous dependence (M2) is:

$$x_t = \lambda e^{h_t} + e^{h_t/2}\varepsilon_t \tag{3.2a}$$

$$h_t = \alpha + \beta h_{t-1} + \sigma v_t \tag{3.2b}$$

$$\begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \sim N(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}) \tag{3.2c}$$

3.3.2 Estimation Results

Table 3.5 reports the estimation results for two models using S & P 500 index. Table 3.6, 3.7, 3.8 report the results using exchange rates. In each table, the top panel shows the results of Model 1 applying FIML, the middle panel displays the estimates of Model 2 using FIML, and estimates of Model 2 employing 3SLS are reported in the bottom panel. In each panel, we report the results using different RVs. The standard errors are reported in brackets[3].

For the equity index, the signs of all estimates are consistent using three different RVs. The estimates of λ and α are similar by using RV1 and RV2, whereas the estimates of β and σ are similar by using RV2 and RV3.

[3] ** means the estimate is significant at 1 percent level, * means it is significant at 5 percent level.

For Model 1, the FIML approach is applied. The estimates of risk premium parameter, λ, are round 0.01, but not significant using different RVs. This is a common finding in estimating SV model with lagged inter-temporal dependence. For example, when Shi (2005) examined the estimation of discrete-time SV-M model using daily realized volatility the estimated λ was not significant. The estimated values of λ and α are very similar when using RV1 and RV2, whereas RV2 and RV3 lead to similar estimates for β and σ. Overall, the estimated β and σ are significant at one percent level using different RVs. The estimated values of ρ are not significant when using RV1/RV2, whereas the estimated α is not significant using RV3.

For Model 2, both FIML and 3SLS procedures are employed. We find these two approaches provide very similar results. This is not surprising, as two methods are asymptotically equivalent. In general, the estimates are significant at one percent level except for those of α using RV3. The estimated values of the persistency parameter β are always above 0.71, consistent with the stylized fact of volatility persistency. The estimates of volatility, σ, are between 0.55 to 0.6, the standard errors are extremely small. However, the value when using RV1 is relatively larger than those from using RV2/RV3 (we also observe this fact in Model 1 estimation), indicating that the "closed part effect" does exist. Similarly as we observed for Model 1, the estimated values of λ and α are very similar when using RV1 and RV2, whereas RV2 and RV3 lead to similar estimates for β and σ. Moreover, the estimated ρ is much higher in absolute value using RV1 than those using RV2/RV3.

Comparing the results for Model 1 and Model 2, we find that the estimate of the correlation coefficient parameter, ρ, is only significant at the one percent level by using RV3 for Model 1, whereas they are always significant at the one percent level in Model 2, suggesting that Model 2 outperforms by capturing correlations between return and volatility processes. Both models yield similar estimates for all the other four parameters. The interesting finding is that both models provide higher β values along with lower α values when using RV3. The estimate of λ is much higher for Model 2 regardless of which estimation process is used, implying that Model 2 has a better fit to this data set.

For the three foreign exchange rates, two approaches of realized volatility lead to similar estimates. The only obvious difference is that no matter which estimation method is applied, for both models, the estimates of σ are approximately ten percent higher by using RV2 which is constructed with the first order autocorrelation of return series taken into account. This finding indicates that first order autocorrelation of intra-day returns is significant. The estimates of Model 2 using FIML are similar as those applying 3SLS again confirming the asymptotic equivalence of the two estimation methods. The estimates of α, β, ρ, and σ are mostly significant at one percent level, whereas the estimates of λ are very small in absolute values and unstable.

Comparing the results for Model 1 and Model 2, we find in general the estimates are similar except for those of estimated ρ. The sign of ρ changes between Model 1 and Model 2 for CAD/USD, DEM/USD, whereas it stays negative for USD/GBP.

Comparing the results using different exchange rates, we find the estimate of β has a higher value, above 0.6, for CAD/USD, and lower values for DEM/USD and USD/GBP which are around 0.4. This observation is consistent with what we observed from summary statistics tables. The ACF values of CAD/USD are much higher than those of USD/GBP and DEM/USD suggesting that our models do a good job in capturing the persistency of volatility series.

The estimated α ranges from 0.6156 to 0.9660 in absolute value for different exchange rates and are all significant.

Comparing the estimates for the S & P 500 index and the three exchange rates, we have several findings. First, the estimated risk premium parameter λ is significant at the one percent level for Model 2 for the equity index, whereas they are not for either model for the three exchange rates. Second, the estimated persistency parameter β is much higher for the equity index (above 0.71) than those for exchange rates (below 0.65), suggesting that volatility of equity return is more persistent. Moreover, these estimates are consistent with the ACF values of daily realized volatility series indicating that both models do a good job in capturing this stylized fact. Third, the estimates of α and ρ are improved in exchange rates, as they are all significant at one percent level. Fourth, the estimates of volatility of volatility, are similar and quite stable for both S & P 500 and exchange rates.

Overall, the standard errors are small suggesting that estimates are quite stable. The estimates of correlation coefficient parameter, ρ, are always within the natural band $[-1, 1]$ and mostly significant suggesting that both models do a reasonable job in capturing correlation between return and volatility processes. In general, both models fit data well, Model 2 does a better job than Model 1 in capturing some stylized facts. With respect to estimation methods, both FIML and 3SLS are appropriate when volatility is treated as an observable variable. And the statistical inference is reliable by using daily realized volatility as a proxy for the true latent volatility in estimating both models.

3.4 Forecasting of Discrete-Time Stochastic Volatility Models Using Realized Volatility

Given the important role that volatility has in financial markets, volatility forecasting has been of long-time interest for financial econometricians. However, the evaluation and comparison of volatility forecasts are complicated by the primary variable of interest, the conditional volatility, which is not directly observable. Consequently, ex-post measurement for the latent variable is needed for forecasting evaluation. Early studies such as Taylor (1987), Akgiray (1989), etc, used squared asset returns as a proxy for the latent volatility process as it had been proved to be a conditionally unbiased estimator. However, it has been recognized that squared return is not only a noisy, but also an inefficient estimator of the actual variance dynamics and will lead to incorrect inferences.

In recent studies realized volatility constructed from high frequency data has been widely used in volatility forecasting evaluation. As discussed in ABDL (2003), volatility forecasting performance has improved significantly as high-frequency volatility turns out to be highly predictable. Among these studies, some include the discrete-time SV model as a competing model in volatility forecast evaluation. For example, Lopez (2001) compared one-step-ahead exchange rate forecasting performance of discrete-time SV models, GARCH models, and simple models. For SV models, as volatility process is not directly observable, the estimation was conducted by using the Kalman filter on both asset return and volatility equations and applying quasi-maximum likelihood methods. The results indicated that forecast evaluation varied correspondingly to the choice of loss functions. Yu (2002) forecasted one-month-ahead volatility of New Zealand stock market, and reported the forecasting performance of nine models including discrete-time SV model, GARCH, random walk, etc. Treating volatility as latent in SV model, Yu (2002) applied the same procedure as Lopez (2001) then plugged the estimates into out-of-sample forecasting equation to generate volatility forecasts. Hol and Koopman (2002) concentrated on the S & P 100 volatility forecasting, using a daily discrete-time SV model as one of the candidates. As daily volatility series were unobserved, Hol and Koopman (2002) constructed the likelihood function using simulation methods then applied exact maximum likelihood methods. The results showed that the daily SV model outperformed the GARCH model. Two common features of these studies are: 1). when estimating discrete-time SV models, volatility is treated as latent, therefore the estimation is either inefficient or computationally complicated in practice; 2). the two disturbances in the discrete-time SV models are assumed to be mutually uncorrelated, both contemporaneously and at all lags.

In the following section, we investigate the forecasting performance of both Model 1 and Model 2 in which the two error terms are allowed to be correlated, either inter-temporally or contemporaneously. We use daily realized volatility not only to evaluate forecasting performance but also in the in-sample estimation. Specifically, we apply the FIML procedure to both Model 1 and Model 2 to obtain in-sample estimates then plug these estimates into the volatility equation of both models to generate one-day-ahead logarithmic volatility forecasts.

3.4.1 Methodology for Forecasting Volatility

We investigate the forecasting performance of four different models including a simple regression model in which the logarithmic of volatility follows an AR(1) process, the logarithmic version of the heterogeneous autoregressive (HAR) model, M1, and M2. We treat the simple regression model as the benchmark, hence we are able to examine: 1). whether capturing asymmetrical behavior between asset return and volatility processes would help volatility forecasting by comparing the forecasting performance of the benchmark with those of M1, M2; 2). whether permitting a long-memory would improve forecasting accuracy by comparing the forecasting performance of the benchmark with that of the HAR model. We have listed M1 and M2 in previous section, below we list the other two candidate models.

Simple Regression Model (SRM)

Both M1 and M2 consist of two equations, namely an asset return equation, and a volatility equation. The correlation between error terms in two equations enables SV models to capture the asymmetric behavior such as leverage effect (or feedback effect), as well as leptokurtosis of financial asset return series. Whether the models permitting these features would help volatility forecasting is of interest in the literature. Corsi, Mittnik, Pigorsch, and Pigorsch (2008) constructed a HAR-GARCH(1,1) with the error terms following a standardized normal inverse Gaussian (NIG) distribution to capture the fat tails of the asset returns, they found that permitting leptokurtosis in the innovation distribution did not help in point forecasting. In our study, to investigate whether discrete-time SV models would improve forecasting performance regression model, we compare the forecasting performance of Model 1, Model 2 with a simple AR (1) process with a constant, which is in fact the logarithmic volatility equation of the SV models.

$$h_t = \alpha + \beta h_{t-1} + \sigma v_t \tag{3.3}$$

where $v_t \sim N(0, 1)$.

The estimation is quite easy to implement, a simple OLS procedure can be applied. After the estimated parameters are obtained, the one-day-ahead forecast of h_{T+1} can be generated as:

$$\hat{h}_{T+1} = \hat{\alpha} + \hat{\beta} h_T$$

Heterogeneous Autoregressive Model (HAR)

One of the most important features of the volatility processes is that volatility is persistent. To capture this long-memory pattern, ABDL (2003) specified the autoregressive fractionally integrated moving average (ARFIMA (p,d,q)) model. However, as Corsi (2009) discussed, fractionally integrated models are nontrivial to estimate and not easily extendible to multivariate processes. Consequently, Corsi (2009) proposed a simple component model, a so called "heterogeneous autoregressive model (HAR)". As Corsi (2009), Corsi, Mittnik, Pigorsch, and Pigorsch (2008) demonstrated, the long-memory pattern could be reproduced by a sum of volatility components constructed over different time horizons. In Corsi (2009), the empirical results showed that both the HAR model's in-sample and out-of-sample performance were strong. Later, Andersen, Bollerslev, and Diebold (2007) included jump measures in the HAR model and showed that it provided a good predictive performance. They also considered the HAR model for the logarithm of volatility, and the results were similar. Following these studies, we consider the logarithmic version of the HAR model as a candidate in our study. The logarithmic version of the HAR model is in fact an extension of the simple regression model by including two extra terms, namely, the weekly logarithmic volatility and the monthly logarithmic volatility. The model is defined as:

$$h_t = \alpha + \beta_1 h_{t-1} + \beta_2 h_{t-5:t-1} + \beta_3 h_{t-22:t-1} + \sigma v_t \tag{3.4}$$

where v_t is Gaussian white noise.

The multi-period logarithmic volatility components are defined by

$$h_{t+1-k:t} = \frac{1}{k}\sum_{i=1}^{k} h_{t-i}$$

More specifically, a weekly volatility is given by the average:

$$h_{t-5:t-1} = \frac{1}{5}(h_{t-1} + h_{t-2} + h_{t-3} + h_{t-4} + h_{t-5})$$

Similarly, a monthly volatility is given by:

$$h_{t-22:t-1} = \frac{1}{22}(h_{t-1} + h_{t-2} + \ldots + h_{t-22})$$

As discussed in Corsi (2009), when using realized volatility the estimation of the HAR model is easy to implement and a standard OLS procedure provides consistent estimators. Then the one-day-ahead forecast of h_{T+1} can be generated as:

$$\hat{h}_{T+1} = \hat{\alpha} + \hat{\beta}_1 h_T + \hat{\beta}_2 h_{T-4:T} + \hat{\beta}_3 h_{T-21:T}$$

We notice that using these four models the forecasted logarithmic volatilities, $\hat{h}_{T+1}$, are generated. While our interest is the one-day-ahead forecast realized volatility. We need to transform these predicted logarithmic volatilities to obtain forecasted volatilities. As first discussed in Granger and Newbold (1976), and then widely accepted in the literature, directly taking the exponential of $\hat{h}_{T+1}$ would induce bias, hence would result in poor forecasts. Some transformation is necessary.

The one-day-ahead forecast error of h_{T+1} is calculated by:

$$e_{t+1} = h_{T+1} - \hat{h}_{T+1} = h_{T+1} - E(h_{T+1|T}) = \sigma v_{T+1}$$

With the bias being taken into account, the one-day-ahead forecasted realized volatility, $\hat{RV}_{T+1}$, is transformed by:

$$\hat{RV}_{T+1} = E(RV_{T+1|T}) = E(e^{h_{T+1|T}}) = E(e^{\hat{h}_{T+1}+\sigma v_{T+1}}) = E(e^{\hat{h}_{T+1}})E(e^{\sigma v_{T+1}})$$

where $E(e^{\sigma v_{T+1}})$ is the moment generating function of disturbance v.

Under the assumption that v follows a standard normal distribution, the moment generating function is given by $e^{\frac{\sigma^2}{2}}$.

Therefore, the one-day-ahead forecasted realized volatility can be computed as:

$$\hat{RV}_{T+1} = exp(\hat{h}_{T+1} + \frac{\sigma^2}{2})$$

where σ^2 is the variance parameter. As it is unknown, we use the sample estimate $\hat{\sigma}^2_T$ to replace σ^2.

3.4.2 Forecasting Evaluation

Testing the null: competing models provide equally accurate forecasts

Obviously the accuracy of volatility forecasting is of great importance in finance. This importance has sparked a lot of interest in both evaluating and improving volatility forecasting performance. The crucial object in evaluating forecasting accuracy is the loss function, $L(y_{t+k}, \hat{y}_{t+k|t})$, in which y_{t+k} denotes the realized value at $t+k$ and $\hat{y}_{t+k|t}$ refers to the forecasted value of $t+k$ based on the information set at t . The loss function measures the "loss" or "cost" associated with various pairs of forecasts and realizations. The smaller the loss, the better the forecast. Usually the loss function is defined as a function of the k-step-ahead forecast error which is the difference between the realized and the forecasted value , i.e. $e_{t+k,t} = y_{t+k} - \hat{y}_{t+k|t}$, or percent error, i.e. $p_{t+k} = (y_{t+k} - \hat{y}_{t+k|t})/y_{t+k}$. For example, the mean error, $ME = \frac{1}{T}\sum_{t=1}^{T} e_{t+k,t}$, or mean percent error, $MPE = \frac{1}{T}\sum_{t=1}^{T} p_{t+k,t}$, provide measures of bias. However, as Diebold and Lopez (1996) pointed out, the shape of the loss function is crucial in measuring forecast accuracy. Different loss functions would result in different rankings of forecast performance. In addition, most commonly used accuracy measures involve strong assumptions such as the loss function is quadratic, symmetric, or the forecast errors are zero mean, Gaussian, serially uncorrelated or contemporaneously uncorrelated, so forth. These assumptions may not be realistic.

Alternatively, Diebold and Mariano (1995) proposed some widely applicable tests. In their tests, the null hypothesis states that there is no difference in the accuracy of two competing forecasts. These tests are based directly on predictive performance, the loss functions can be neither quadratic nor symmetric, the forecast errors can be non-gaussian, non-zero mean, or serially correlated. The basic procedure of these tests is, let the time t loss be an arbitrary function of the realization and the forecast, $g(e_{t+k,t}) = g(y_{t+k}, \hat{y}_{t+k|t})$, for two competing models, the null hypothesis is that forecasts from these two models are equally accurate, i.e. $H_0 : E[g(e_{it+k,t})] = E[g(e_{jt+k,t})]$, or $E[d_{t+k}] = 0$, where $d_{t+k} = g(e_{it+k,t}) - g(e_{jt+k,t})$ refers to the loss differential. In fact, the null hypothesis is equivalent to that when the population mean of the loss-differential series is 0. If the null is rejected, the forecasting method that yields the smallest loss is preferred. As the Diebold and Mariano (1995) tests relax strong assumptions, they are robust and widely applicable. In this subsection, we apply their tests to decide whether the competing models provide equally accurate forecasts.

1. Asymptotic Test

Assume the loss-differential series $\{d_t\}, t = 1, \ldots, T$, is covariance stationary and has short memory. According to central limit theory, when the sample size is large, the sample mean of the loss differential, $\bar{d}$, approximately follows a normal distribution. Let the asymptotic distribution of sample mean loss differential $\bar{d}$ be :

$$\sqrt{T}(\bar{d} - \mu) \longrightarrow_d N(0, 2\pi f_d(0))$$

where $\bar{d}$ and μ denote the sample and population means of d_t, respectively. $f_d(0) = \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty}\gamma_d(\tau)$ is the spectral density of the loss differential at frequency 0, with $\gamma_d(\tau) = E[(d_t - \mu)(d_{t-\tau} - \mu)]$ being the auto-covariance of d_t at τ.

The null hypothesis of two forecasts being equally accurate can be tested using a standard normal statistic:

$$S_1 = \frac{\bar{d}}{\sqrt{2\pi \hat{f}_d(0)/T}} \sim N(0, 1)$$

in which $2\pi\hat{f}_d(0)$ is a consistent estimate of $2\pi f_d(0)$ and can be calculated as:

$$2\pi\hat{f}_d(0) = \sum_{\tau=-(T-1)}^{T-1} 1(\tfrac{\tau}{s(T)})\tfrac{1}{T}\sum_{t=|\tau|+1}^{T}(d_t - \bar{d})(d_{t-|\tau|} - \bar{d})$$

where $s(T) = h - 1$ is the truncation lag and $1(\frac{\tau}{s(T)})$ is the lag window, taking the value 1 if $|\frac{\tau}{s(T)}| \leq 1$, and 0 otherwise. As our study focuses on one-step-ahead forecasting, i.e. $h = 1$, hence $s(T) = 0$, as a result for any $\tau \neq 0$, $1(\frac{\tau}{s(T)}) = 0$ since $|\frac{\tau}{s(T)}| > 1$. In this situation, S_1 is calculated only when $\tau = 0$.

2. The Sign Test

Sometimes it is the case that only a few forecast-error observations are available. When the number of the forecast error is small, a finite-sample test such as the sign test can be conducted.

The null of the sign test is that the median loss differential is zero, i.e. $H_0 : med(g(e_{it}) - g(e_{jt})) = 0$. Clearly the sign test and the asymptotic test will coincide with each other when the loss differential is distributed symmetrically.

The test statistic is the following:

$$S_2 = \frac{\sum_{t=1}^{T} 1(d_t) - T/2}{\sqrt{T/4}} \sim_a N(0,1)$$

$1(d_t) = 1$ for $d_t > 0$, and 0 otherwise. This test is based on the assumption that $\{d_t\}|_{t=1}^{T}$ is i.i.d. distributed. Being a fair game, the probability of a positive loss differential is equal to 0.5. With the sample size being T, the number of positive loss differentials i follows a binomial distribution with mean $Tp = T/2$ and variance $Tp(1-p) = T/4$. As the sample size becomes larger, S_2 approaches to a standard normal distribution.

3. Wilcoxon's Signed-Rank Test

Another finite-sample test introduced by Diebold and Mariano (1995) is the Wilcoxon's signed-rank test. This test is a related distribution-free procedure, and assumes that the loss differential is i.i.d. and symmetric. The test statistic is the sum of the ranks of the absolute values of the positive observations,

$$W = \sum_{t=1}^{T} 1(d_t) rank(|d_t|)$$

The idea of this test is that if the distribution of the loss differential is symmetric around zero, a very large (or small) sum of the ranks of the absolute values of the positive observations would be unlikely. When the sample size is large, the standardized W follows a standard normal distribution:

$$S_3 = \frac{\sum_{t=1}^{T} 1(d_t) rank(|d_t|) - \frac{T(T+1)}{4}}{\sqrt{T(T+1)(2T+1)/24}} \sim_a N(0,1)$$

Forecasting Evaluation

Diebold and Mariano (1995)'s tests are valid for a very wide class of loss functions. However, the null of these tests is equally accurate forecasts, therefore the results do not tell which model provides more accurate forecast. Therefore we need to evaluate the point forecasts of each candidate model.

There are numerous criteria available for forecasting performance evaluation. However, as Diebold and Lopez (1996), Lopez (2001) etc. demonstrated, it is not obvious which measure is more appropriate. Rankings of forecast accuracy may be different across different measures. Therefore, instead of only focusing on one single measure, we employ several different measures in this study. In the volatility forecasting literature, the most commonly used measure is mean squared error, $MSE = \frac{1}{T}\sum_{t=1}^{T} e_{t+k,k}^2$, which depends on the second moment structure of the joint distribution of the realized and forecasted series.

$$E(MSE) = E[(y_{t+k} - \hat{y}_{t+k,t})^2] = var(y_{t+k} - \hat{y}_{t+k,t}) + (E[y_{t+k}] - E[\hat{y}_{t+k,t}])^2$$

Often the square root of MSE is used to preserve units yielding the root mean squared error, $RMSE$.

Another commonly used measure is mean absolute error, $MAE = \frac{1}{T}\sum_{t=1}^{T} |y_{t+k} - \hat{y}_{t+k,t}|$, which depends on the first moment of the joint distribution of the realized and forecasted series.

$$E(MAE) = E[|y_{t+k} - \hat{y}_{t+k,t}|]$$

Both MSE and MAE are mathematically simple, however, as Brailsford and Faff (1996) pointed out, these two measures are symmetrical which may not be realistic in practice.

Same as MSE and MAE, the Theil's-U statistic is constructed under the assumption of symmetry. But unlike the previous two measures, the random walk model is treated as a benchmark model in the Theil's-U statistic. Hence the Theil's-U provides a relative accuracy measure by comparing the forecasts with that using the random walk model. Intuitively, the Theil's-U is 1 when the random walk model is applied; the Theil's-U being less than 1 indicates that the model of interest provides better forecasts than the random walk model. The Theil's-U is invariant to scalar transformation.

As Bollerslev et al. (1994) discussed, the loss functions constructed under the assumption of symmetry do not penalize the method for zero and negative variance estimates which are clearly counterfactual. Therefore, a more natural loss function for volatility models may be the mean percentage squared errors, i.e. $MPSE = \frac{1}{T}\sum_{t=1}^{T}(y_{t+k} - \hat{y}_{t+k,t})^2/y_{t+k}^2$, or $QLIKE = \frac{1}{T}\sum_{t=1}^{T}(log(y_{t+k}) + \frac{\hat{y}_{t+k,t}}{y_{t+k}})$, the loss function implicit in the Gaussian likelihood. Patton (2006) proved, while many loss functions existed, $QLIKE$ was one of commonly used loss functions that belong to a family of loss functions robust to noise in the volatility proxy. Further, Patton and Sheppard (2007) showed that $QLIKE$, comparing with RMSE, had more power in differentiating between forecasts. In Hansen and Lunde (2005), they employed $QLIKE$ to evaluate the forecasting performance of various models.

Following most studies, we use four criteria to evaluate forecast accuracy of the competing models. Specifically, let σ_t^2 (or RV_t) be the true realized volatility (or variance) at time t, $\hat{\sigma}_t^2$ (or $\hat{RV}_t$) be the forecasted realized volatility. The four measures are defined as:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{\sigma}_t^2 - \sigma_t^2)^2} \tag{3.5}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|\hat{\sigma}_t^2 - \sigma_t^2| \tag{3.6}$$

$$Theil - U = \frac{\sum_{t=1}^{T}(\hat{\sigma}_t^2 - \sigma_t^2)^2}{\sum_{t=1}^{T}(\sigma_{t-1}^2 - \sigma_t^2)^2} \tag{3.7}$$

$$QLIKE = \frac{1}{T}\sum_{t=1}^{T}(log(\hat{\sigma}^2) + \frac{\sigma_t^2}{\hat{\sigma}^2}) \tag{3.8}$$

Empirical Results

In this subsection, we perform one-day-ahead realized volatility forecasts. The data sets are the same as we used in the estimation section: daily return and realized volatility series of S & P 500 index, CAD/USD, USD/GBP, and DEM/USD. For each set of data, we have to choose a period for in-sample estimation and a period for out-of-sample realized volatility forecasting. We consider two methods to obtain parameter estimates. One method is the so called "rolling sample" method. That is, we keep the sample size constant at T hence when the new data arrives, we discard the least recent data. Alternatively, to examine whether including richer information would significantly improve volatility forecasting performance, we keep increasing sample size, i.e. we use all the observations available in the estimation. We compare the estimation results as well as the forecasted volatility series and find these two methods provide very similar results. This finding suggests that recent information weighs much more in estimating and forecasting therefore we only report the results using "rolling sample".

The sample of S & P 500 index covers the period from November 12th, 1997 through July 28th, 2006, consisting of 2187 daily observations. We start in-sample estimation using data from November 12th, 1997 to July 28th, 2005, consequently the first out-of-sample daily forecasted volatility is obtained in July 29, 2005. We then re-estimate the models using data from November 13th, 1997 to July 29th, 2005 to obtain sequential forecasted volatility. The procedure is repeated 252 times. As a result, the forecasted volatility series from July 29th, 2005 through July 28th, 2006 are obtained.

The samples of the three exchange rates, CAD/USD, USD/GBP, and DEM/USD, cover the period from April 13th, 1998 through July 28th, 2006, consisting of 2126 observations. We again aim to obtain one year of one-day-ahead forecasted volatilities. Hence the in-sample estimation starts from April 13th, 1998 to July 28, 2005, we then plug the estimates into the models to obtain the forecasted volatility in July 29, 2005. We roll over the sample, and repeat the procedure for 258 times. The forecasted volatility series from July 29th, 2005 to July 28th, 2006 are obtained.

Figures 3.26-3.34 provide a graphical summary of the performances of the competing models over the entire out-of-sample period in forecasting realized volatilities. These figures illustrate the actual realized volatilities (solid line) and forecasted values from competing models (in the top figure, dotted lines are forecasts from M1, dashed lines are forecasts from M2; in the bottom figure, dots are forecasts from HAR, dashed lines are forecasts from SRM). Visual inspection suggests that, in general, all the competing models do a good job of capturing both the low-frequency and the high-frequency movements in the realized volatilities. We next proceed to a more thorough statistical evaluation of the forecasts.

3.4.3 Results for null: two models provide equally accurate forecasts

As pointed out by Diebold and Mariano (1995), even optimal forecast errors are serially correlated. To examine the level of serial correlation of forecast errors, we compute the ACF values and report these values in Table 3.9-3.12. We find no matter which data set we use, the autocorrelation of forecast error series is significant at lag one and starts to diminish from lag two. This finding is consistent with Diebold and Mariano (1995)'s results. We then apply the well known Diebold and Mariano (1995) forecast comparison tests, including the asymptotic test, the sign test and the Wilcoxon's signed-rank test. As discussed in Diebold and Mariano (1995), the loss function $g(e_{t+k,t})$ can be an arbitrary function of the forecast error $e_{t+k,t}$, without loss of generality, we set $g(e_{t+k,t}) = e_{t+k,t}$, consequently, the loss differential $d_{t+k} = e_{1t+k,t} - e_{2t+k,t}$.

Table 3.13 report the test statistics of equal forecast accuracy. We report the values of $S1$ (the asymptotic test), $S2$ (the sign test), $S3$ (the Wilcoxon's signed-rank test) for each pair of models for S & P 500 index and exchange rates. For example, in column 3, we report all three testing results for the null hypothesis: HAR model and M1 provide equally accurate forecasting, column 4 reports results for the null: HAR model and M2 provide equally accurate forecasting, so forth.

We find all the values of S_1 from the asymptotic test are very large, suggesting that null hypothesis is rejected at .01 percent confidence level. The results from the sign test along with those from the Wilcoxon's signed-rank test confirm that the null hypothesis should be rejected. Based on these findings, we draw the conclusion that all the four competing models provide unequally accurate forecasts.

3.4.4 Forecasting Evaluation

Diebold and Mariano (1995) tests provide us information that the competing models provide unequally accurate forecasts. We then extend our study by employing four different loss functions to evaluate the forecasting performance. Table 3.14-3.17 report the value along with ranking of all four competing models under RMSE, MAE, $Theil-U$, and $QLIKE$.

From the examination of these tables, we find, in general, the rankings are consistent when different RV_s are forecasted. On the other hand, the values and rankings vary across different criteria and different data sets. For example, when S & P 500 data are used, RMSE statistic indicates that Model 1 ranks the second, while $Theil-U$ statistic suggests that this model provides the poorest forecasts among four competing models. Under RMSE, when S & P 500 data are used, HAR ranks the first, Model 1 second, simple regression model third, and Model 2 ranks fourth, whereas when currency exchange rate data are used, HAR ranks first, and all the other three models rank second. For S & P 500 index, the marginal differences between Model 1 and simple regression model are generally small, while the differences between Model 1 and Model 2 are relatively much larger. For foreign exchange rates, the marginal differences among Model 1, Model 2, and simple regression model are very small, and sometimes there is no difference.

It is noted that no matter whether we use S & P 500 index data or currency exchange rate data, the HAR model always provides the best point forecasts among the four competing models, hence ranks the first according to all criteria. This finding is consistent with other studies, such as Corsi (2009), Andersen, Bollerslev, and Diebold (2007), implying that time series model capturing long-memory feature of volatility process is able to provide accurate forecasts. Moreover, we find sometimes the marginal differences between HAR and the second position are very large. For example, for RV1 of S & P 500 index, the marginal difference in the $Theil - U$ between HAR and second position is approximately 100 percent. The large marginal difference between HAR and other models indicates that the time series model capturing long-memory feature could significantly improve the out-of-sample volatility forecasting performance.

As we mentioned, we are interested in investigating whether capturing leverage (or feedback) effect as well high-kurtosis would improve the out-of-sample volatility forecasting performance. Comparing the forecasting performances of Model 1, Model 2, and the simple regression model, we find that the results vary with different data sets. Specifically, when S & P 500 index data are used, we find according to most criteria (except that for $Theil - U$), Model 1 in which leverage effect is captured provides better forecasting performance than the simple regression model. However, capturing feedback effect does not seem to help in point forecasting as Model 2 provides poorer forecasts than the simple regression model. When currency exchange rates are used, generally neither Model 1 nor Model 2 can provide better performance than the simple regression model. This finding suggests that the forecasts performance of both Model 1 and Model 2 are sensitive to different data set. Overall, it seems that capturing feedback effect and high kurtosis does not improve the accuracy of volatility point forecasts. This finding is consistent with Corsi, Mittnik, Pigorsch, and Pigorsch (2008).

We are also interested in comparing the forecasting performance of Model 1 and Model 2, as our primary interest is to investigate the estimation and forecasting performance of these two models. We find the results are different across different data sets and different criteria. For S & P 500 index data, Model 1 outperforms Model 2 under RMSE, MAE, and $QLIKE$, however, according to $Theil - U$, Model 2 always provide better forecast. In general, the marginal difference between these two models are not negligible under all criteria. For the exchange rates, we notice that in general these two models rank the same. According to some criteria, Model 1 performs better, however, the marginal differences are extremely small, hence not significant.

In summary, the forecasting evaluations show that modeling the long-memory behavior of volatility results in an improvement in forecast accuracy, in contrast, permitting leptokurtosis and an asymmetric relationship between return and volatility processes in the model does not seem to help significantly in point forecasting.

3.5 Conclusion and Future Research

In Chapter 2, we focused on investigating the estimation of discrete-time SV models using traditional methods including FIML and 3SLS when volatility is observed. In this chapter, we extended our study to empirically investigate the estimation and forecasting performance of SV models using realized volatility. We used high frequency data to construct daily realized volatilities of both equity market index and exchange rates, and applied them in the estimation of Model 1 and Model 2. The evidence showed that, in general, the estimates were stable. We then focused on the one-day-ahead volatility point forecasts.

We considered four different models to examine whether allowing asymmetric relationships between return and volatility and leptokurtosis, or modeling the long-memory behavior of volatility, would result in an improvement in forecast accuracy. The results of Diebold and Mariano (1995)'s tests rejected the null that two models provided equally accurate forecasts. The results of forecasting measures generally confirmed these tests. However, allowing asymmetric behavior and leptokurtosis did not seem to improve point forecasts, whereas modeling long-memory behavior seemed do. Moreover, the forecasts performance of Model 1 and Model 2 are sensitive to different data set.

For continuous time SV models, the lagged inter-temporal and contemporaneous dependencies will coincide. However, estimation of continuous-time SV model is another challenge. Although various methods have been developed, in general, the estimation is computationally demanding. In Chapter 4, we will use realized volatility as well as model-free implied volatility, which is computed from option prices, as two different proxies for the latent volatility, and apply a consistent-approximate maximum likelihood approach in the estimation the continuous-time SV model.

3.6 Appendix

:

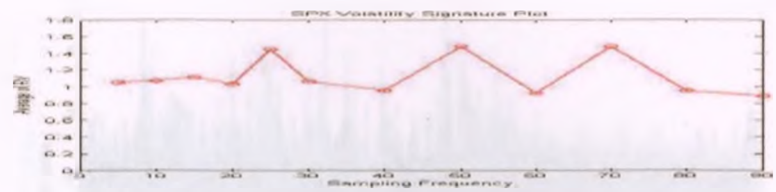

Figure 3.1: SPX Volatility Signature Plot

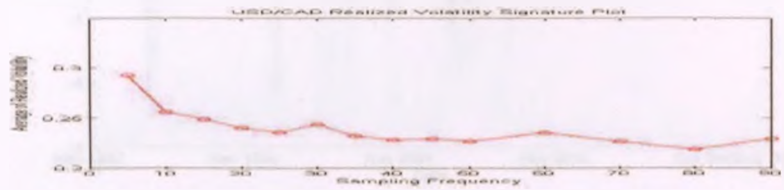

Figure 3.2: CAD Volatility Signature Plot

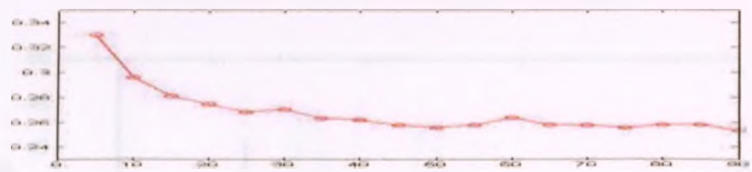

Figure 3.3: GBP Volatility Signature Plot

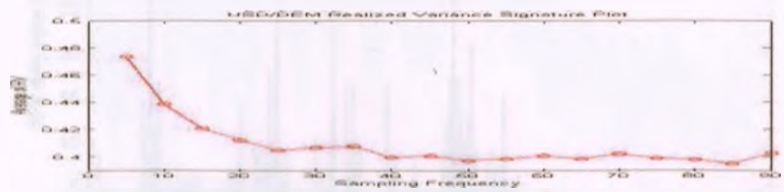

Figure 3.4: DEM Volatility Signature Plot

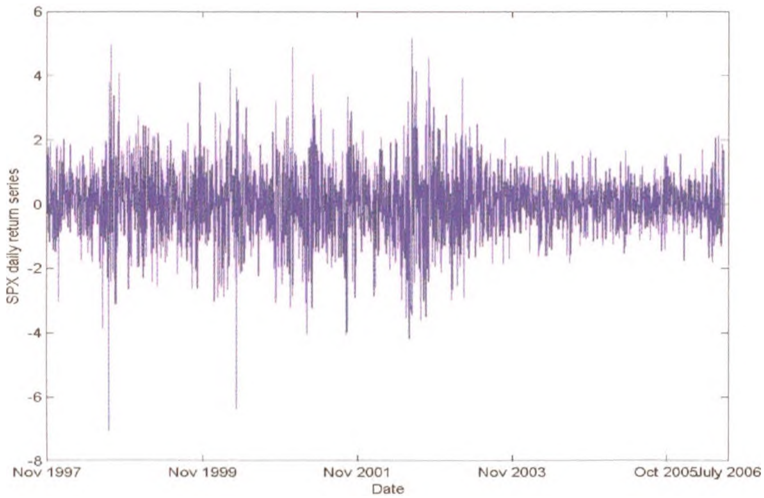

Figure 3.5: SPX Daily Return Plot

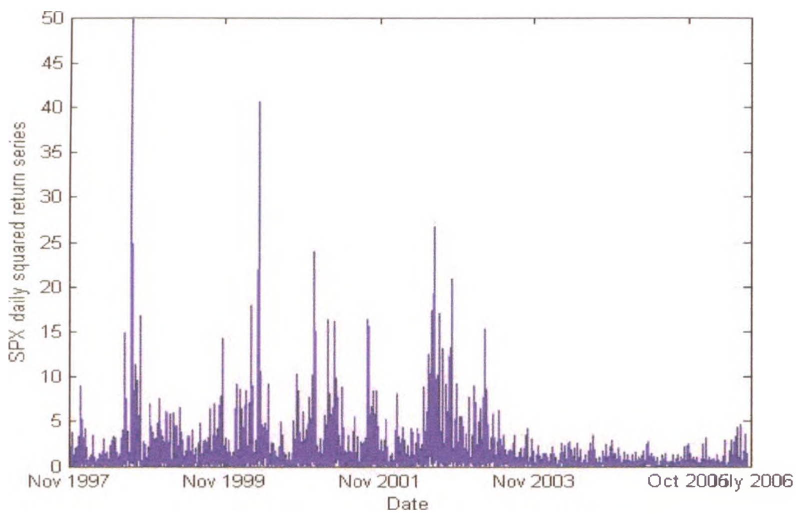

Figure 3.6: SPX Daily Squared Return Plot

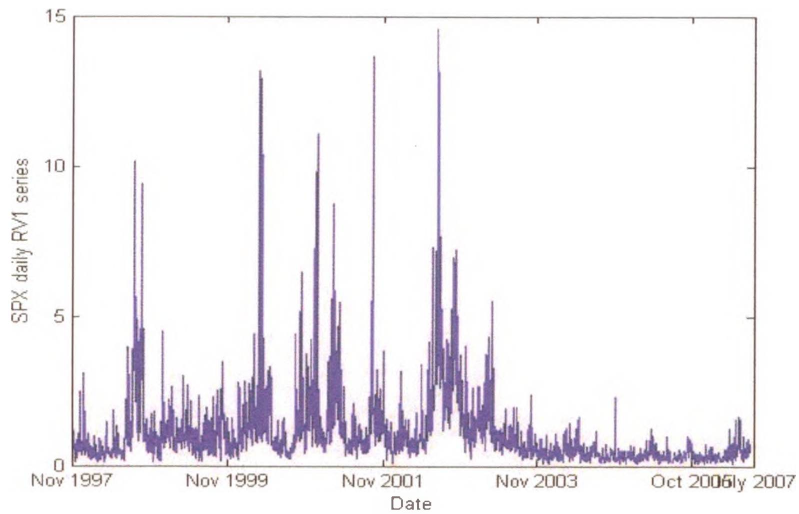

Figure 3.7: SPX Daily RV1 Plot

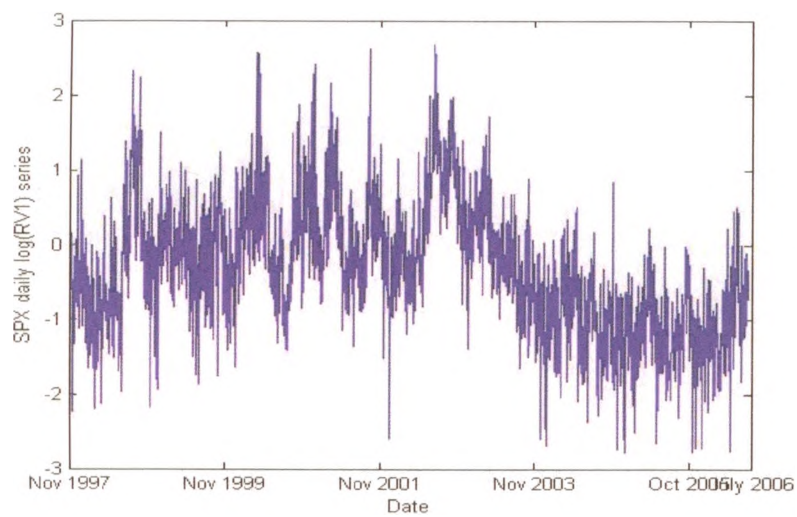

Figure 3.8: SPX Daily ln(RV1) Plot

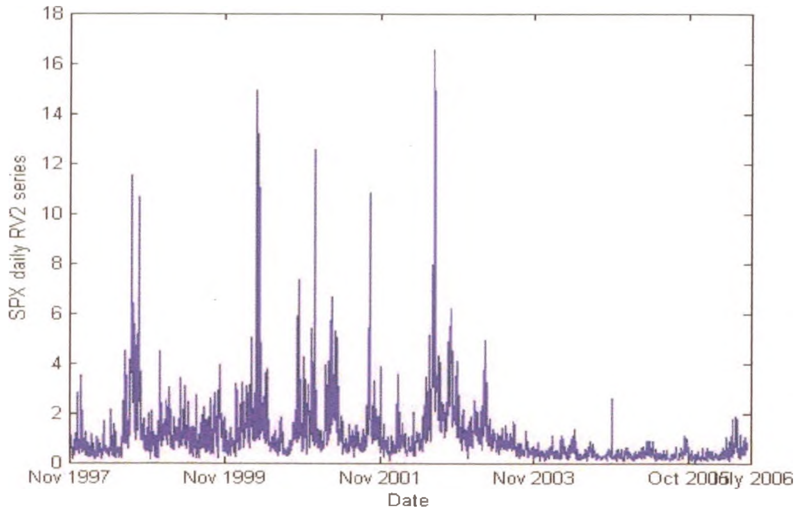

Figure 3.9: SPX Daily RV2 Plot

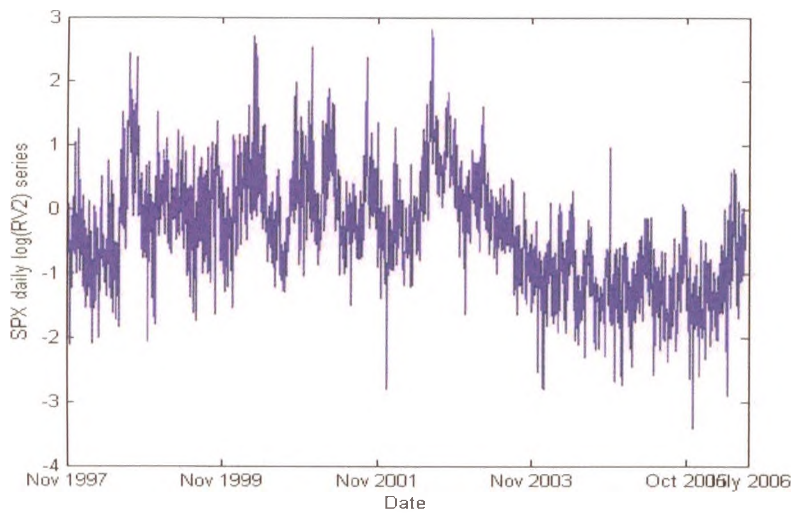

Figure 3.10: SPX Daily ln(RV2) Plot

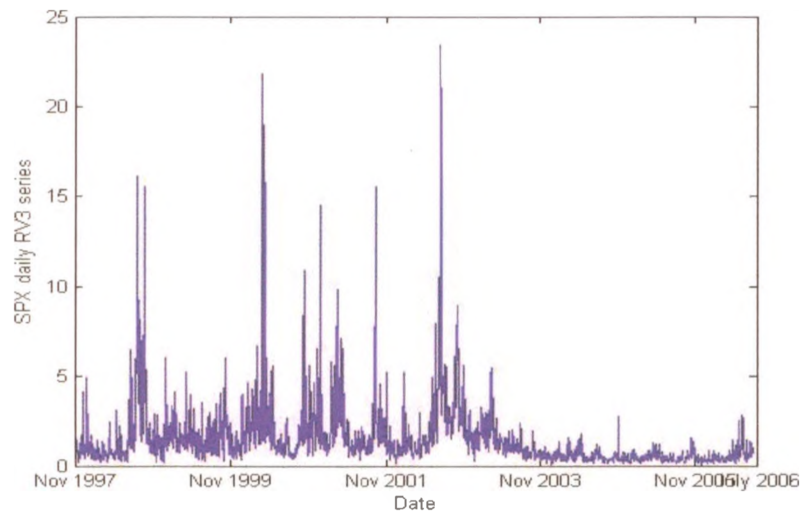

Figure 3.11: SPX Daily RV3 Plot

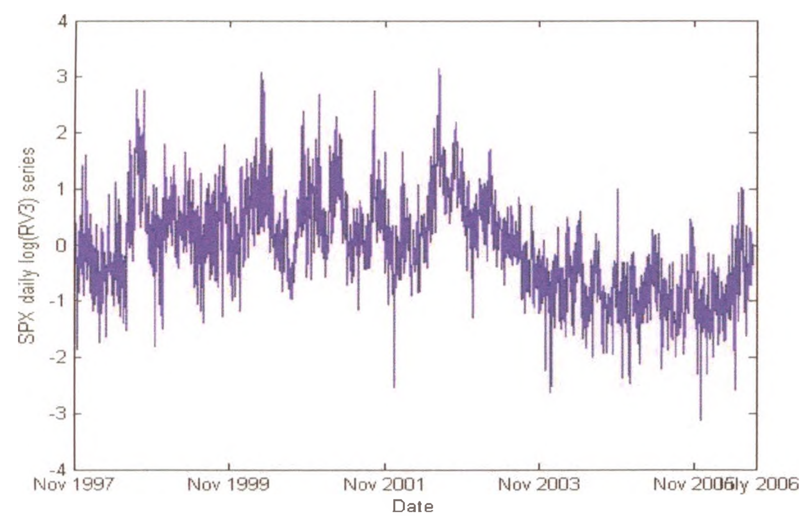

Figure 3.12: SPX Daily ln(RV3) Plot

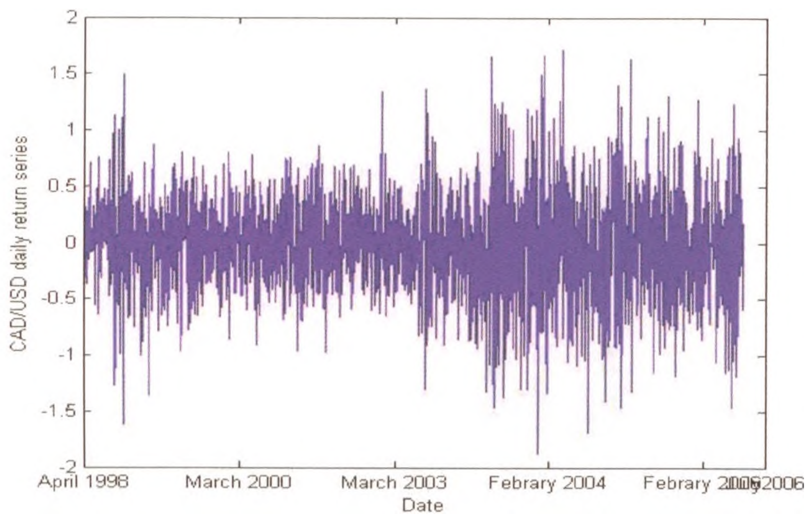

Figure 3.13: CAD Daily Return Plot

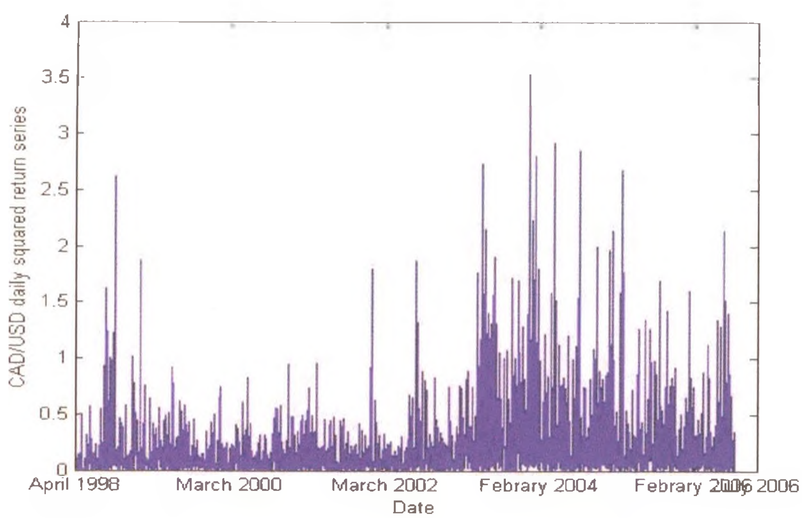

Figure 3.14: CAD Daily Squared Return Plot

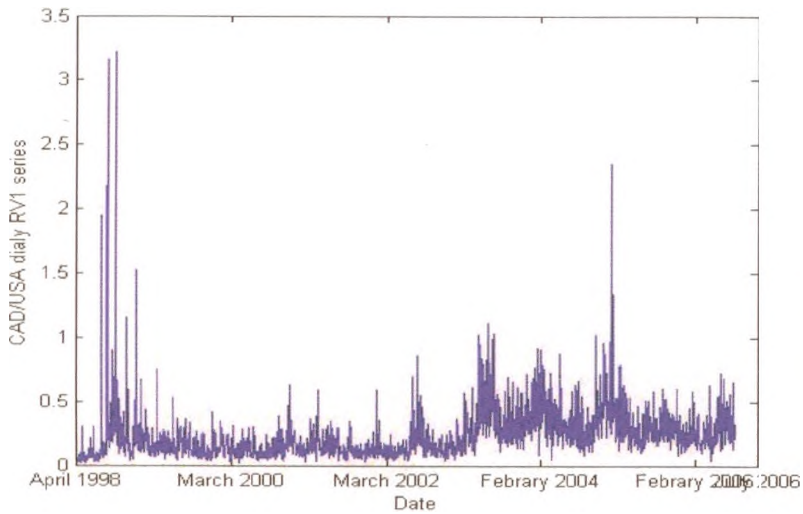

Figure 3.15: CAD Daily RV1 Plot

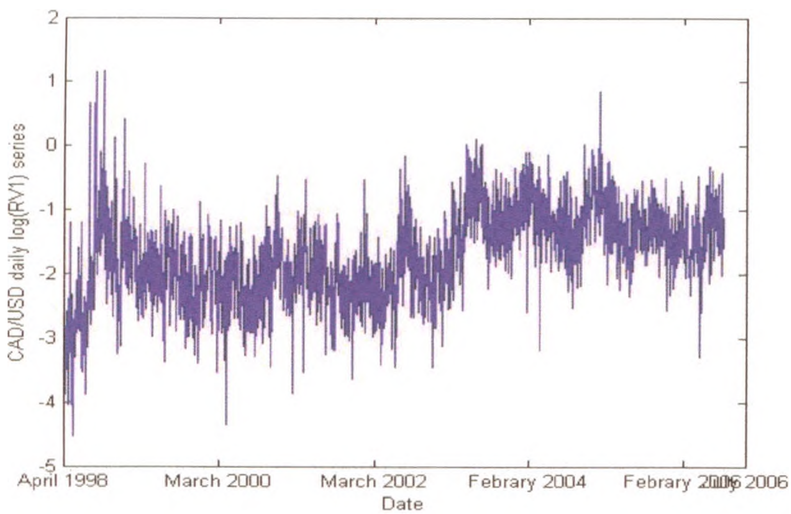

Figure 3.16: CAD Daily ln(RV1) Plot

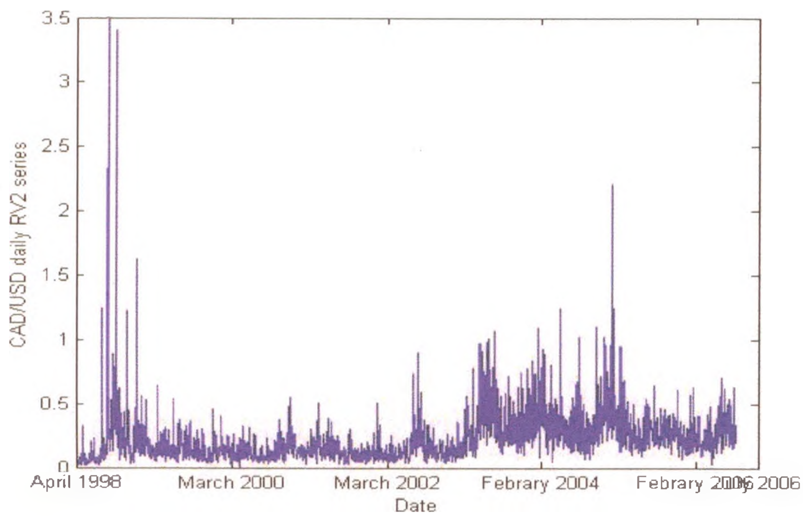

Figure 3.17: CAD Daily RV2 Plot

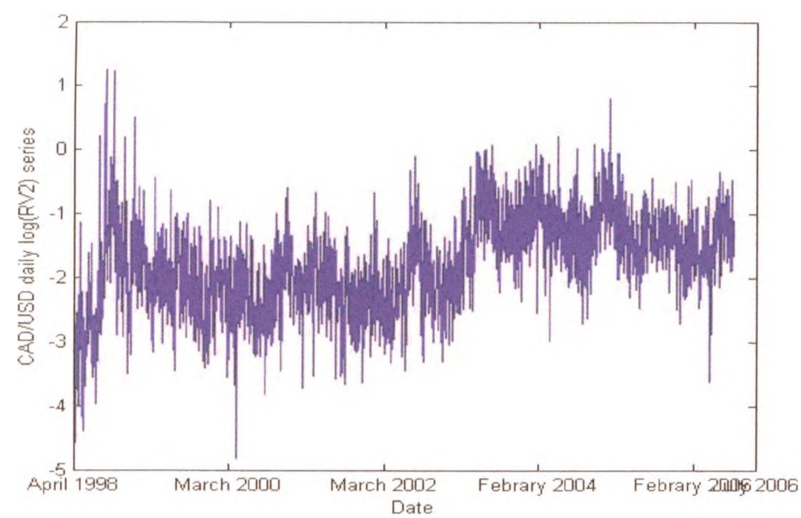

Figure 3.18: CAD Daily ln(RV2) Plot

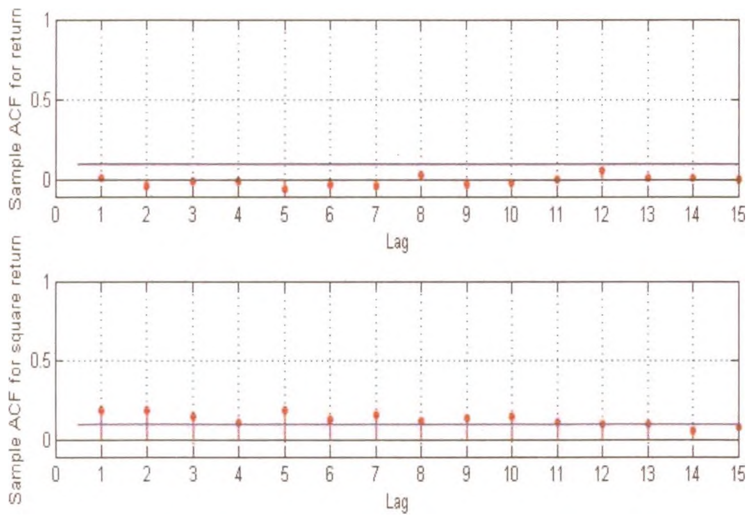

Figure 3.19: SPX ACF plot of Daily Return and Squared Return

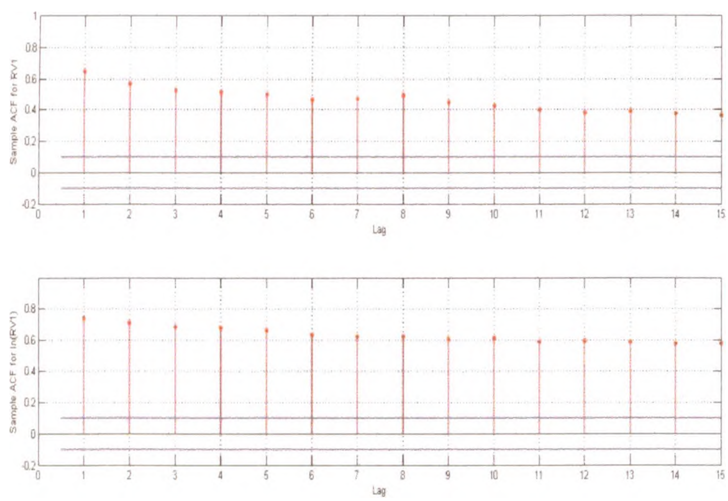

Figure 3.20: SPX ACF plot of Daily RV1 and ln(RV1)

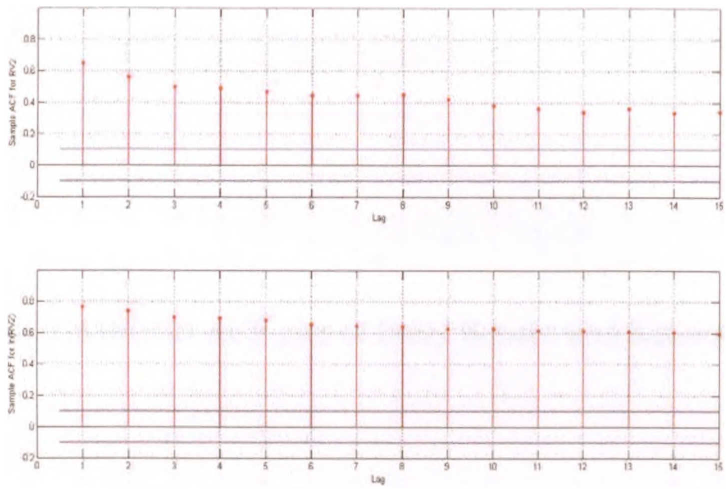

Figure 3.21: SPX ACF plot of Daily RV2 and ln(RV2)

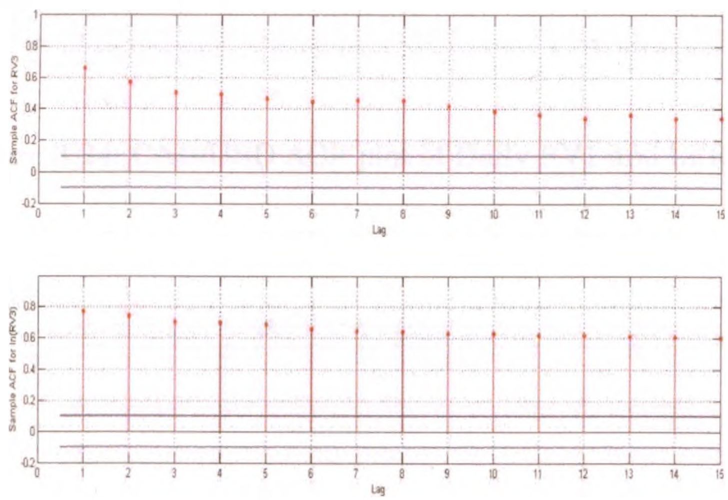

Figure 3.22: SPX ACF plot of Daily RV3 and ln(RV3)

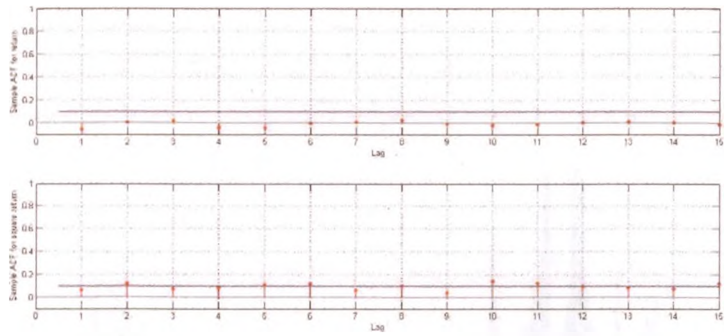

Figure 3.23: CAD/USD ACF plot of Daily Return and Squared Return

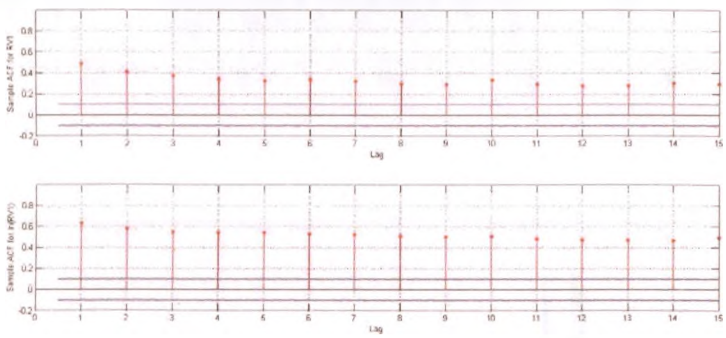

Figure 3.24: CAD/USD ACF plot of Daily RV1 and ln(RV1)

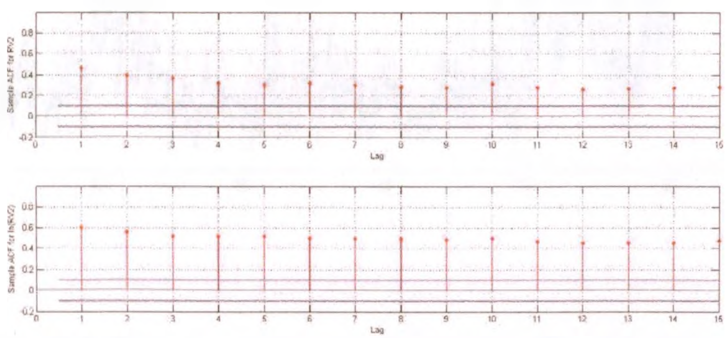

Figure 3.25: CAD/USD ACF plot of Daily RV2 and ln(RV2)

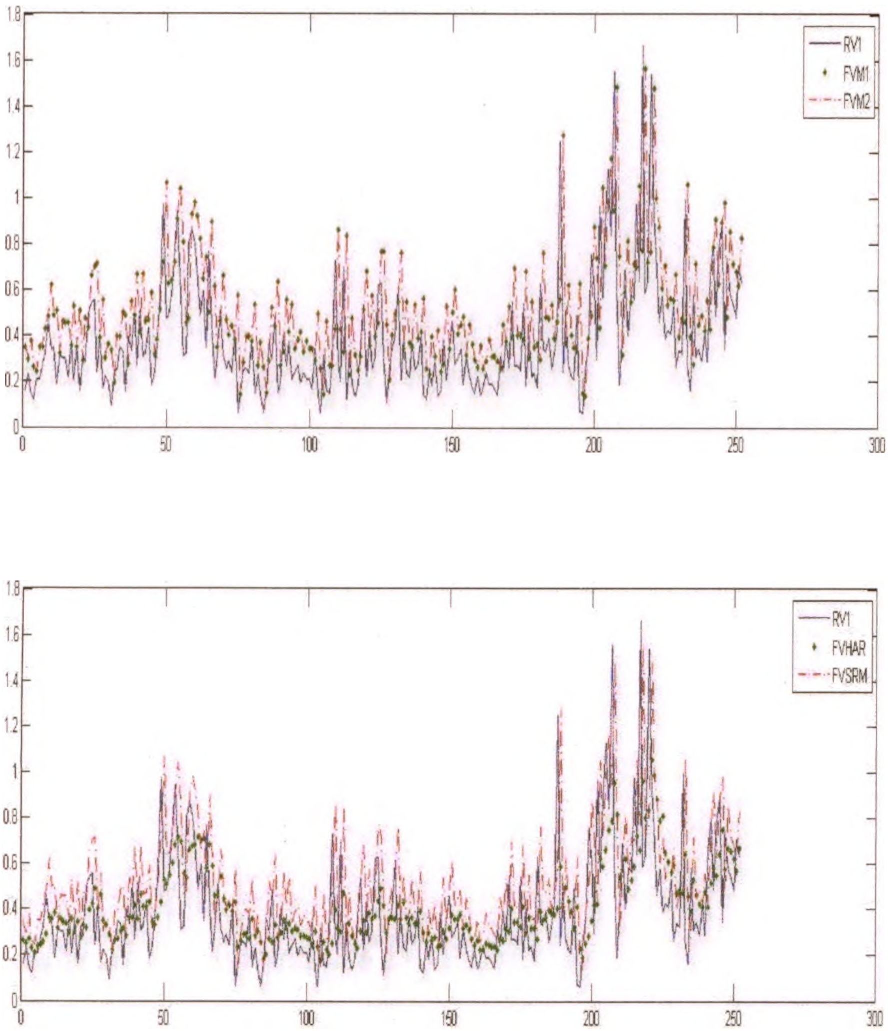

Figure 3.26: SPX RV1, Forecasted Volatility by M1, M2, SRM, HAR plot

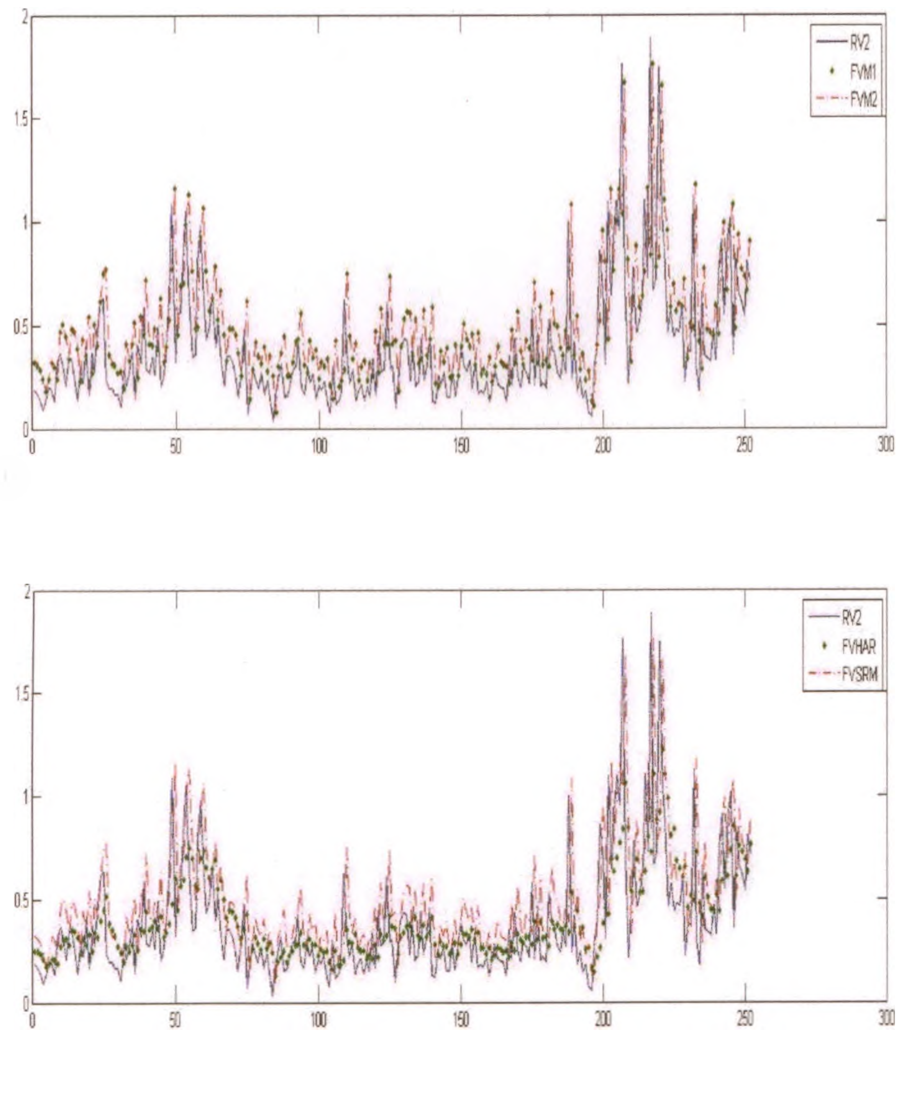

Figure 3.27: SPX RV2, Forecasted Volatility by M1, M2, SRM, HAR plot

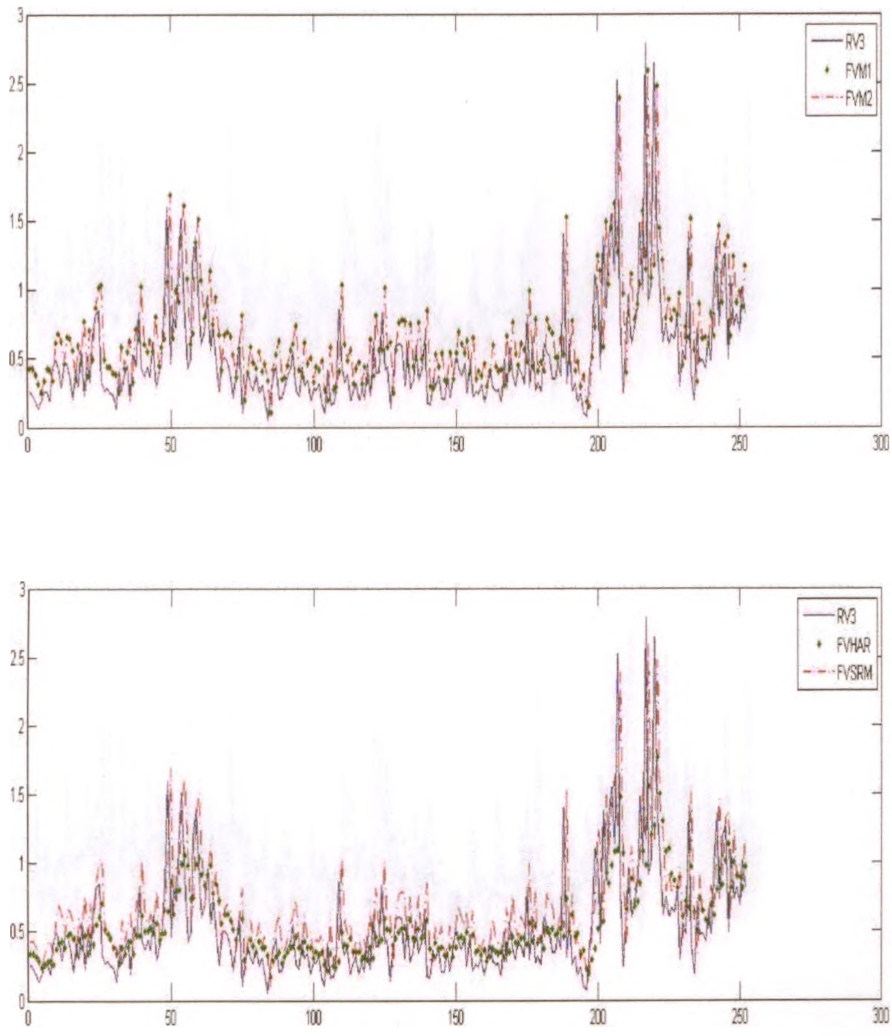

Figure 3.28: SPX RV3, Forecasted Volatility by M1, M2, SRM, HAR plot

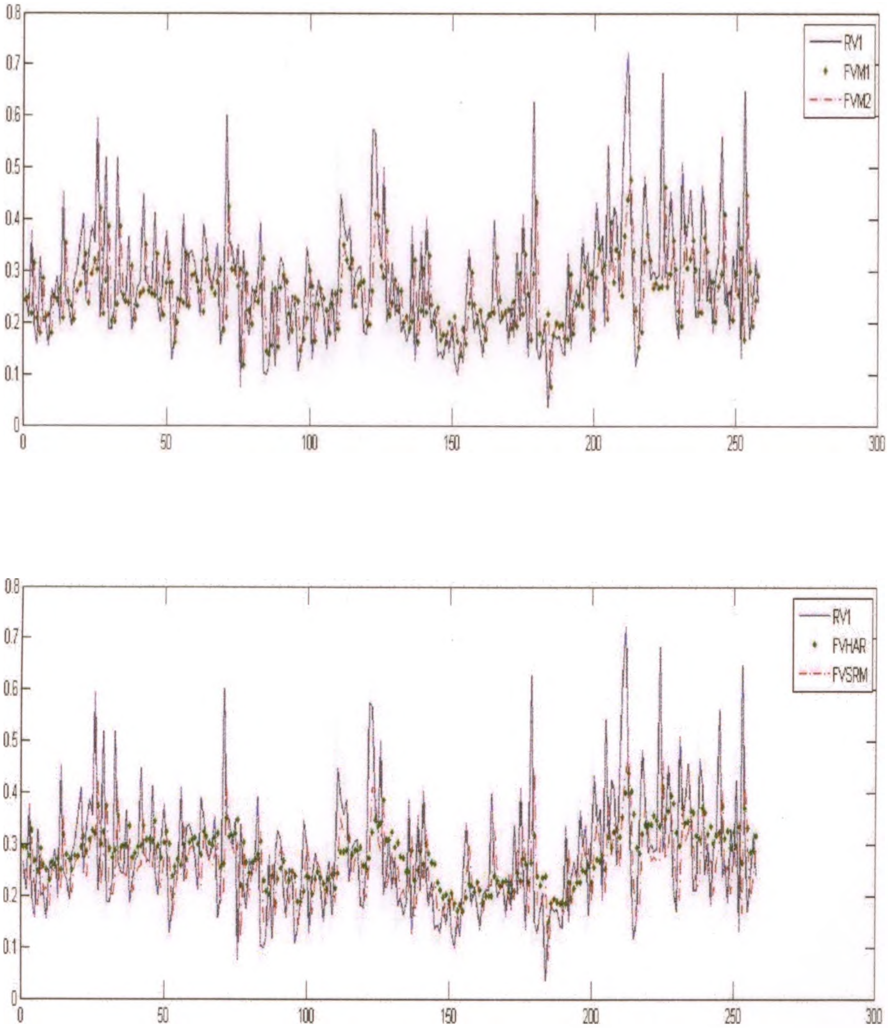

Figure 3.29: CAD/USD RV1, Forecasted Volatility by M1, M2, SRM, HAR plot

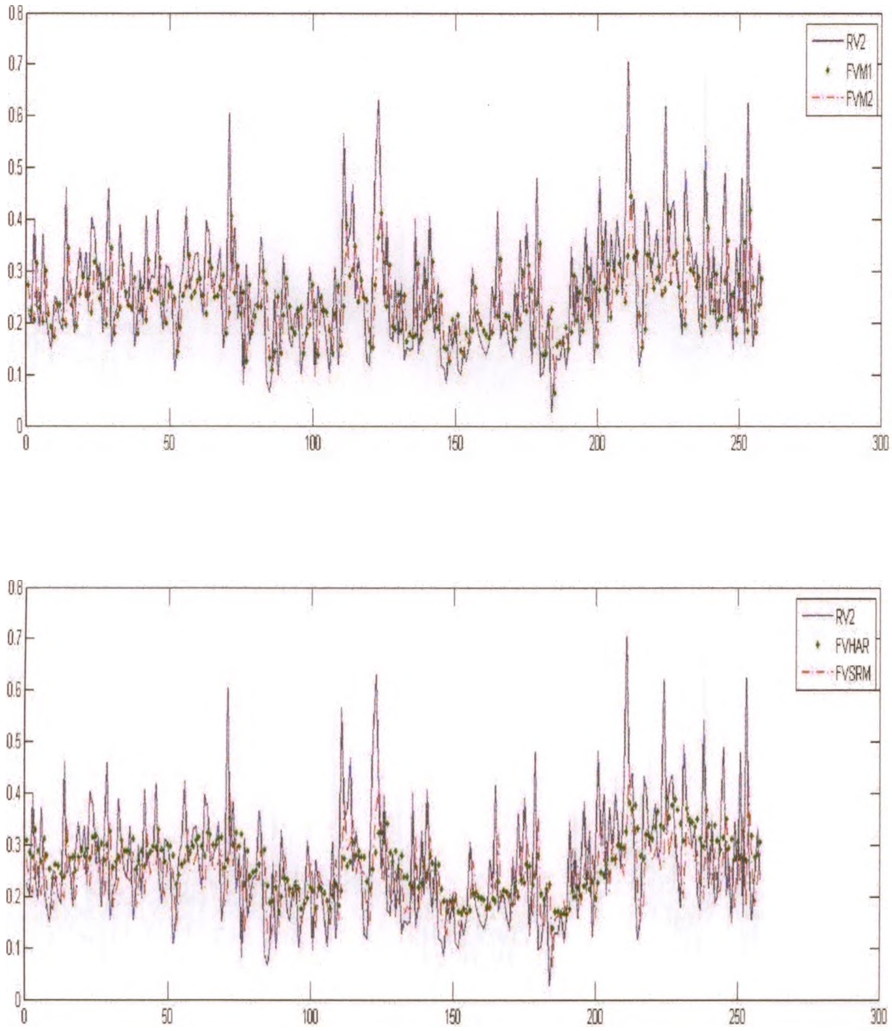

Figure 3.30: CAD/USD RV2, Forecasted Volatility by M1, M2, SRM, HAR plot

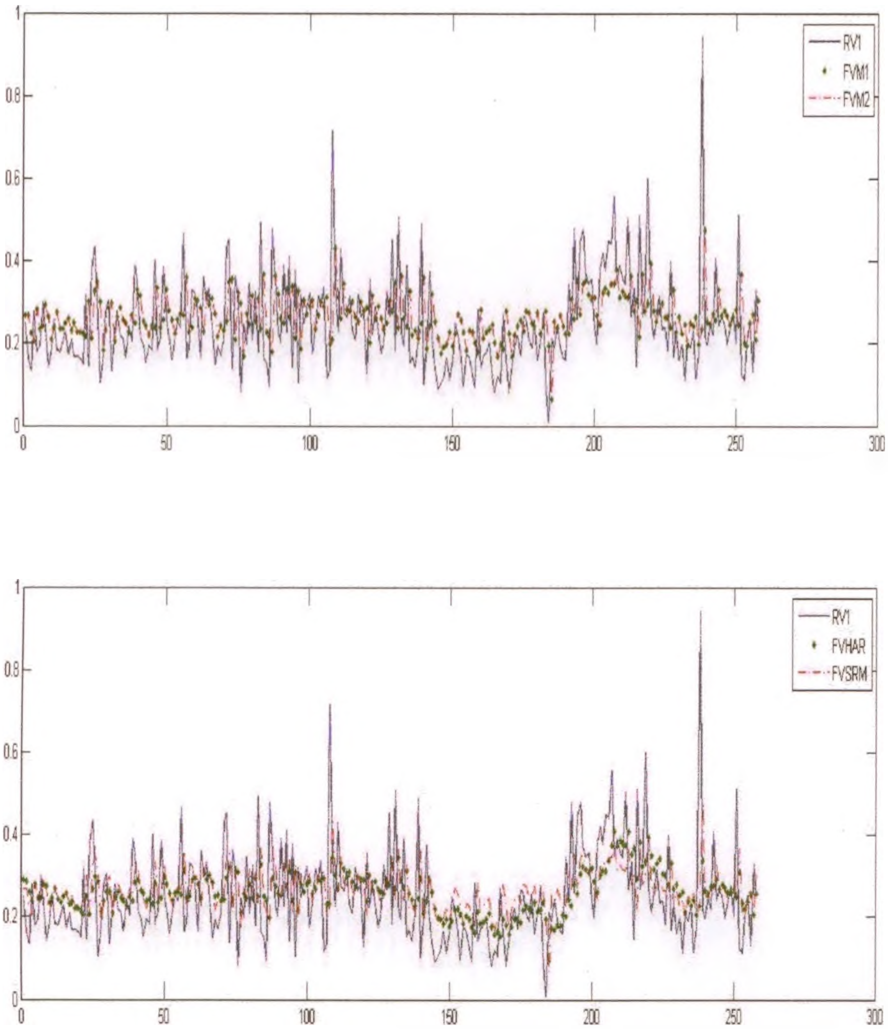

Figure 3.31: USD/GBP RV1, Forecasted Volatility by M1, M2, SRM, HAR plot

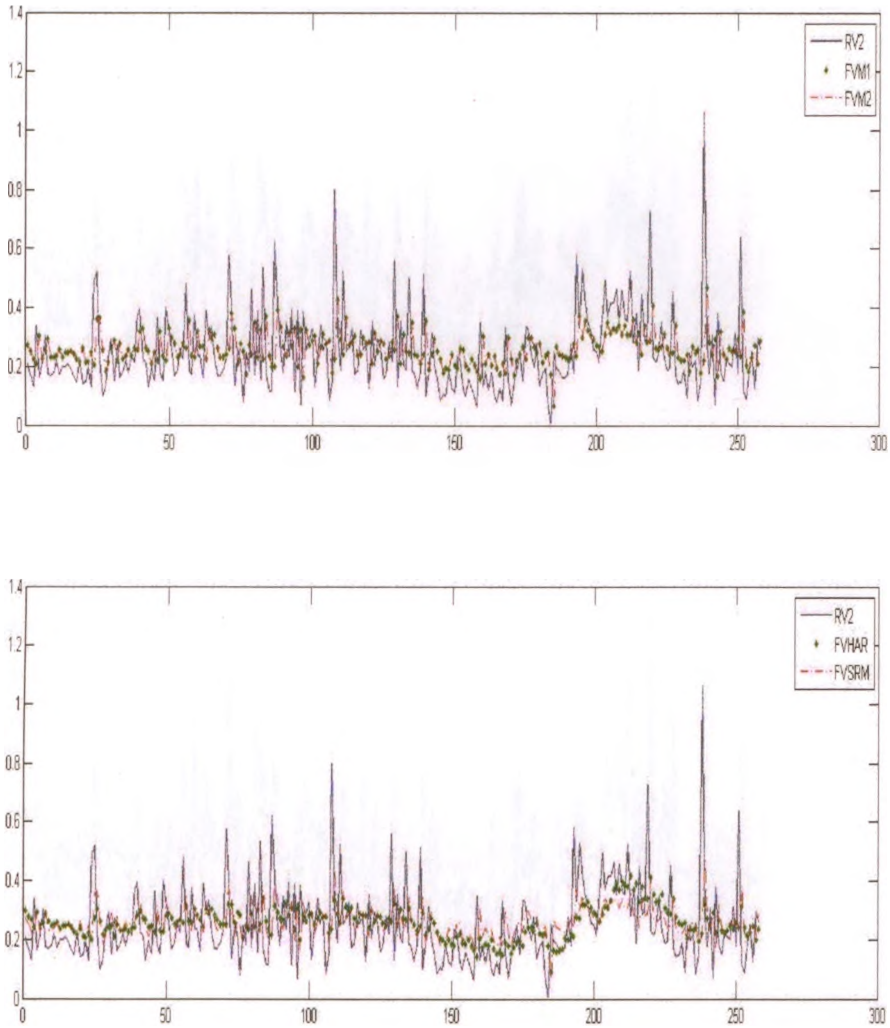

Figure 3.32: USD/GBP RV2, Forecasted Volatility by M1, M2, SRM, HAR plot

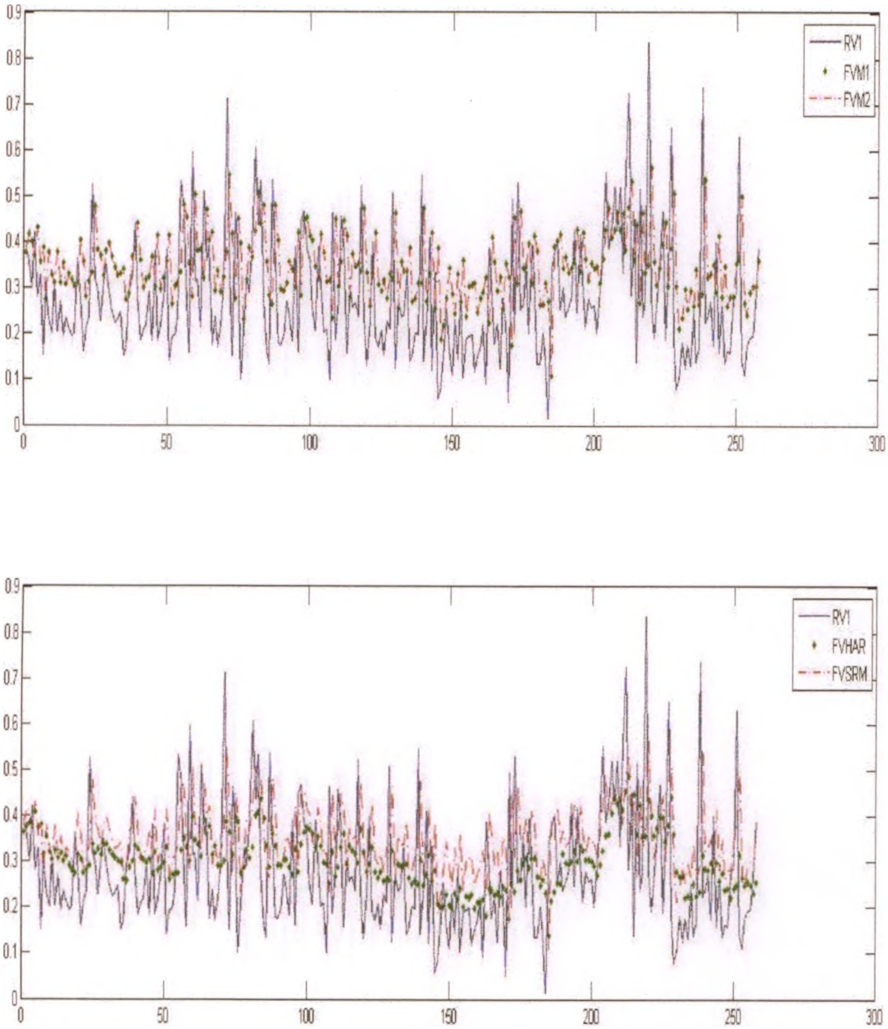

Figure 3.33: DEM/USD RV1, Forecasted Volatility by M1, M2, SRM, HAR plot

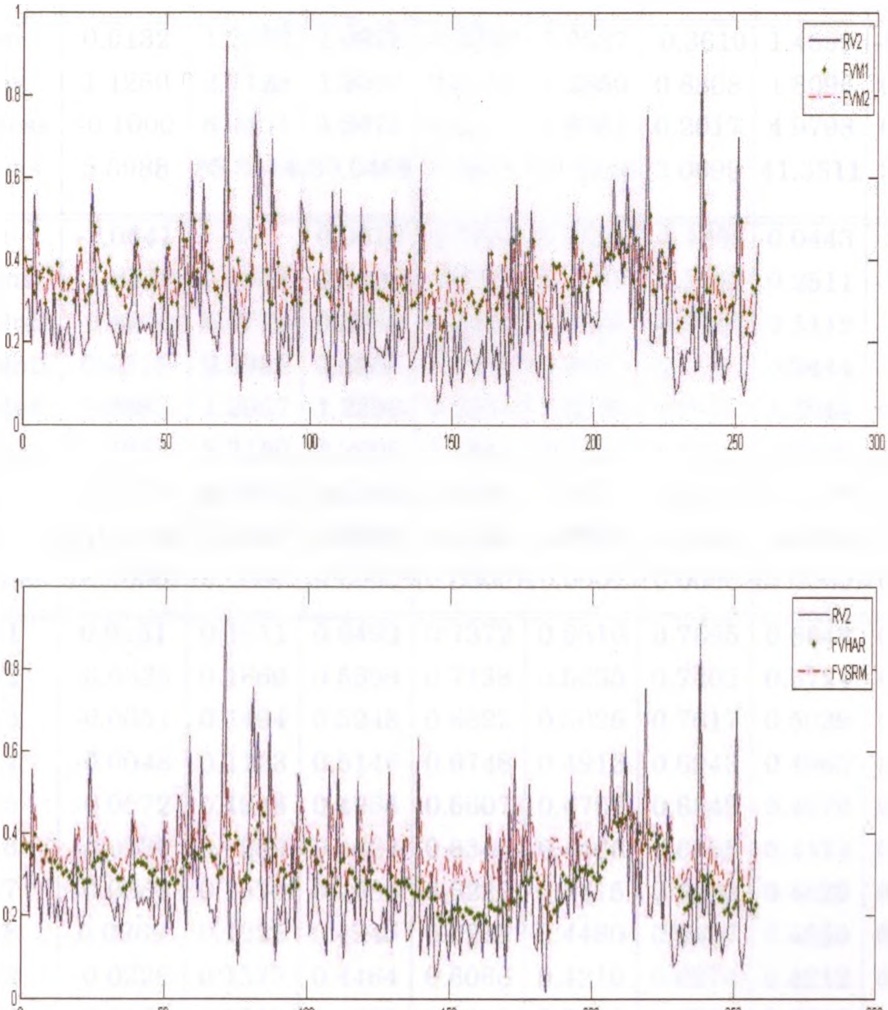

Figure 3.34: DEM/USD RV2, Forecasted Volatility by M1, M2, SRM, HAR plot

Table 3.1: Summary statistics of return and volatility time series for S& P 500 during the period of November 12th 1997 through July 28th 2006

	x_t	x_t^2	RV1	ln(RV1)	RV2	ln(RV2)	RV3	ln(RV3)
Mean	0.0132	1.2652	1.0829	-0.3346	1.0527	-0.3610	1.4599	-0.0397
St.Dev.	1.1250	2.7122	1.3000	0.8779	1.2850	0.8808	1.8096	0.8847
Skewness	-0.1000	6.9801	4.2475	0.2618	4.9081	0.2017	4.9798	0.2136
Kurtosis	5.5988	85.3414	30.0465	3.0625	40.4126	3.0699	41.3811	3.1004
Min	-7.0641	0	0.0616	-2.7871	0.0325	-3.4265	0.0443	-3.1163
5pct Qntl.	-1.8328	0.0033	0.1881	-1.6709	0.1793	-1.7187	0.2511	-1.3819
25pct Qntl.	-0.6200	0.0770	0.3865	-0.9506	0.3706	-0.9927	0.5112	-0.6709
50pct Qntl.	0.0513	0.3931	0.6966	-0.3615	0.6893	-0.3721	0.9444	-0.0572
75pct Qntl.	0.6383	1.2957	1.2396	0.2148	1.2229	0.2012	1.6942	0.5272
95pct Qntl.	1.7633	5.3159	3.2695	1.1846	3.1350	1.1426	4.3199	1.4632
Max	5.1720	49.9015	14.5626	2.6785	16.5271	2.8050	23.4172	3.1535
JB Stat.	616.6740	634400	136060	15.2198	147960	18.5032	143000	17.4705
(p-value)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0002)	(0.0000)	(0.0000)	(0.0000)
ACF1	0.0151	0.1911	0.6493	0.7372	0.6510	0.7695	0.6644	0.7755
ACF2	-0.0325	0.1869	0.5699	0.7138	0.5635	0.7402	0.5724	0.9434
ACF3	-0.0051	0.1494	0.5248	0.6822	0.5026	0.7017	0.5039	0.7048
ACF4	-0.0048	0.1123	0.5146	0.6748	0.4913	0.6943	0.4965	0.6977
ACF5	-0.0572	0.1946	0.4954	0.6607	0.4707	0.6842	0.4670	0.6859
ACF6	-0.0239	0.1289	0.4624	0.6345	0.4466	0.6565	0.4473	0.6589
ACF7	-0.0353	0.1576	0.4686	0.6229	0.4475	0.6482	0.4529	0.6488
ACF8	0.0269	0.1225	0.4945	0.6218	0.4480	0.6407	0.4550	0.6423
ACF9	-0.0226	0.1373	0.4464	0.6066	0.4210	0.6274	0.4212	0.6286
ACF10	-0.0175	0.1549	0.4235	0.6091	0.3806	0.6303	0.3857	0.6324
ACF11	-0.0023	0.1113	0.3993	0.5894	0.3625	0.6149	0.3638	0.6161
ACF12	0.0651	0.1061	0.3828	0.5908	0.3387	0.6144	0.3415	0.6165
ACF13	0.0144	0.0992	0.3921	0.5890	0.3632	0.6106	0.3622	0.6135
ACF14	0.0091	0.0651	0.3759	0.5790	0.3353	0.6048	0.3389	0.6081
ACF15	0.0007	0.0849	0.3640	0.5750	0.3398	0.5978	0.3399	0.6004

Table 3.2: Summary statistics of return and volatility time series for CAD/USD during the period of April 13th 1998 through July 28th 2006

	x_t	x_t^2	RV1	ln(RV1)	RV2	ln(RV2)
Mean	-0.0106	0.1971	0.2356	-1.6999	0.2266	-1.7686
St.Dev.	0.4440	0.3441	0.2036	0.7104	0.2104	0.7513
Skewness	-0.0565	3.8078	4.9406	-0.0030	5.2940	-0.0052
Kurtosis	4.0417	22.6428	55.2296	3.1249	62.6838	3.0969
Min	-1.8806	0	0.0107	-4.5375	0.0080	-4.8283
5pct Qntl.	-0.7463	0.0007	0.0599	-2.8157	0.0513	-2.9701
25pct Qntl.	-0.2746	0.0149	0.1101	-2.2064	0.1011	-2.2916
50pct Qntl.	0	0.0706	0.1812	-1.7079	0.1714	-1.7638
75pct Qntl.	0.2603	0.2207	0.3017	-1.1983	0.2949	-1.2211
95pct Qntl.	0.6968	0.8349	0.5697	-0.5626	0.5551	-0.5886
Max	1.7097	3.5367	3.2218	1.1699	3.4868	1.2490
JB Stat.	96.5572	39235	279240	1.6872	241280	1.2283
(p-value)	(0.0000)	(0.0000)	(0.0000)	(0.5165)	(0.0000)	(0.6729)
ACF1	-0.0551	0.0682	0.4903	0.6370	0.4651	0.6094
ACF2	0.0068	0.1245	0.4142	0.5844	0.3979	0.5658
ACF3	0.0264	0.0744	0.3740	0.5487	0.3695	0.5247
ACF4	-0.0401	0.0865	0.3394	0.5475	0.3161	0.5255
ACF5	-0.0409	0.1154	0.3267	0.5434	0.3001	0.5213
ACF6	-0.0035	0.1207	0.3413	0.5334	0.3164	0.5080
ACF7	0.0083	0.0636	0.3246	0.5244	0.3014	0.5006
ACF8	0.0268	0.1027	0.2995	0.5063	0.2830	0.4927
ACF9	-0.0096	0.0422	0.2904	0.5044	0.2741	0.4848
ACF10	-0.0186	0.1402	0.3330	0.5073	0.3151	0.4971
ACF11	-0.0122	0.1251	0.2927	0.4847	0.2788	0.4687
ACF12	0.0038	0.0995	0.2774	0.4729	0.2594	0.4557
ACF13	0.0126	0.0886	0.2780	0.4705	0.2655	0.4594
ACF14	0.0017	0.0772	0.3028	0.4683	0.2735	0.4547
ACF15	-0.0117	0.1233	0.2928	0.4906	0.2754	0.4743

Table 3.3: Summary statistics of return and volatility time series for USD/GBP during the period of April 13th 1998 through July 28th 2006

	x_t	x_t^2	RV1	ln(RV1)	RV2	ln(RV2)
Mean	0.0043	0.2613	0.2699	-1.4581	0.2604	-1.5249
St.Dev.	0.5113	0.4155	0.1762	0.5405	0.1864	0.5969
Skewness	0.0127	3.1347	5.4828	-0.0939	5.0204	-0.1158
Kurtosis	3.5256	16.7031	82.2027	4.3860	67.4919	4.1533
Min	-1.7300	0	0.0093	-4.6777	0.0064	-5.0515
5pct Qntl.	-0.8400	0.0009	0.0983	-2.3193	0.0828	-2.4913
25pct Qntl.	-0.3100	0.0196	0.1640	-1.8079	0.1498	-1.8985
50pct Qntl.	0.0100	0.0961	0.2342	-1.4518	0.2173	-1.5267
75pct Qntl.	0.3200	0.3136	0.3301	-1.1084	0.3223	-1.1323
95pct Qntl.	0.8500	1.1449	0.5510	-0.5960	0.5630	-0.5745
Max	2.0500	4.2025	3.7471	1.3210	3.7139	1.3121
JB Stat.	26.4158	20877	445200	102.8899	148630	63.4536
(p-value)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
ACF1	-0.0116	0.0302	0.3781	0.4173	0.3477	0.3813
ACF2	-0.0191	0.0457	0.2486	0.3139	0.2401	0.2903
ACF3	0.0015	0.0101	0.2211	0.2798	0.2048	0.2695
ACF4	-0.0028	-0.0034	0.2253	0.2938	0.1998	0.2685
ACF5	0.0275	0.0466	0.2604	0.3195	0.2451	0.3010
ACF6	0.0083	0.0636	0.2291	0.2982	0.2055	0.2719
ACF7	0.0127	0.0426	0.1907	0.2622	0.1641	0.2414
ACF8	0.0270	0.0728	0.2094	0.2680	0.2102	0.2535
ACF9	-0.0010	0.0595	0.2299	0.2670	0.1930	0.2298
ACF10	-0.0067	0.1055	0.2174	0.2764	0.1934	0.2549
ACF11	-0.0372	0.0378	0.1597	0.2324	0.1367	0.2150
ACF12	-0.0101	0.0217	0.1790	0.2266	0.1723	0.2095
ACF13	0.0130	0.0307	0.1535	0.2161	0.1311	0.1944
ACF14	-0.0067	0.0307	0.1324	0.1922	0.0911	0.1694
ACF15	-0.0120	0.0488	0.1573	0.2237	0.1329	0.1992

Table 3.4: Summary statistics of return and volatility time series for DEM/USD during the period of April 13th 1998 through July 28th 2006

	x_t	x_t^2	RV1	ln(RV1)	RV2	ln(RV2)
Mean	-0.0082	0.4115	0.4051	-1.0982	0.3973	-1.1415
St.Dev.	0.6415	0.6537	0.3159	0.6080	0.3228	0.6519
Skewness	-0.0691	3.0549	5.1188	0.1330	4.8802	0.0261
Kurtosis	3.5198	15.2073	60.9138	3.7995	55.5388	3.5908
Min	-2.4300	0	0.0166	-4.0984	0.0179	-4.0230
5pct Qntl.	-1.0800	0.0009	0.1283	-2.0535	0.1106	-2.2023
25pct Qntl.	-0.3900	0.0289	0.2205	-1.5119	0.2059	-1.5806
50pct Qntl.	-0.0100	0.1521	0.3271	-1.1176	0.3193	-1.1418
75pct Qntl.	0.4000	0.4900	0.4942	-0.7047	0.4876	-0.7183
95pct Qntl.	1.0510	1.7161	0.9067	-0.0980	0.9286	-0.0741
Max	2.2000	5.9049	5.8532	1.7670	5.7451	1.7483
JB Stat.	25.0283	16510	690540	35.9701	330660	19.4825
(p-value)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)
ACF1	-0.0574	-0.0163	0.3840	0.4455	0.3222	0.3732
ACF2	-0.0030	0.0139	0.3231	0.4003	0.2951	0.3588
ACF3	-0.0093	0.0515	0.2710	0.3556	0.2445	0.3333
ACF4	0.0099	0.0210	0.2572	0.3366	0.2294	0.3036
ACF5	0.0166	0.0779	0.3032	0.3726	0.2638	0.3368
ACF6	0.0088	0.0329	0.2430	0.3207	0.2106	0.2786
ACF7	-0.0045	0.0558	0.2122	0.2988	0.1769	0.2611
ACF8	0.0086	0.0329	0.2083	0.2964	0.1678	0.2506
ACF9	-0.0281	0.0288	0.2273	0.2982	0.1980	0.2607
ACF10	0.0060	0.0525	0.1946	0.2942	0.1743	0.2682
ACF11	-0.0212	0.0070	0.1635	0.2528	0.1407	0.2219
ACF12	0.0234	0.0411	0.2105	0.2676	0.1989	0.2439
ACF13	0.0205	0.0611	0.1575	0.2627	0.1433	0.2357
ACF14	-0.0145	0.0372	0.1583	0.2614	0.1273	0.2139
ACF15	-0.0036	0.0181	0.1972	0.2925	0.1803	0.2619

Table 3.5: Estimation Results for S & P 500 Index

Estimation Results for S & P 500 Index					
Estimation Results of Model 1 Using FIML					
Parameter	λ	α	β	ρ	σ
RV1	**0.0123**	**-0.0885**	**0.7385**	**0.0089**	**0.5929**
	(0.0206)	(0.0136)**	(0.0144)**	(0.0052)	(0.0000)**
RV2	0.0127	−0.0839	0.7725	0.0239	0.5622
	(0.0208)	(0.0130)**	(0.0136)**	(0.0049)**	(0.0001)**
RV3	**0.0092**	**-0.0103**	**0.7775**	**0.0186**	**0.5581**
	(0.0177)	(0.0119)	(0.0135)**	(0.0041)**	(0.0000)**
Estimation Results of Model 2 Using FIML					
Parameter	λ	α	β	ρ	σ
RV1	0.0847	−0.0898	0.7193	−0.2709	0.5931
	(0.0205)**	(0.0131)**	(0.0139)**	(0.0063)**	(0.0003)**
RV2	0.0707	−0.0845	0.7528	−0.2252	0.5623
	(0.0208)**	(0.0127)**	(0.0133**	(0.0062)**	(0.0003)**
RV3	0.0485	−0.0061	0.7638	−0.1813	0.5582
	(0.0177)**	(0.0118)	(0.0133)**	(0.0060)**	(0.0002)**
Estimation Results of Model 2 Using 3SLS					
Parameter	λ	α	β	ρ	σ
RV1	0.0799	−0.0890	0.7197	−0.2694	0.5931
	(0.0205)**	(0.0135)**	(0.0140)**	(0.0062)**	(0.0003)**
RV2	0.0714	−0.0846	0.7527	−0.2254	0.5623
	(0.0208)**	(0.0130)**	(0.0133)**	(0.0062)**	(0.0003)**
RV3	0.0498	−0.0062	0.7637	−0.1817	0.5582
	(0.0177)**	(0.0119)	(0.0133)**	(0.0061)**	(0.0002)**

Note1: The standard errors are reported in brackets.
Note2: $**$ means 1% significance, $*$ means 5% significance.

Table 3.6: Estimation Results for CAD/USD

Estimation Results for CAD/USD					
Estimation Results of Model 1 Using FIML					
Parameter	λ	α	β	ρ	σ
RV1	−0.0452 (0.0447)	−0.6164 (0.0308)**	0.6373 (0.0167)**	−0.0310 (0.0004)**	0.5477 (0.0000)**
RV2	−0.0470 (0.0456)	−0.6894 (0.0331)**	0.6103 (0.0172)**	−0.0368 (0.0008)**	0.5958 (0.0000)**
Estimation Results of Model 2 Using FIML					
Parameter	λ	α	β	ρ	σ
RV1	−0.0691 (0.0447)*	−0.6156 (0.0308)**	0.6378 (0.0167)**	0.0444 (0.0057)**	0.5477 (0.0000)**
RV2	−0.0845 (0.0456)*	−0.6891 (0.0330)**	0.6103 (0.0172)**	0.0635 (0.0061)**	0.5958 (0.0000)**
Estimation Results of Model 2 Using 3SLS					
Parameter	λ	α	β	ρ	σ
RV1	−0.0757 (0.0447)*	−0.6156 (0.0308)**	0.6378 (0.0167)**	0.0453 (0.0057)**	0.5477 (0.0000)**
RV2	−0.0896 (0.0456)*	−0.6892 (0.0330)**	0.6103 (0.0172)**	0.0642 (0.0061)**	0.5958 (0.0000)**

Note1: The standard errors are reported in brackets.
Note2: ** means 1% significance, * means 5% significance.

Table 3.7: Estimation Results for GBP/USD

Estimation Results for USD/GBP					
Estimation Results of Model 1 Using FIML					
Parameter	λ	α	β	ρ	σ
RV1	0.0186 (0.0415)	−0.8788 (0.0309)**	0.4014 (0.0197)**	−0.0363 (0.0003)**	0.5056 (0.0000)**
RV2	0.0194 (0.0423)	−0.9666 (0.0330)**	0.3704 (0.0220)**	−0.0462 (0.0003)**	0.5642 (0.0002)**
Estimation Results of Model 2 Using FIML					
Parameter	λ	α	β	ρ	σ
RV1	0.0223 (0.0415)	−0.8784 (0.0309)**	0.4017 (0.0197)**	−0.0067 (0.0052)	0.5056 (0.0000)**
RV2	0.0250 (0.0423)	−0.9660 (0.0330)**	0.3708 (0.0200)**	−0.0097 (0.0057)	0.5642 (0.0000)**
Estimation Results of Model 2 Using 3SLS					
Parameter	λ	α	β	ρ	σ
RV1	0.0206 (0.0415)	−0.8784 (0.0309)**	0.4017 (0.0197)**	−0.0065 (0.0052)	0.5056 (0.0000)**
RV2	0.0232 (0.0423)	−0.9660 (0.0330)**	0.3708 (0.0200)**	−0.0094 (0.0057)	0.5642 (0.0000)**

Note1: The standard errors are reported in brackets.
Note2: ** means 1% significance, * means 5% significance.

Table 3.8: Estimation Results for DEM/USD

Estimation Results for DEM/USD					
Estimation Results of Model 1 Using FIML					
Parameter	λ	α	β	ρ	σ
RV1	−0.0197 (0.0341)	−0.6089 (0.0244)**	0.4452 (0.0194)**	0.0145 (0.0003)**	0.5445 (0.0000)**
RV2	−0.0201 (0.0344)	−0.7150 (0.0264)**	0.3733 (0.0201)**	0.0197 (0.0005)**	0.6049 (0.0000)**
Estimation Results of Model 2 Using FIML					
Parameter	λ	α	β	ρ	σ
RV1	−0.0077 (0.0341)	−0.6083 (0.0244)**	0.4457 (0.0194)**	−0.0298 (0.0058)**	0.5445 (0.0000)**
RV2	−0.0047 (0.0344)	−0.7147 (0.0264)**	0.3735 (0.0201)**	−0.0346 (0.0063)**	0.6049 (0.0000)**
Estimation Results of Model 2 Using 3SLS					
Parameter	λ	α	β	ρ	σ
RV1	−0.0049 (0.0341)	−0.6084 (0.0244)**	0.4457 (0.0194)**	−0.0303 (0.0058)**	0.5445 (0.0000)**
RV2	−0.0019 (0.0344)	−0.7147 (0.0264)**	0.3735 (0.0201)**	−0.0351 (0.0063)**	0.6049 (0.0000)**

Note1: The standard errors are reported in brackets.
Note2: $**$ means 1% significance, $*$ means 5% significance.

Table 3.9: ACF of Forecast Error for SPX

		M 1	M 2	HAR	SRM
$RV1$	ACF1	-0.4648	-0.4628	-0.4710	-0.4648
	ACF2	-0.0214	-0.0200	-0.0296	-0.0214
	ACF3	0.0640	0.0651	0.0650	0.0641
	ACF4	-0.0596	-0.0583	-0.0661	-0.0595
	ACF5	0.0106	0.0113	0.0153	0.0106
	ACF6	0.0560	0.0572	0.0451	0.0560
	ACF7	-0.0088	-0.0078	-0.0105	-0.0088
	ACF8	-0.0901	-0.0886	-0.0932	-0.0900
	ACF9	-0.0434	-0.0417	-0.0552	-0.0433
	ACF10	0.1139	0.1148	0.1202	0.1140
$RV2$	ACF1	-0.4461	-0.4440	-0.4514	-0.4458
	ACF2	-0.0390	-0.0371	-0.0511	-0.0387
	ACF3	0.0479	0.0485	0.0544	0.0480
	ACF4	-0.0598	-0.0582	-0.0693	-0.0596
	ACF5	0.1145	0.1150	0.1175	0.1145
	ACF6	-0.0538	-0.0525	-0.0592	-0.0536
	ACF7	0.0014	0.0028	-0.0066	0.0016
	ACF8	-0.0949	-0.0935	-0.0943	-0.0947
	ACF9	-0.0277	-0.0258	-0.0409	-0.0275
	ACF10	0.1523	0.1529	0.1619	0.1524
$RV3$	ACF1	-0.4590	-0.4577	-0.4652	-0.4588
	ACF2	-0.0279	-0.0268	-0.0380	-0.0278
	ACF3	0.0609	0.0610	0.0714	0.0609
	ACF4	-0.0846	-0.0836	-0.0955	-0.0845
	ACF5	0.1346	0.1348	0.1370	0.1346
	ACF6	-0.0641	-0.0634	-0.0688	-0.0640
	ACF7	-0.0048	-0.0040	-0.0120	-0.0047
	ACF8	-0.0764	-0.0758	-0.0723	-0.0762
	ACF9	-0.0639	-0.0628	-0.0761	-0.0638
	ACF10	0.1905	0.1906	0.2011	0.1905

Table 3.10: ACF of Forecast Error for CAD/USD

		M 1	M 2	HAR	SRM
$RV1$	ACF1	-0.2812	-0.2802	-0.3995	-0.2803
	ACF2	-0.0804	-0.0800	-0.1151	-0.0801
	ACF3	0.0041	0.0043	-0.0181	0.0043
	ACF4	0.0915	0.0917	0.0801	0.0916
	ACF5	-0.0164	-0.0161	-0.0482	-0.0161
	ACF6	0.0214	0.0216	-0.0168	0.0217
	ACF7	0.0913	0.0915	0.0698	0.0915
	ACF8	0.0261	0.0264	-0.0020	0.0264
	ACF9	-0.0230	-0.0228	-0.0429	-0.0228
	ACF10	0.0811	0.0811	0.0710	0.0811
$RV2$	ACF1	-0.3093	-0.3072	-0.4504	-0.3075
	ACF2	-0.0070	-0.0063	-0.0377	-0.0065
	ACF3	0.0173	0.0177	-0.0065	0.0177
	ACF4	0.0441	0.0445	0.0236	0.0444
	ACF5	0.0255	0.0261	-0.0025	0.0260
	ACF6	-0.0054	-0.0048	-0.0394	-0.0049
	ACF7	0.0660	0.0664	0.0380	0.0663
	ACF8	0.0532	0.0537	0.0194	0.0537
	ACF9	-0.0215	-0.0209	-0.0571	-0.0210
	ACF10	0.1339	0.1341	0.1143	0.1341

Table 3.11: ACF of Forecast Error for USD/GBP

		M 1	M 2	HAR	SRM
$RV1$	ACF1	-0.3160	-0.3164	-0.5088	-0.3165
	ACF2	0.0047	0.0046	-0.0340	0.0046
	ACF3	0.0635	0.0634	0.0310	0.0633
	ACF4	0.0736	0.0735	0.0195	0.0734
	ACF5	0.0299	0.0298	0.0070	0.0298
	ACF6	-0.0028	-0.0029	-0.0524	-0.0029
	ACF7	-0.0174	-0.0175	-0.0507	-0.0175
	ACF8	0.1147	0.1147	0.0897	0.1147
	ACF9	-0.0433	-0.0434	-0.0954	-0.0434
	ACF10	0.0966	0.0965	0.0556	0.0965
$RV2$	ACF1	-0.2892	-0.2898	-0.5048	-0.2899
	ACF2	0.0117	0.0116	-0.0133	0.0116
	ACF3	0.0307	0.0306	0.0185	0.0306
	ACF4	0.0115	0.0114	-0.0334	0.0114
	ACF5	0.0841	0.0841	0.0710	0.0840
	ACF6	-0.0149	-0.0151	-0.0529	-0.0151
	ACF7	-0.0567	-0.0568	-0.0920	-0.0568
	ACF8	0.1639	0.1640	0.1521	0.1639
	ACF9	-0.1015	-0.1017	-0.1667	-0.1017
	ACF10	0.1469	0.1468	0.1232	0.1468

Table 3.12: ACF of Forecast Error for DEM/USD

		M 1	M 2	HAR	SRM
$RV1$	ACF1	-0.3440	-0.3444	-0.4709	-0.3440
	ACF2	-0.0184	-0.0186	-0.0434	-0.0184
	ACF3	0.0408	0.0408	0.0260	0.0409
	ACF4	0.0012	0.0011	-0.0326	0.0012
	ACF5	0.0836	0.0836	0.0955	0.0836
	ACF6	-0.1203	-0.1204	-0.1583	-0.1203
	ACF7	0.0867	0.0866	0.0668	0.0867
	ACF8	0.1292	0.1293	0.1199	0.1292
	ACF9	-0.1426	-0.1427	-0.1759	-0.1426
	ACF10	0.0962	0.0962	0.0844	0.0962
$RV2$	ACF1	-0.3043	-0.3047	-0.4857	-0.3041
	ACF2	-0.0037	-0.0037	-0.0233	-0.0037
	ACF3	0.0546	0.0546	0.0356	0.0546
	ACF4	-0.0365	-0.0365	-0.0659	-0.0365
	ACF5	0.1216	0.1216	0.1379	0.1215
	ACF6	-0.1309	-0.1310	-0.1551	-0.1309
	ACF7	0.0820	0.0820	0.0786	0.0820
	ACF8	0.0734	0.0734	0.0593	0.0734
	ACF9	-0.1068	-0.1069	-0.1391	-0.1067
	ACF10	0.1272	0.1272	0.1170	0.1272

Table 3.13: DM95 Testing Results

		$HAR/M1$	$HAR/M2$	HAR/SRM	$M1/M2$	$M1/SRM$	$M2/SRM$
S & P 500 Index							
$RV1$	$S1$	-12.7118	-13.6952	-12.7758	-29.7868	-22.3646	29.3468
	$S2$	-10.2050	-10.9610	-10.2050	-14.3626	-14.7406	14.2367
	$S3$	-13.7640	-13.7636	-13.7640	-13.7614	-13.7613	-13.7601
$RV2$	$S1$	-11.2425	-12.2712	-11.4040	-27.2796	-30.5505	26.4509
	$S2$	-10.2050	-10.8350	-10.2050	-13.8587	-14.6146	13.7327
	$S3$	-13.7635	-13.7632	-13.7635	-13.7614	-13.7613	-13.7602
$RV3$	$S1$	-11.0204	-11.7444	-11.1318	-26.6787	-30.0223	25.8595
	$S2$	-9.8271	-10.7090	-9.8271	-13.9847	-14.6146	13.8587
	$S3$	-13.7643	-13.7641	-13.7643	-13.7614	-13.7613	-13.7602
CAD/USD Rate							
$RV1$	$S1$	7.7303	7.7806	7.7816	4.8679	6.8880	2.6699
	$S2$	6.5993	6.5993	6.5993	-4.2335	-1.1206	-8.3425
	$S3$	-13.9213	-13.9213	-13.9213	-13.9239	-13.9239	-13.9239
$RV2$	$S1$	7.9647	8.0597	8.0481	6.7675	7.3519	-2.3053
	$S2$	6.3502	6.3502	6.3502	0.8716	1.4942	-9.5876
	$S3$	-13.9212	-13.9212	-13.9212	-13.9238	-13.9238	-13.9239
USD/GBP Rate							
$RV1$	$S1$	-3.5391	-3.5419	-3.5427	-2.7468	-2.9877	-2.1239
	$S2$	-2.4903	-2.4903	-2.4903	-7.9689	-6.5993	-14.5682
	$S3$	-13.9244	-13.9244	-13.9244	-13.9239	-13.9239	-13.9239
$RV2$	$S1$	-3.3986	-3.4085	-3.4104	-4.0870	-4.2209	-2.5248
	$S2$	-2.1206	-2.1206	-2.1206	-5.3541	-4.3580	-14.4437
	$S3$	-13.9244	-13.9244	-13.9244	-13.9239	-13.9239	-13.9239
DEM/USD Rate							
$RV1$	$S1$	-15.0553	-15.0342	-15.0589	3.3469	-5.1509	-4.9627
	$S2$	-11.4553	-11.4553	-11.4553	-3.3619	-15.9379	-10.3347
	$S3$	-13.9251	-13.9251	-13.9251	-13.9239	-13.9239	-13.9239
$RV2$	$S1$	-14.8047	-14.7909	-14.8198	2.2563	-9.4026	-5.0296
	$S2$	-11.7044	-11.7044	-11.7044	-4.2335	-14.4437	-9.0896
	$S3$	-13.9252	-13.9252	-13.9252	-13.9239	-13.9239	-13.9239

Table 3.14: Forecasting Performance of Four Models for SPX

			$M1$	$M2$	HAR	SRM
$RV1$	RMSE	Value	0.2851	0.2871	0.2130	0.2852
		Rank	2	4	1	3
	MAE	Value	0.2151	0.2180	0.1575	0.2153
		Rank	2	4	1	3
	$Theil-U$	Value	0.8875	0.8651	0.4392	0.8863
		Rank	4	2	1	3
	$QLIKE$	Value	0.0428	0.0456	-0.0017	0.0430
		Rank	2	4	1	3
$RV2$	RMSE	Value	0.2842	0.2856	0.2213	0.2844
		Rank	2	4	1	3
	MAE	Value	0.2069	0.2102	0.1557	0.2074
		Rank	2	4	1	3
	$Theil-U$	Value	0.9024	0.8795	0.4924	0.8992
		Rank	4	2	1	3
	$QLIKE$	Value	-0.0042	-0.0016	-0.0467	-0.0038
		Rank	2	4	1	3
$RV3$	RMSE	Value	0.4050	0.4061	0.3149	0.4052
		Rank	2	4	1	3
	MAE	Value	0.2898	0.2928	0.2162	0.2903
		Rank	2	4	1	3
	$Theil-U$	Value	0.9058	0.8895	0.4695	0.9036
		Rank	4	2	1	3
	$QLIKE$	Value	0.3113	0.3130	0.2690	0.3116
		Rank	2	4	1	3

Table 3.15: Forecasting Performance of Four Models for CAD/USD

			$M1$	$M2$	HAR	SRM
$RV1$	RMSE	Value	0.1207	0.1207	0.1126	0.1207
		Rank	2	2	1	2
	MAE	Value	0.0898	0.0897	0.0861	0.0897
		Rank	3	2	1	2
	$Theil-U$	Value	0.3267	0.3248	0.2064	0.3249
		Rank	4	2	1	3
	$QLIKE$	Value	-0.2674	-0.2674	-0.2875	-0.2674
		Rank	3	2	1	3
$RV2$	RMSE	Value	0.1187	0.1186	0.1108	0.1186
		Rank	3	2	1	2
	MAE	Value	0.0917	0.0916	0.0856	0.0916
		Rank	3	2	1	2
	$Theil-U$	Value	0.3034	0.3002	0.2218	0.3006
		Rank	4	2	1	3
	$QLIKE$	Value	-0.3257	-0.3258	-0.3460	-0.3258
		Rank	3	2	1	2

Table 3.16: Forecasting Performance of Four Models for USD/GBP

			$M1$	$M2$	HAR	SRM
$RV1$	RMSE	Value	0.1216	0.1216	0.1149	0.1216
		Rank	2	2	1	2
	MAE	Value	0.0917	0.0917	0.0860	0.0917
		Rank	2	2	1	2
	$Theil-U$	Value	0.1952	0.1957	0.1886	0.1958
		Rank	2	3	1	4
	$QLIKE$	Value	-0.3739	-0.3738	-0.3865	-0.3738
		Rank	2	3	1	3
$RV2$	RMSE	Value	0.1416	0.1416	0.1351	0.1416
		Rank	2	2	1	2
	MAE	Value	0.1051	0.1051	0.0992	0.1051
		Rank	2	2	1	2
	$Theil-U$	Value	0.1529	0.1534	0.1521	0.1535
		Rank	2	3	1	4
	$QLIKE$	Value	-0.3797	-0.3796	-0.3947	-0.3796
		Rank	2	3	1	3

Table 3.17: Forecasting Performance of Four Models for DEM/USD

			$M1$	$M2$	HAR	SRM
$RV1$	RMSE	Value	0.1589	0.1589	0.1358	0.1589
		Rank	2	2	1	2
	MAE	Value	0.1316	0.1316	0.1066	0.1316
		Rank	2	2	1	2
	$Theil-U$	Value	0.2797	0.2803	0.1888	0.2797
		Rank	2	3	1	2
	$QLIKE$	Value	-0.2476	-0.2476	-0.2737	-0.2476
		Rank	2	2	1	2
$RV2$	RMSE	Value	0.1770	0.1770	0.1541	0.1770
		Rank	2	2	1	2
	MAE	Value	0.1481	0.1481	0.1203	0.1482
		Rank	2	2	1	3
	$Theil-U$	Value	0.1861	0.1866	0.1573	0.1859
		Rank	3	4	1	2
	$QLIKE$	Value	-0.2676	-0.2676	-0.2977	-0.2676
		Rank	2	2	1	2

3.7 Bibliography

AKGIRAY, V. (1989): "Conditional Heteroscedasticity in time series of stock returns: Evidence and Forecast," *Journal of Business*, 62, 55–80.

ANDERSEN, T., AND G. T. BOLLERSLEV (1998): "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885–905.

ANDERSEN, T. G., T. BOLLERSLEV, F. DIEBOLD, AND P. LABYS (2001): "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.

——— (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71(2), 579–625.

ANDERSEN, T. G., T. BOLLERSLEV, AND F. X. DIEBOLD (2007): "Roughing It Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility," *Review of Economics and Statistics*, 89, 701–20.

ANDERSEN, T. G., T. BOLLERSLEV, AND N. MEDDAHI (2005): "Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities," *Econometrica*, 73(1), 279–96.

ANDERSON, T. G., T. BOLLERSLEV, P. F. CHRISTOFFERSEN, AND F. X. DIEBOLD (2005): "Volatility and Correlation Forecasting," *Handbook of Economic Forecasting*, G. Elliott, C.W.J.Granger and A.Timmermann ed.s, North Holland Press, Amsterdam.

ARMSTRONG, J. S., AND R. FILDERS (1995): "On the Selection of Error Measures for Comparisons among Forecasting Methods," *Journal of Forecasting*, 14, 67–71.

BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2002): "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society*, B64, 253–280.

BOLLERSLEV, T., R. F. ENGLE, AND D. NELSON (1994): "ARCH Models," *Handbook of Econometrics*, IV(Engle, R.F., McFadden, D. L. (eds) Elsevier Science B.V. Amsterdam), 2961–3038.

BOLLERSLEV, T., M. GIBSON, AND H. ZHOU (2010): "Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities," *Journal of Econometrics*, forthcoming, doi: 10.1016/j.jeconom.2010.03.033.

BOLLERSLEV, T., AND H. ZHOU (2006): "Volatility Puzzles: A Simple Framework for Gauging Return-Volatility Regressions," *Journal of Econometrics*, 131(1-2), 123–50.

BRAILSFORD, T. J., AND R. W. FAFF (1996): "An Evaluation of Volatility Forecasting Technique," *Journal of Banking and Finance*, 20, 419–38.

CHRISTOFFERSEN, P. F., AND F. DIEBOLD (1987): "Optimal Prediction under Asymmetric Loss," *Econometric Theory*, 13, 808–817.

CORNISH, R. (2007): "A Comparison of the Properties of Realized Variance for the FTSE 100 and FTSE 250 Equity Indices," *Forecasting Volatility in the Financial Markets*, 3ed, Satchell, S. and Knight. J., Elsevier Butterworth-Heinemann.

CORSI, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, pp. 1–23.

CORSI, F., S. MITTNIK, C. PIGORSCH, AND U. PIGORSCH (2008): "The Volatility of Realized Volatility," *Econometric Reviews*, 27(1-3), 46–78.

DE BRUIN, P., AND P. H. FRANSES (1999): "Forecasting Power-Transformed Time Series Data," *Journal of Applied Statistics*, 26(7), 807–15.

DIEBOLD, F. X., AND J. A. LOPEZ (1996): "Forecast Evaluation and Combination," Working paper.

DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–63.

GRANGER, C. W. J. (1996): "Can We Improve the Perceived Quality of Economic Forecasts?," *Journal of Applied Econometrics*, 11(5), 455–473.

GRANGER, C. W. J., AND P. NEWBOLD (1976): "Forecasting Transformed Series," *Journal of the Royal Statistical Society*, 38(2), 189–203.

GRANGER, C. W. J., AND M. H. PESARAN (1999): "A Decision Theoretic Approach to Forecast Evaluation," Working paper.

GRANGER, C. W. J., AND S. H. POON (2003): "Forecasting Volatility in Financial Markets: A Review," *Journal of Economic Literature*, XLI, 478–539.

HANSEN, P. R., AND A. LUNDE (2005): "A Forecast Comparison of Volatility Models: Does Anything Beat A GARCH(1,1)?," *Journal of Applied Econometrics*, 20(7), 873–89.

HOL, E., AND S. J. KOOPMAN (2002): "Stock Index Volatility Forecasting with High Frequency Data," Free U. Amsterdam.

HUWANG, P., J. L. KNIGHT, AND S. E. SATCHELL (1999): "Forecasting Volatility Using LINEX Loss Functions," Working paper.

JACQUIER, E., N. G. POLSON, AND P.ROSSI (1994): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371–417.

LIU, C., AND J. M. MAHEU (2008): "Forecasting Realized Volatility: A Bayesian Model Averaging Approach," Working paper.

LOPEZ, J. A. (2001): "Evaluating the Predictive Accuracy of Volatility Models," *Journal of Forecasting*, 20(2), 87–109.

MARTENS, M. (2002): "Measuring and Forecasting S & P 500 index Futures Volatility Using High-Frequency Data," *Journal of Futures Markets*, 22, 497–518.

MCALEER, M., AND M. C. MEDEIROS (2008): "Realized Volatility: A Review," *Econometric Reviews*, 27(1-3), 10–45.

MERTON, R. C. (1980): "On Estimation the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–61.

PATTON, A. (2010): "Volatility Forecast Comparison Using Imperfect Volatility Proxies," *Journal of Econometrics*, forthcoming.

PATTON, A., AND K. SHEPPARD (2007): "Evaluating Volatility Forecasts," *Handbook of Financial Time Series*, Andersen, T.G., R.A.Davis, J.P. Kreiss, and T. Mikosch, eds. Springer-Verlag.

YU, J. (2002): "Forecasting Volatility in the New Zealand Stock Market," *Applied Financial Economics*, 12, 193–202.

Chapter 4

Estimation of Continuous-Time Stochastic Volatility Model When Volatility is observed: A Monte Carlo Study

4.1 Introduction

The continuous-time SV model was first introduced by Hull and White (1987), Johnson and Shanno (1987), and Scott (1987), and Wiggins (1987), motivated by the desire to accurately model option prices where the underlying asset price volatility was believed to be stochastic. Viewed as the limit of the discrete-time SV model, the continuous-time SV model deals with a system of stochastic differential equations (SDE). In general, the representation can be written as:

$$dS_t/S_t = \mu_t dt + \sigma_t(V_t)dB_{1t}$$

$$dV_t = \gamma_t dt + \delta_t dB_{2t}$$

$$dB_{1t}dB_{2t} = \rho_t dt$$

where S_t refers to the financial asset price process while V_t is the volatility state variable. As in the discrete-time SV model, V_t is an unobserved latent variable, and is stochastic following a diffusion process. B_{1t}, B_{2t} are two standard Brownian motion processes which are possibly correlated with $cov(dB_{1t}, dB_{2t}) = \rho_t dt$[1] and $\mu_t, \sigma_t, \gamma_t, \delta_t$, and ρ_t are coefficient functions.

The continuous time SV model has been widely used in finance for modeling asset prices, including stock prices, interest rates, and exchange rates. For example, the square-root diffusion process has been used by Cox, Ingersoll, and Ross (1985) to model the nominal interest rates, and by Bailey and Stulz (1989) and Heston (1993) to model the conditional volatility of asset returns. The affine jump-diffusion process has often been used to model the dynamics of asset returns, index returns, exchange rates, etc. As the continuous-time SV model can explain some empirical features of the joint time-series behavior of stock and option prices, such as time varying volatility, fat tails of asset return distribution, etc., it has dominated the option pricing literature since the mid-1980s.

However, traditional inference for the continuous-time SV model has been viewed as difficult for some time. One challenge is that the continuous sample of observations is unavailable and thus often requiring the model to be discretized and introducing a significant discretization error. The estimation is even more difficult as one of the state variables, the volatility process, can not be directly observed. Thus the estimation of the SV model with only the time series of stock prices being observed requires the elimination of the unobserved variables. In addition, except for a few cases, the transition density does not have a closed form expression hence maximum likelihood method is not

[1]A standard Brownian motion $B_t : (t \geq 0)$ is a stochastic process having (1)continuous paths, (2) stationary, independent increments, (3)$B_t \sim N(0, t)$, for all $t \geq 0$.

directly available. Numerous competing estimation strategies have been proposed in the literature. Most of these approaches are moment-based and based on simulation. For example, Chan, Karolyi, Longstaff, and Sanders (1992), Hansen and Scheinkman (1996) proposed the generalized method of moment (GMM) approach; Duffie and Singleton (1993) advanced the simulated methods of moments (SMM) estimation; Gallant and Tauchen (1996) and Gallant and Long (1997) developed the efficient methods of moments (EMM); Jacquier, Polson, and Rossi (1994), Eraker (1998), Kim, Shephard, and Chib (1998) developed the Markov Chain Monte Carlo (MCMC) estimation methods, etc. All of these estimation methods yield consistent estimates, however, in practice, the estimation is computationally demanding or involves discretizaion error or is based on simulation methods.

It has been noticed that for the affine diffusion and affine jump diffusion processes, although the transition density functions are unknown, the corresponding conditional characteristic functions (CCF) can be derived explicitly. Based on this finding, some new strategies have been developed. For example, Singleton (2001) proposed to use the CCF along with simulation. Chacko and Viceira (2003) constructed a GMM estimator based on the unconditional mean of the difference between the CCF and its empirical counter part. Jiang and Knight (2002) exploited the explicit functional form of the unconditional joint characteristic function and developed an estimation procedure based on the empirical joint characteristic function. Because of the one-to-one correspondence between the distribution function and the characteristic function, the empirical characteristic function (ECF) contains the same amount of information as the empirical distribution function. Hence with a judiciously chosen weight function, the ECF approach not only provides consistent estimators, but also can achieve high efficiency close to that of maximum likelihood estimation.

Despite the theoretical appeal of the ECF method, as volatility can not be observed directly, it has to be integrated out of the joint characteristic function. Moreover, the lack of general solution for the choice of optimal weight function increases the computational burden.

In their recent study, Jiang and Knight (2010) advanced the ECF approach by proposing an analytical approximation of the optimal weight function via an Edgeworth/Gram-Charlier expansion of the logarithmic transition density. Their procedure is similar to the approximate maximum likelihood estimation (AMLE) method introduced by Ait-Sahalia (2002), however, unlike the AMLE, their method ensures the consistency of the estimation hence is named consistent AMLE (hereafter C-AMLE). Jiang and Knight (2010) proposed to apply the C-AMLE approach widely to the Markov models where the transition density is unknown, including both the discrete-time and the continuous-time SV models. They illustrated the application of the C-AMLE method via the Monte Carlo simulations. In their Monte Carlo study, they investigated the C-AMLE estimation for the univariate diffusion process, however, estimation of the bivariate diffusion processes was unexamined. For the bivariate diffusion processes, when both the asset return series and the volatility series are observed, the implementation of the C-AMLE is fairly straightforward and computationally easy since the volatility state variable does not have to be integrated out of the model. Motivated by the advantage of the C-AMLE procedure and easy implementation by treating volatility as an observable variable, in this chapter, we extend Jiang and Knight (2010)'s study by examining the C-AMLE estimation performance for the affine continuous-time SV model via a Monte Carlo experiment in which both asset return and volatility state variables are observed. In our Monte Carlo study, we simulate the volatility process from its unconditional density function and employ the almost exact simulation method to generate the asset return process, at both daily and monthly frequencies. The moments calculated from our simulations are very close to the true moments, indicating that our simulations are accurate. We then apply the C-AMLE method for the affine continuous-time SV model using simulated observations at both daily and monthly frequencies, the results suggest that the C-AMLE method does a good job at recovering the true parameters.

The rest of this paper is organized as follows. In Section 4.2 we discuss the affine continuous-time SV model specification and its dynamic statistical properties. Section 4.3 discusses the estimation of the affine continuous-time SV model when volatility is latent as well as estimation when volatility is observed. In Section 4.4 we discuss the procedure of data generation and present the Monte Carlo experiment results. A brief conclusion is contained in the last section.

4.2 Affine Continuous-Time SV Model and Its Statistical Properties

4.2.1 Model Specification

Let x_t be the time-t logarithmic price of the risky asset or portfolio, i.e. $x_t = log(p_t)$, and V_t be the asset return volatility process, the affine continuous-time SV model consists of a system of SDEs:

$$dx_t = \mu dt + \sqrt{V_t} dB_{1t} \tag{4.2a}$$

$$dV_t = \beta(\alpha - V_t)dt + \sigma\sqrt{V_t} dB_{2t} \tag{4.2b}$$

$$dB_{1t}dB_{2t} = \rho dt \tag{4.2c}$$

where the first SDE describes the dynamics of the risky financial asset, and the second SDE describes dynamics of the asset variance. It is often the case that while the volatility process is stationary, the whole affine diffusion process is non-stationary as the logarithmic asset price process is first difference stationary. After transformation, the vector of (s_t, V_t) is a stationary process, where $s_t = \Delta x_t = x_t - x_{t-\Delta}$ refers to the asset return process.

In both SDEs, the drift functions as well as the diffusion functions have an "affine" structure[2]. That is, $\mu(x_t) = \mu, \mu(V_t) = \beta(\alpha - V_t)$, and $\sigma^2(x_t) = V_t, \sigma^2(V_t) = \sigma^2 V_t$.

In the first SDE, as in Singleton (2001), and Jiang and Knight (2002), we specify the drift term of the asset return process as a constant μ, which is the rate of return of the financial asset or portfolio.

In the second SDE, the volatility process follows a square-root diffusion process, and is often called a CIR process since it was first used by Cox, Ingersoll, and Ross (1985) to model interest rates. The specification of the instantaneous volatility process guarantees the volatility being non-negative. Specifically, the parameter $\alpha > 0$ determines the long-run unconditional mean of the volatility process. As time t tends to infinity, the expected value of V_t tends to α. The parameter β refers to the degree of mean reversion for the volatility process, i.e. how long it takes to converge to the long-run mean. β can be interpreted as presenting the degree of "volatility clustering", and is often assumed to be positive so the process displays the property of mean reversion and is stationary. σ is the volatility of volatility influencing the kurtosis of the distribution.

[2]Intuitively, the drift terms refers to the time trend of the processes, and the diffusion terms represent the variance of the processes.

If $\sigma = 0$, volatility is deterministic and stock prices are lognormal, $\sigma > 0$ results in fat-tailed distributions which is one of the stylized facts of asset return series. dB_{1t} and dB_{2t} are two different standard Brownian Motion processes with the instantaneous $corr(dB_{1t}, dB_{2t}) = \rho$ which measures the level of asymmetry of the conditional volatility. If $\rho > 0$, then asset return and volatility are positively correlated, this will create a fat right-tailed distribution. If $\rho < 0$, then there exists a "leverage effect", i.e. volatility will increase as return decreases, resulting in a fat left-tailed distribution.

The affine continuous-time SV model was proposed by Heston (1993) (hence is often called the Heston model). It has been widely used in the empirical finance literature. The aforementioned features of this model enable it to support various shapes of the density functions, hence can capture many stylized facts of financial time series. In addition, a computationally convenient feature of this model is that it provides a closed-form solution for the European option prices and the CCF of the asset return, making it more tractable and easier to implement than other continuous-time models. There are numerous studies examining estimation of the affine continuous-time SV model, in general, volatility is treated as latent. In this chapter, we extend the study in the literature by examining estimation of this model treating volatility as an observed variable.

4.2.2 Statistical Properties of the Affine Diffusion Processes

The dynamic properties of the diffusion process are determined by its transition density function. The square-root process, known as the CIR model, is one of the two well-known diffusion processes (the other is Ornstein-Uhlenbeck process) having an explicit transition density function. As Cox, Ingersoll, and

Ross (1985) discussed, among others, the distribution of V_t conditional on initial value V_0 is a noncentral chi-squared distribution:

$$f(V_t|V_0 = v_0) = \frac{\sigma^2(1 - e^{-\beta t})}{4\beta}\chi_d'^2(\lambda) \tag{4.3}$$

where $\chi_d'^2(\lambda)$ denotes a noncentral chi-squared random variable. The parameter d refers to the degrees of freedom and λ represents the noncentrality, with

$$d = \frac{4\alpha\beta}{\sigma^2} \tag{4.4}$$

$$\lambda = \frac{4\beta e^{-\beta t}}{\sigma^2(1 - e^{-\beta t})}V_0 \tag{4.5}$$

The conditional first and second moments are given by:

$$E[V_t|V_0 = v_0] = v_0 e^{-\beta t} + \alpha(1 - e^{-\beta t}) \tag{4.6}$$

$$Var[V_t|V_0 = v_0] = \frac{v_0\sigma^2}{\beta}(e^{-\beta t} - e^{-2\beta t}) + \frac{\alpha\sigma^2}{2\beta}(1 - e^{-\beta t})^2 \tag{4.7}$$

As Jiang and Knight (2002) showed, with the mean reversion parameter $\beta > 0$, the process is stationary, its marginal distribution is in fact a gamma distribution:

$$f(V_t = v_t) = (\frac{\omega^s}{\Gamma(s)})v_t^{s-1}e^{-\omega v_t} \tag{4.8}$$

where $\omega = \frac{2\beta}{\sigma^2}$, $s = \frac{2\alpha\beta}{\sigma^2}$.

The explicit expression of the first four unconditional moments are readily derived as:

$$E[V_t] = \alpha \tag{4.9a}$$

$$Var[V_t] = \frac{\alpha\sigma^2}{2\beta} \tag{4.9b}$$

$$E[(V_t - \alpha)^3] = \frac{\alpha(2\alpha^2\beta^2 + 3\alpha\beta\sigma^2 + \sigma^4)}{2\beta^2} \tag{4.9c}$$

$$E[(V_t - \alpha)^4] = \frac{\alpha(4\alpha^3\beta^3 + 12\alpha^2\beta^2\sigma^2 + 11\alpha\beta\sigma^4 + 3\sigma^6)}{4\beta^3} \tag{4.9d}$$

Clearly the third moment of the volatility process is nonzero indicating that the volatility distribution is asymmetric. In fact, the long-run mean parameter α and mean reversion parameter β both being positive leads to the positive skewness of volatility process. The fourth moment is also positive, demonstrating that the volatility distribution has fatter tails comparing with the normal distribution. These theoretical features are consistent with the typical empirical findings about the asset return volatility process.

It is noted that although the volatility process has an explicit transition density function, the explicit form of transition density function for the asset return process is unavailable. Alternatively, the unconditional characteristic function of the return process can be derived. Consequently the dynamic statistical properties of asset return process can be analyzed. As shown in Jiang and Knight (2002), the unconditional characteristic function of the asset returns is:

$$\phi(r; s_t) = e^C(1 - \frac{\sigma^2 D}{2\beta})^{\frac{-2\alpha\beta}{\sigma^2}} \tag{4.10}$$

where

$C = ir\mu + \frac{\alpha\beta}{\sigma^2}(b - h - 2\ln(\frac{1-ge^{-h}}{1-g}))$

$D = \frac{b-h}{\sigma^2}\frac{1-e^{-h}}{1-ge^{-h}}$ and

$b = \beta - i\rho\sigma r, h = \sqrt{b^2 + \sigma^2 r^2}, g = \frac{b-h}{b+h}.$

Based on the unconditional characteristic function, the first four unconditional moments of asset return process can be derived:

$$E[s_t] = \mu \tag{4.11a}$$

$$Var[s_t] = \alpha \tag{4.11b}$$

$$E[(s_t - \mu)^3] = \frac{3\alpha\rho\sigma(e^{-\beta} + \beta - 1)}{\beta^2} \tag{4.11c}$$

$$E[(s_t - \mu)^4] = 3\alpha^2 + \frac{3\alpha\sigma^2(e^{-\beta} + \beta - 1 + 4((2+\beta)e^{-\beta} + \beta - 2)\rho^2)}{\beta^3} \tag{4.11d}$$

The sign of the third moment is determined by the sign of the correlation coefficient parameter ρ as $e^{-\beta} + \beta - 1$ is always positive[3]. The asset return distribution is asymmetric when $\rho \neq 0$. In particular, if $\rho < 0$ then the asset return distribution is negatively skewed. The fourth moment has positive sign. Moreover, $E[(s_t-\mu)^4]-3Var(s_t)^2 = E[(s_t-\mu)^4]-3\alpha^2 = \frac{3\alpha\sigma^2(e^{-\beta}+\beta-1+4((2+\beta)e^{-\beta}+\beta-2)\rho^2)}{\beta^3} > 0$ indicates that the return distribution has fatter tails comparing with the distribution of a normally distributed random variable. These features are consistent with the empirical findings of the asset returns in the financial markets.

We study the statistical properties of the asset return and volatility processes via the explicit theoretical expression of their moments. Later, in our Monte Carlo simulation, we will apply these expression to examine the accuracy of our simulation by comparing the moments calculated from our simulation with those computed from true parameter values.

[3]For example, when $\beta = 0.2$, $e^{-\beta} + \beta - 1 = 0.018731$, when $\beta = 0.5$, $e^{-\beta} + \beta - 1 = 0.106531$, and when $\beta = 0.8$, $e^{-\beta} + \beta - 1 = 0.249329$.

4.3 ECF Estimation of the Affine Continuous-Time SV Model

Over the last two decades, a number of econometric methods have been developed to estimate the parameters of the continuous-time SV models including the affine diffusion process. Some are based on the moment conditions, others on simulations, or Bayesian method, etc. Broto and Ruiz (2004) provided an excellent survey about these methods. Since in Chapter one we already had a brief discussion about these approaches, in this section, we focus on discussing ECF estimation of the affine continuous-time SV model with volatility being latent and estimation with volatility being observed.

4.3.1 ECF Estimation of the Affine Continuous-Time SV Model When Volatility Is Latent

For the affine diffusion and affine jump diffusion processes, the analytical form CCF can be derived. Based on the characteristic functions, Singleton (2001), Jiang and Knight (2002), and Chacko and Viceira (2003) proposed the empirical characteristic function (ECF) procedure. The basic idea of the ECF method is to match the analytical CCF derived from the model to its empirical counterpart, ECF, which is calculated from the data. The advantage of using this approach is that it can achieve almost the same asymptotic efficiency as the ML method, while it avoids the difficulties inherent in deriving and maximizing the likelihood function.

Among these studies, Singleton (2001), and Chacko and Viceira (2003) used the conditional characteristic functions. In Singleton (2001), he proposed an

SMM procedure based on the CCF. The estimation is efficient, however, the stochastic process is simulated hence an approximation error is induced due to the discretization of the data generating process. Chacko and Viceira (2003) applied a GMM method based on the CCF which did not require discretization of the stochastic process. However, their method does not condition on all of the information available hence is not as efficient as that in Singleton (2001) or Jiang and Knight (2002). Unlike these two studies, Jiang and Knight (2002) used the unconditional joint characteristic functions and developed an efficient estimation procedure . In their study, Jiang and Knight (2002) employed both GMM and ECF procedures to estimate the affine continuous-time SV model. As they demonstrated, with the availability of an analytical expression for the unconditional joint characteristic function, the exact unconditional moments and cross moments of the state variable are readily derived, consequently applying GMM in this situation does not involve model discretization. On the other hand, choosing the optimal weighting function, the ECF method is able to provide asymptotically efficient estimates as maximum likelihood does. Both approaches do not involve model discretization nor simulation. However, in GMM approach, as volatility is latent, only the unconditional moments of asset return processes are used in the estimation. Also in ECF approach, as volatility is unobserved, it has to be integrated out of the joint characteristic function. The problem is that even though the bivariate stochastic process is a Markov process, the marginal return process is not, and in this case, there is no general solution for the choice of optimal weight function leading to increased computational burden in the estimation.

4.3.2 ECF Estimation of the Affine Continuous-Time SV Model When Volatility Is Observed

In their recent study, Jiang and Knight (2010) investigated the ECF estimation for the Markov models where the transition density function is unknown. In their ECF approach, a close-form approximation of the optimal weight function is derived by approximating the logarithmic transition density of the Markov process via the multivariate Edgeworth/Gram-Charlier expansion, hence the estimation is based on the analytical conditional cumulants. Their approach is similar to the approximate maximum likelihood estimation (AMLE) method proposed by Ait-Sahalia (2002) in which the Hermite polynomials are used to approximate the transition function. The major difference is that the Jiang and Knight (2010) approach guarantees the consistency while the AMLE procedure does not. As Jiang and Knight (2010) demonstrated, the C-AMLE method can be widely applied to Markov models, including the discrete-time SV models, the univariate discrete-time non-linear non-Gaussian process, the univariate continuous-time Gaussian process, the square-root diffusion process, the bivariate continuous-time Gaussian process, and the affine continuous-time SV model, etc. Jiang and Knight (2010) investigated the performance of the C-AMLE for the univariate continuous-time square-root diffusion process and the bivariate discrete-time SV models via a Monte Carlo study. The evidence showed that overall C-AMLE estimator had desirable finite sample performance in comparison with the GMM and the QML approaches.

With respect to the affine continuous-time SV model, when the asset return process together with return volatility process are both observable, the C-AMLE procedure is fairly straightforward and can be easily implemented. We now outline the C-AMLE approach.

Let $f(S_{t+1}|S_t;\theta)$ be the transition density function, in which S_t is a vector of state variables consisting of the asset return state variable x_t and the return volatility state variable v_t in the affine diffusion framework, i.e. $S_t = (x_t, v_t)$, and $\theta = \{\mu, \alpha, \beta, \rho, \sigma\}$ be a vector of unknown parameters that need to be estimated, define the CCF as:

$$\phi(r, S_{t+1}|S_t;\theta) = E[e^{ir'S_{t+1}}|S_t] = \int e^{ir'S_{t+1}} f(S_{t+1}|S_t;\theta)dS_{t+1}$$

where $r = \{r_1, r_2\}$ is the argument of the CCF.

The ECF is the sample counterpart of the CCF defined as:

$$\frac{1}{T}\sum e^{ir'S_{t+1}} = \int e^{ir'S_{t+1}} f_t(S_{t+1}|S_t;\theta)dS_{t+1}$$

where $f_t(S_{t+1}|S_t;\theta)$ is an empirical conditional density function.

Following Singleton (2001), and Jiang and Knight (2010), a consistent estimator based on the ECF can be derived by solving the following equation:

$$\frac{1}{T}\sum_{t=1}^{T}\int\dots\int \omega(r,t|S_t;\theta)(e^{ir'S_{t+1}} - \phi(r, S_{t+1}|S_t;\theta))dr = 0 \qquad (4.12)$$

where $\omega(r,t|S_t;\theta)$ represents the weight function.

As shown in the Lemma 1 in Jiang and Knight (2010), the optimal weight function can be derived by

$$\omega(r,t|S_t;\theta) = \frac{1}{2\pi}\int\dots\int \frac{\partial \ln f(S_{t+1}|S_t;\theta)}{\partial\theta} e^{-ir'S_{t+1}} dS_{t+1} \qquad (4.13)$$

where $lnf(S_{t+1}|S_t;\theta)$ is the logarithmic transition density function of the affine diffusion process. The above weight function is optimal in the sense that it leads to the estimators being equivalent to MLE. Clearly the optimal weight function depends on the logarithm of the transition density function. The issue is for the affine diffusion process, the transition density function does not have a closed form. Therefore, Jiang and Knight (2010) proposed an analytical approximation to the logarithmic transition density function via the Gram-Charlier/Edgeworth expansion:

$$
\begin{aligned}
\ln \hat{f}(S_{t+1}|S_t;\theta) = & \ln f_0(S_{t+1}|S_t;\theta) \\
& + \frac{1}{6}K^{i,j,k}h_{ijk}(S_{t+1}|S_t;\theta) \\
& + \frac{1}{24}K^{i,j,k,l}h_{ijkl}(S_{t+1}|S_t;\theta) \\
= & -\frac{N}{2}ln2\pi - \frac{1}{2}ln|\lambda^{i,j}| - \frac{1}{2}(S^i_{t+1} - \lambda^i)(S^j_{t+1} - \lambda^j)\lambda_{i,j} \\
& + \frac{1}{6}K^{i,j,k}h_{ijk}(S_{t+1}|S_t;\theta) \\
& + \frac{1}{24}K^{i,j,k,l}h_{ijkl}(S_{t+1}|S_t;\theta)
\end{aligned}
$$

where f_0 is a multivariate normal density. $K^{i,j,k}$ and $K^{i,j,k,l}$ represent the third and fourth order conditional cumulants while h_{ijk} and h_{ijkl} are the third and fourth order Hermite polynomial tensors[4].

The derivative of the $\ln \hat{f}(S_{t+1}|S_t;\theta)$ is given by:

$$
\begin{aligned}
\frac{\partial \ln \hat{f}(S_{t+1}|S_t;\theta)}{\partial\theta} = & -\frac{1}{2|\lambda^{i,j}|}\frac{\partial|\lambda^{i,j}|}{\partial\theta} + \frac{\partial\lambda^i}{\partial\theta}h_i - \frac{1}{2}(S^i_{t+1} - \lambda^i)(S^j_{t+1} - \lambda^j)\frac{\partial\lambda_{i,j}}{\partial\theta} \\
& + \frac{1}{6}[\frac{\partial K^{i,j,k}}{\partial\theta}h_{ijk} + K^{i,j,k}\frac{\partial h_{i,j,k}}{\partial\theta}] \\
& + \frac{1}{24}[\frac{\partial K^{i,j,k,l}}{\partial\theta}h_{ijkl} + K^{i,j,k,l}\frac{\partial h_{i,j,k,l}}{\partial\theta}]
\end{aligned}
$$

The optimal weight function is then approximated as:

$$
\hat{\omega}(r,t|S_t;\theta) = \frac{1}{2\pi}\int\ldots\int\frac{\partial \ln \hat{f}(S_{t+1}|S_t;\theta)}{\partial\theta}e^{-ir'S_{t+1}}dS_{t+1}
$$

[4] i, j, k, l take values 1 or 2 since there are two state variables x_{t+1}, v_{t+1}. The expression of f_0, the closed-form expression of the CGF, along with the expression of $K^{i,j,k}$, $K^{i,j,k,l}$, h_{ijk}, and h_{ijkl} are showed in Jiang and Knight (2010) and appendix 1 of this paper.

Employing the approximate optimal weight function, the C-AMLE procedure can be written as:

$$\frac{1}{T}\sum_{t=1}^{T}\int \hat{\omega}(r,t|S_t;\theta)(e^{ir'S_{t+1}}-\phi(r,S_{t+1}|S_t;\theta))dr=0 \tag{4.14}$$

From the above expression, the approximate ML estimator can be derived from:

$$\frac{1}{T}\sum_{t=1}^{T}(\frac{\partial \ln \hat{f}(S_{t+1}|S_t;\theta)}{\partial\theta}-E(\frac{\partial \ln \hat{f}(S_{t+1}|S_t;\theta)}{\partial\theta}|S_t))=0 \tag{4.15}$$

The estimating equation above is based on the approximate likelihood function, however, unlike the AMLE this approach includes the expectation of the approximate score thus ensures the consistency of the proposed estimator hence is the consistent AMLE.

Based on the analytical conditional cumulants, the system of estimation equations is following:

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T-1}\{\frac{\partial K^i}{\partial\theta}h_i-\frac{1}{2}\frac{\partial\lambda_{i,j}}{\partial\theta}[(S^i_{t+1}-K^i)(S^j_{t+1}-K^j)-K^{i,j}]\\
&+\frac{1}{6}[\frac{\partial K^{i,j,k}}{\partial\theta}(h_{ijk}-E(h_{ijk}|S_t))+3K^{i,j,k}((\bar{h}_ih_jh_k-E(\bar{h}_ih_jh_k|S_t))\\
&-\bar{z}_i(h_jh_k-E(h_jh_k|S_t))-\bar{h}_i\lambda_{j,k}-h_i\frac{\partial\lambda_{j,k}}{\partial\theta})]\\
&+\frac{1}{24}[\frac{\partial K^{i,j,k,l}}{\partial\theta}(h_{ijkl}-E(h_{ijkl}|S_t))+K^{i,j,k,l}(4(\bar{h}_ih_jh_kh_l-E(\bar{h}_ih_jh_kh_l|S_t))\\
&-4\bar{z}_i(h_jh_kh_l-E(h_jh_kh_l|S_t))-12(\bar{h}_ih_j\lambda_{k,l}-E(\bar{h}_ih_j\lambda_{k,l}|S_t))\\
&+12\bar{z}_ih_j\lambda_{k,l}-6(h_ih_j-E(h_ih_j|S_t))\frac{\partial\lambda_{k,l}}{\partial\theta})]\}=0
\end{aligned} \tag{4.16}$$

As θ is a vector consisting of five parameters, there are five equations correspondingly. Plugging the observations of $S_t=(x_t,v_t)$ into these equations, the solution yields the estimates for μ, α, β, ρ, and σ. The detailed discussion about these estimation equations is provided in the Appendix one.

4.4 Monte Carlo Experiments

In this section, we investigate via simulation the performance of the C-AMLE method for the affine continuous-time SV model. In our simulation, we treat the volatility as an observed state variable, hence we simulate both asset return and volatility observations and use these observations in the estimation.

The simulation of the continuous-time SV models has been a difficulty. As the transition density function is unknown, the continuous sampling observations are unable to be directly generated. The models usually have to be discretized. The most straightforward way is to discretize both asset return and return volatility processes by applying a first-order Euler scheme. However, when using direct Euler scheme, one has to investigate how to deal with negative values of the volatility process. Handling negative values in the wrong way would lead to extremely biased schemes. It is also noticed that the Euler scheme does not use any information of the analytical properties of the volatility process, hence may not capture the stylized facts the time series.

As we discussed in the previous section, in the affine continuous-time SV framework, the volatility process follows the square root stochastic process. The conditional volatility, i.e. $V_t|V_u$ for some $t > u$, follows a non-central chi-squared distribution with d degrees of freedom and non-centrality parameter λ.

$$V_t = \frac{\sigma^2(1-e^{-\beta(t-u)})}{4\beta}\chi^2_d(\lambda), t > u$$

where $\chi_d^2(\lambda)$ denotes the noncentral chi-squared random variable with:

$$d = \frac{4\alpha\beta}{\sigma^2}$$

$$\lambda = \frac{4\beta e^{-\beta(t-u)}}{\sigma^2(1-e^{-\beta(t-u)})} V_u$$

Given V_u, V_t is distributed as a constant term, $\frac{\sigma^2(1-e^{-\beta(t-u)})}{4\beta}$, times a noncentral chi-squared random variable. As Broadie and Kaya (2006) demonstrated, the non-central chi-squared distributed random variable, $\chi_d^2(\lambda)$, for $d > 1$, can be generated by first generating a central chi-squared random variable χ_{d-1}^2 and an independent standard normal random variable Z, then setting:

$$\chi_d^2(\lambda) = (Z + \sqrt{\lambda})^2 + \chi_{d-1}^2$$

Thus the stochastic volatility process can be exactly simulated from its distribution. The estimator is unbiased.

Set $u = 0$, then V_t is noncentral chi-squared distributed conditional on the initial value V_0. There are two ways to set initial value V_0. The first option is that we can simply set $V_0 = \alpha$ since α is the unconditional long run mean of the volatility process. Alternatively, as Jiang and Knight (2002) demonstrated, if the volatility process displayed the property of mean reversion, i.e. $\beta > 0$, then this volatility process was stationary and followed a gamma distribution, that is, $f(V_t = v_t) = \frac{\omega}{\Gamma(s)} v_t^{s-1} e^{-\omega v_t}$, where $\omega = \frac{2\beta}{\sigma^2}$, $s = \frac{2\alpha\beta}{\sigma^2}$, with mean α, and variance $\frac{\alpha\sigma^2}{2\beta}$. Based on their theoretical study, we first generate volatility from the gamma distribution then set it as our initial value V_0.

Simulating the asset return process is not as straightforward as simulating the volatility process, as the transition density function for the asset return process is unknown. Broadie and Kaya (2006) proposed an exact simulation method. Broadie and Kaya scheme is theoretically appealing, however, the practical implementation requires great computational effort. Alternatively, Van Haastrecht and Pelsser (2008) introduced an efficient approximation of the exact scheme. Van Haastrecht and Pelsser (2008) approach may not be as accurate as Broadie and Kaya (2006) method theoretically, however their approach is very easy to implement and as they showed, highly accurate. Hence we follow van Haastrecht and Pelsser (2008)'s study to generate the asset return process.

Allowing an instantaneous correlation between the asset return and the volatility processes, the SDE for the logarithmic stock prices can be written as:

$$dx_t = \mu dt + \sqrt{V_t}(\rho dB_{1t} + \sqrt{1-\rho^2} dB_{2t}) \tag{4.17}$$

The solution for the system SDEs can be obtained:

$$x_t = x_u + \mu\tau + \rho \int_u^t \sqrt{V_s} dB_{2s} + \sqrt{1-\rho^2} \int_u^t \sqrt{V_s} dB_{1s} \tag{4.18a}$$

$$V_t = V_u + \beta\alpha\tau - \beta \int_u^t V_s ds + \sigma \int_u^t \sqrt{V_s} dB_{2s} \tag{4.18b}$$

where $\tau = t - u$ for $u < t$ represents the time between observations.

The integrated variance process $\int_u^t V_s ds$ in the volatility equation can be approximated by using a drift interpolation method:

$$\int_u^t V_s ds \approx \tau \frac{V_t + V_u}{2}$$

Rearranging the volatility equation, $\int_u^t \sqrt{V_s} dB_{2s}$ can be computed as:

$$\int_u^t \sqrt{V_s} dB_{2s} \approx \frac{1}{\sigma}(V_t - V_s - \beta\alpha\tau + \beta\tau \frac{V_t + V_u}{2})$$

As the process for V_t is independent of the standard Brownian motion B_{1t}, the distribution of $\int_u^t \sqrt{V_s} dB_{1s}$ in the logarithmic asset price equation, given the path generated by V_t, is a normal distribution with mean 0 and variance $\int_u^t V_s ds$ which is approximated as $\tau \frac{V_t+V_u}{2}$.

$$\int_u^t \sqrt{V_s} dB_{1s} \sim_{approx} \sqrt{\tau \frac{V_t+V_u}{2}} Z_2$$

where Z_2 follows a standard normal distribution.

Hence, the sample path for logarithmic asset prices can be generated as:

$$x_t = x_u + \mu\tau - \frac{\rho\beta\alpha}{\sigma}\tau + (V_t + V_u)\frac{\rho(1+\frac{\beta\tau}{2})}{\sigma} - V_u\frac{2\rho}{\sigma} + \sqrt{\frac{(1-\rho^2)\tau}{2}}\sqrt{V_t + V_u} Z_2 \quad (4.19)$$

And the asset return process can be obtained by $r_\tau = x_t - x_{t-\tau}$.

The true parameter values are set as $\alpha = 0.867, \beta = 0.269, \rho = -.5, \sigma = 0.613$, and $\mu = 0.059$. These values are similar to those obtained in the empirical study of Jiang and Knight (2002). We set two sampling intervals, i.e. $\tau = 1/252, 1/12$ to generate daily and monthly observations respectively. When setting $\tau = 1/252$, we simulate 30 observations for each day, then the first 29 out of every 30 observations are discarded, leaving only the 30th observation at a daily frequency. The sample length is set as n=252 which corresponds to approximately one year of daily data. When setting $\tau = 1/12$, we simulate $22 \times 30 = 660$ observations for each month, then the first 659 out of every 660 observations are discarded, leaving only the 660th observation at a monthly frequency. The sample length is set as n=360 representing 30 years' transitions at the monthly frequency. Each simulated data series is initialized with the stock price at 100 hence the initial logarithmic stock price is $x_0 = 4.6052$. We first generate an initial 2,000 observations, then discard the first 1999 observations, only keep the last observations as the starting point to generate the data series used for the estimation. By doing so, we can mitigate the start-up effect. The number of replications in the Monte Carlo simulation is 10,000.

Figure 4.1 plots the mean of 10,000 sample paths along with the first 100 individual sample paths for the asset return processes when $\tau = 1/252$. Visual inspection shows that all 100 individual sample paths fluctuate around the mean path either above or below. The simulated asset return takes either a positive or a negative value with the mean being very close to zero. Figure 4.2 plots the mean of 10,000 sample paths and the first 100 individual sample paths for the return volatility processes when $\tau = 1/252$. All the volatility values are above zero hence the mean of the simulated return volatility values is positive. The individual sample paths fluctuate around the mean path. Figure 4.3-4.4 plot the mean of 10,000 sample paths along with the first 100 individual sample paths for the return and volatility processes when $\tau = 1/12$, and they display similar patterns. Overall, visual inspection suggests that our simulation method for the asset return and volatility processes is appropriate. We then proceed to a more thorough examination of our simulation accuracy.

Based on the conditional moment generating function (MGF), taking the expectation we can obtain the unconditional MGF, then the first four unconditional moments of both asset returns and volatilities can be derived based on the relationship between the moments and cumulants[5]. As these moments are explicit functions of the parameters $\mu, \alpha, \beta, \sigma, \rho$, and τ, plugging the true values into the expressions, we can calculate the true unconditional moments of both asset return and volatility. Also we can calculate the first four moments of our simulated asset return process as well as volatility process. From the comparison of the true moments and the moments from our simulated observations,

[5]The detailed derivatives are provided in Appendix 2.

we can examine the accuracy of our simulations.

Table 4.1 reports the true moments and the moments calculated from our simulation when $\tau = 1/252$. The top panel reports the first four moments of the asset return and the volatility processes calculated from the explicit expression of moments by plugging true parameters $\alpha = 0.867, \beta = 0.269, \rho = -0.5, \sigma = 0.613, \mu = 0.059, \tau = 1/252$. The bottom panel shows the first four moments of the generated return and volatility processes. We find all the four moments calculated from our simulations are very close to the true moments, suggesting that our simulations recover the properties of the distributions of the asset return and volatility processes.

Table 4.2 reports the true moments and the moments calculated from our simulation when $\tau = 1/12$. Same as in Table 4.1, the top panel reports the first four moments of the asset return and the volatility processes calculated from true parameters $\alpha = 0.867, \beta = 0.269, \rho = -0.5, \sigma = 0.613, \mu = 0.059, \tau = 1/12$. The bottom panel shows the first four moments of the two simulated processes. Comparing the true moments of return process in Table 4.1 and Table 4.2, we find both the mean and variance values of the monthly return process are much larger than those of daily return process. In addition, the skewness and kurtosis values of monthly return process are much larger than those of daily return process in absolute value, indicating that the monthly return process is more volatile, more negatively skewed and has fatter tails. Overall, all the four moments calculated from our simulation are very close to the true moments, suggesting that our simulations do a good job at capturing the properties of the distributions of the asset return and volatility processes.

We then estimate the model parameters applying the C-AMLE method. For each set of asset return and volatility observations, the C-AMLE yields the estimates of the five parameters. The estimation procedure is repeated 10000 times, we then calculate the mean, median, standard deviation, and the seventy fifth percentile. Table 4.3 reports the result when $\tau = 1/252$ and Table 4.4 reports the result when $\tau = 1/12$. The second column shows the true values, and third column reports the means of estimated parameters. The fourth to sixth columns report the medians, standard deviations, and the 75th percentiles, respectively.

In Table 4.3, most of the means of the estimates are very close to the true values. For example, the mean of the estimated α is 0.86, which is approximately one percent lower than the true value $\alpha = 0.867$; the mean of the estimated mean reversion parameter β is about three percent lower than the true value. The medians of the estimates are also close to the true parameters. The only exception is that for μ. We notice that the sign of the mean of estimates are negative whereas the true parameter takes a positive value. However, we observe that the median of the estimates of μ is not only positive but also very close to the true value, moreover, the 75th percentile of the estimates are close to the true parameter, suggesting that there exist some negative extreme values which distort the mean of the estimates. The standard deviation value is very large, confirming the above finding. For other estimates, the values of the standard deviations are quite small, suggesting that the estimates are stable.

From examination of Table 4.4, we find the estimates of α, ρ, σ are very close to the true parameters, while when using simulated monthly observations, the mean-reverting parameter β is relatively difficult to estimate, this finding is consistent with that in Jiang and Knight (2010). Same as using simulated daily observations, the parameter μ is the most difficult parameter to estimate. Comparing the values of mean, median, standard deviation and seventy fifth percentile, we find that in general the estimates are close to the true parameters, suggesting that the C-AMLE does a good job at recovering true parameters when volatility variable is directly observed.

Overall, the Monte Carlo result suggests that the C-AMLE procedure does a good job at estimating the affine continuous-time SV model when volatility is observed.

4.5 Conclusion and Extension

In this chapter, we investigated the estimation performance of the C-AMLE via a Monte Carlo study to the affine continuous-time SV model when volatility is observed. We generated volatility process based on its conditional density function and asset return process applying an efficient approximation of the exact scheme at both daily and monthly frequencies. Evidence showed that our simulation of return and volatility processes were accurate. We applied the C-AMLE approach for the affine continuous-time SV model using simulated observations at both daily and monthly frequencies, the results suggested that the C-AMLE did a good job at recovering the true parameters. When the C-AMLE procedure is applied to a univariate process, one does not have to worry about latent variable, estimation is very simple. In the bivariate system, volatility is latent, consequently, the implementation of the C-AMLE procedure is complicated. However, when treating volatility as observed, the implementation of the C-AMLE procedure is straightforward, and the estimates are reliable.

In next chapter, we continue our study of estimation of the affine continuous-time SV model via an empirical study. In particular, we will employ two different volatility proxies, namely, realized volatility and model-free implied volatility in our estimation, hence examine the performance of the C-AMLE method for the affine diffusion process when a volatility proxy is used.

4.6 Appendix

Appendix 1: C-AMLE Estimation of Affine SV Model:

The joint characteristic function of $(x_{t+\tau}, v_{t+\tau})$ conditional on the information structure F_t has the closed form as:

$$\psi(r_1, r_2; x_{t+\tau}, v_{t+\tau}|x_t, v_t)$$

$$\begin{aligned} &= E(exp(ir_1 x_{t+\tau} + ir_2 v_{t+\tau})|x_t, v_t) \\ &= exp(C(\tau; r_1, r_2) + D1(\tau; r_1, r_2)' x_t + D2(\tau; r_1, r_2)' v_t) \end{aligned}$$

where

$C(\tau; r_1, r_2) = (ir_1\mu + i\alpha\beta r_2)\tau + \frac{\alpha\beta}{\sigma^2}((b-h)\tau - 2ln(\frac{1-ge^{-h\tau}}{1-g}))$

$D1(\tau; r_1, r_2) = ir_1$

$D2(\tau; r_1, r_2) = ir_2 + \frac{b-h}{\sigma^2}\frac{1-e^{-h\tau}}{1-ge^{-h\tau}}$

with $b = \beta - \rho\sigma ir_1 - \sigma^2 ir_2$, $h = (b^2 + \sigma^2(r_1^2 + 2\rho\sigma r_1 r_2 + \sigma^2 r_2^2 + 2i\beta r_2))^{1/2}$, and $g = \frac{b-h}{b+h}$.

Let θ be the vector of parameters, i.e. $\theta = (\mu, \alpha, \beta, \rho, \sigma)$.

The cumulant generating function (CGF) thus is

$$\phi(r_1, r_2; x_{t+\tau}, v_{t+\tau}|x_t, v_t; \theta) = ln\psi(-ir_1, -ir_2; x_{t+\tau}, v_{t+\tau}|x_t, v_t; \theta)$$

Therefore the conditional cumulants can be derived by taking the derivatives of the CGF with respect to r_1 and r_2:

For $i = 1, 2, j = 1, 2, k = 1, 2, l = 1, 2$

$K^i = \frac{\partial}{\partial r_i}\psi(r_1, r_2; x_{t+\tau}, v_{t+\tau}|x_t, v_t)|r_1 = r_2 = 0$

$K^{i,j} = \frac{\partial^2}{\partial r_i \partial r_j}\psi(r_1, r_2; x_{t+\tau}, v_{t+\tau}|x_t, v_t)|r_1 = r_2 = 0$

$K^{i,j,k} = \frac{\partial^3}{\partial r_i \partial r_j \partial r_k}\psi(r_1, r_2; x_{t+\tau}, v_{t+\tau}|x_t, v_t)|r_1 = r_2 = 0$

$K^{i,j,k,l} = \frac{\partial^4}{\partial r_i \partial r_j \partial r_k \partial r_l}\psi(r_1, r_2; x_{t+\tau}, v_{t+\tau}|x_t, v_t)|r_1 = r_2 = 0$

The explicit expressions of the cumulants are calculated as:

$K^1 = \mu\tau$

$K^2 = \alpha(1 - e^{-\beta\tau}) + e^{-\beta\tau}v_t$

$K^{1,1} = \alpha\tau + \alpha\frac{1-e^{\beta\tau}}{\beta e^{\beta\tau}} + \frac{e^{\beta\tau}-1}{\beta e^{\beta\tau}}v_t$

$K^{1,2} = \frac{e^{-\beta\tau}}{\beta}(\alpha(e^{\beta\tau} - 1 - \beta\tau) + \beta\tau v_t)\rho\sigma$

$K^{2,2} = \frac{\alpha\sigma^2}{2\beta}(1 - e^{-\beta\tau})^2 + \frac{\sigma^2}{\beta}e^{-\beta\tau}(1 - e^{-\beta\tau})v_t$

$K^{1,1,1} = \frac{3e^{-\beta\tau}}{\beta^2}(\alpha(2 + \beta\tau - e^{\beta\tau}(2 - \beta\tau)) - (1 - e^{\beta\tau} + \beta\tau)v_t)\rho\sigma$

$K^{1,1,2} = \frac{e^{-2\beta\tau}}{2\beta^2}(2(1 - e^{\beta\tau}(1 - \beta\tau - \beta^2\tau^2\rho^2))v_t - \alpha(1 - e^{2\beta\tau}(1 + 4\rho^2) + 2e^{\beta\tau}(\beta\tau + 2\rho^2 + 2\beta\tau\rho^2 + \beta^2\tau^2\rho^2)))\sigma^2$

$K^{1,2,2} = \frac{e^{-2\beta\tau}}{\beta^2}(\alpha(e^{\beta\tau} - 1)(e^{\beta\tau} - 1 - \beta\tau) - (1 + 2\beta\tau - e^{\beta\tau}(1 + \beta\tau))v_t)\rho\sigma^3$

$K^{2,2,2} = \frac{\alpha\sigma^4}{2\beta^2}(1 - e^{-\beta\tau})^3 + \frac{3\sigma^4}{2\beta^2}e^{-\beta\tau}(1 - e^{-\beta\tau})^2 v_t$

$K^{1,1,1,1} = 3\sigma^2(\alpha - 2v_t)\frac{(1-e^{\beta\tau})^2}{2\beta^3 e^{2\beta\tau}} + 6\sigma^2(\alpha - v_t)\frac{\tau^2\rho^2\beta^2 - (2\rho^2+1)(e^{\beta\tau} - \beta\tau - 1)}{\beta^3 e^{\beta\tau}} + 3\alpha\sigma^2\frac{1 + e^{\beta\tau}(\beta\tau - 1) - 8\rho^2(e^{\beta\tau} - 1) + 4\rho^2\tau\beta(1 + e^{\beta\tau})}{\beta^3 e^{\beta\tau}}$

$K^{1,1,1,2} = -\frac{3\alpha\rho\sigma^3}{2\beta^3}(e^{-\beta\tau} - 1)(2\beta\tau e^{-\beta\tau} + 3e^{-\beta\tau} - 3) - \frac{\alpha\rho\sigma^3}{\beta^3}(3\beta^2\tau^2\rho^2 e^{-\beta\tau} + 3\beta^2\tau^2 e^{-\beta\tau} + 9e^{-\beta\tau} + 6\rho^2 e^{-\beta\tau} - 9 + 9\beta\tau e^{-\beta\tau} + 6\beta\tau\rho^2 e^{-\beta\tau} - 6\rho^2 + \beta^3\tau^3\rho^2 e^{-\beta\tau}) + \frac{\rho\sigma^3}{\beta^3}e^{-\beta\tau}(6e^{-\beta\tau} - 6 + 6\beta\tau e^{-\beta\tau} + 3\beta^2\tau^2 + \beta^3\tau^3\rho^2)v_t$

$K^{1,1,2,2} = \frac{e^{-3\beta\tau}}{2\beta^3}((-3 + 4e^{\beta\tau}(1 - \beta\tau - \rho^2 - 2\beta\tau\rho^2 - 2\beta^2\tau^2\rho^2) - e^{2\beta\tau}(1 - 2\beta\tau - 4\rho^2 - 4\beta\tau\rho^2 - 2\beta^2\tau^2\rho^2))v_t + \alpha(1 + e^{3\beta\tau}(1 + 6\rho^2) - e^{2\beta\tau}(1 + 12\rho^2 + 2\beta^2\tau^2\rho^2 + 2\beta(\tau + 4\tau\rho^2)) + e^{\beta\tau}(-1 + 6\rho^2 + 4\beta^2\tau^2\rho^2 + 2\beta(\tau + 4\tau\rho^2))))\sigma^4$

$K^{1,2,2,2} = -\frac{3\rho\sigma^5}{2\beta^3}(-\alpha + 3\alpha e^{-\beta\tau} - 2\alpha\beta\tau e^{-2\beta\tau} + \alpha\beta\tau e^{-3\beta\tau} + \alpha\beta\tau e^{-\beta\tau} - 3\alpha e^{-2\beta\tau} +$

$\alpha e^{-3\beta\tau} - 3e^{-3\beta\tau}\beta\tau v_t - 2e^{-3\beta\tau}v_t + 4e^{-2\beta\tau}v_t + 4e^{-2\beta\tau}\beta\tau v_t - 2e^{-\beta\tau}v_t - \beta\tau e^{-\beta\tau}v_t)$

$K^{2,2,2,2} = \frac{3\alpha\sigma^6}{4\beta^3}(1 - e^{-\beta\tau})^4 + \frac{3\sigma^6}{\beta^3}e^{-\beta\tau}(1 - e^{-\beta\tau})^3 v_t$

Denote $\Omega = \begin{pmatrix} K^{1,1} & K^{1,2} \\ K^{1,2} & K^{2,2} \end{pmatrix}$, $\lambda_{i,j}$ can be calculated:

$\lambda_{1,1} = (\Omega^{-1})_{1,1}$

$\lambda_{1,2} = \lambda_{2,1} = (\Omega^{-1})_{1,2} = (\Omega^{-1})_{2,1}$

$\lambda_{2,2} = (\Omega^{-1})_{2,2}$

Therefore[6],

$h_i = \lambda_{i,j}(S^j_{t+\tau} - K^j)$

$h_{i,j} = h_i h_j - \lambda_{i,j}$

$h_{i,j,k} = h_i h_j h_k - h_i \lambda_{j,k} - h_j \lambda_{i,k} - h_k \lambda_{i,j}$

$h_{i,j,k,l} = h_i h_j h_k h_l - h_i h_j \lambda_{k,l} - h_i h_k \lambda_{j,l} - h_i h_l \lambda_{j,k} - h_j h_k \lambda_{i,l} - h_j h_l \lambda_{i,k} - h_k h_l \lambda_{i,j} + \lambda_{i,j}\lambda_{k,l} + \lambda_{i,k}\lambda_{j,l} + \lambda_{i,l}\lambda_{j,k}$

$\bar{h}_i = \frac{\partial \lambda_{i,j}}{\partial \theta}(S^j_{t+\tau} - K^j)$

$\bar{z}_i = \lambda_{i,j}\frac{\partial K^j}{\partial \theta}$

For our model, Let $S_{t+\tau} = (x_{t+\tau}, v_{t+\tau})$ denote the vector of state variables. Let $\tau = 1$.

Let the initial approximating function be the multivariate norma density with mean vector $K^i, i = 1, 2$, and variance-covariance matrix $\lambda^{i,j}, i = 1, 2, j = 1, 2$.

$$f_0(S_{t+1}|S_t;\theta) = (2\pi)^{-N/2}|\lambda^{i,j}|^{-1/2}exp(-\tfrac{1}{2}(S^i_{t+1} - K^i)(S^j_{t+1} - K^j)\lambda_{i,j})$$

[6]According to tensor notation, any index repeated once as a subsript and once as a superscript is interpreted as sums over these repeated scripts. For example, $h_i = \lambda_{i,j}(S^j - K^j) = \lambda_{i,1}(x_{t+\tau} - K^1) + \lambda_{i,2}(v_{t+\tau} - K^2)$, $\bar{h}_i = \frac{\partial \lambda_{i,j}}{\partial \theta}(S^j_{t+\tau} - K^j) = \frac{\partial \lambda_{i,1}}{\partial \theta}(x_{t+\tau} - K^1) + \frac{\partial \lambda_{i,2}}{\partial \theta}(v_{t+\tau} - K^2)$, $\bar{z}_i = \lambda_{i,j}\frac{\partial K^j}{\partial \theta} = \lambda_{i,1}\frac{\partial K^1}{\partial \theta} + \lambda_{i,2}\frac{\partial K^2}{\partial \theta}$.

Following McCullage (1987), Jiang and Knight (2010), using the tensor notation, the log multivariate density $\ln f(S_{t+1}|S_t;\theta)$ can be approximated by using the general Gram-Charlier/Edgeworth expansion:

$$\begin{aligned}
\ln f(S_{t+1}|S_t;\theta) = & \ln f_0(S_{t+1}|S_t;\theta) \\
+ & \frac{1}{6}K^{i,j,k}h_{ijk} \\
+ & \frac{1}{24}K^{i,j,k,l}h_{ijkl} \\
= & \ln f_0(S_{t+1}|S_t;\theta) \\
+ & \frac{1}{6}[K^{1,1,1}h_{111} + 3K^{1,1,2}h_{112} + 3K^{1,2,2}h_{122} + K^{2,2,2}h_{222}] \\
+ & \frac{1}{24}[K^{1,1,1,1}h_{1111} + 4K^{1,1,1,2}h_{1112} + 6K^{1,1,2,2}h_{1122} \\
& +4K^{1,2,2,2}h_{1222} + K^{2,2,2,2}h_{2222}]
\end{aligned}$$

The approximate score function is derived:

$$
\begin{aligned}
\frac{\partial \ln f(S_{t+1}|S_t;\theta)}{\partial\theta} &= \frac{\partial \ln f_0(S_{t+1}|S_t;\theta)}{\partial\theta} \\
&+ \frac{1}{6}[\frac{\partial K^{i,j,k}}{\partial\theta}h_{ijk} + K^{i,j,k}\frac{\partial h_{ijk}}{\partial\theta}] \\
&+ \frac{1}{24}[\frac{\partial K^{i,j,k,l}}{\partial\theta}h_{ijkl} + K^{i,j,k,l}\frac{\partial h_{ijkl}}{\partial\theta}] \\
&= \frac{\partial \ln f_0(S_{t+1}|S_t;\theta)}{\partial\theta} \\
&+ \frac{1}{6}[\frac{\partial K^{1,1,1}}{\partial\theta}h_{111} + K^{1,1,1}\frac{\partial h_{111}}{\partial\theta} \\
&+ 3\frac{\partial K^{1,1,2}}{\partial\theta}h_{112} + 3K^{1,1,2}\frac{\partial h_{112}}{\partial\theta} \\
&+ 3\frac{\partial K^{1,2,2}}{\partial\theta}h_{122} + 3K^{1,2,2}\frac{\partial h_{122}}{\partial\theta} \\
&+ \frac{\partial K^{2,2,2}}{\partial\theta}h_{222} + K^{2.2.2}\frac{\partial h_{222}}{\partial\theta}] \\
&+ \frac{1}{24}[\frac{\partial K^{1,1,1,1}}{\partial\theta}h_{1111} + K^{1,1,1,1}\frac{\partial h_{1111}}{\partial\theta} \\
&+ 4\frac{\partial K^{1,1,1,2}}{\partial\theta}h_{1112} + 4K^{1,1,1,2}\frac{\partial h_{1112}}{\partial\theta} \\
&+ 6\frac{\partial K^{1,1,2,2}}{\partial\theta}h_{1122} + 6K^{1,1,2,2}\frac{\partial h_{1122}}{\partial\theta} \\
&+ 4\frac{\partial K^{1,2,2,2}}{\partial\theta}h_{1222} + 4K^{1,2,2,2}\frac{\partial h_{1222}}{\partial\theta} \\
&+ \frac{\partial K^{2,2,2,2}}{\partial\theta}h_{2222} + K^{2,2,2,2}\frac{\partial h_{2222}}{\partial\theta}]
\end{aligned}
$$

As $E(\frac{\partial \ln f_0(S_{t+1}|S_t;\theta)}{\partial\theta}) = 0$, the difference between the approximate score function and its expectation is:

$$
\frac{\partial \ln f(S_{t+1}|S_t;\theta)}{\partial\theta} - E(\frac{\partial \ln f(S_{t+1}|S_t;\theta)}{\partial\theta})
$$

$$
\begin{aligned}
= \quad & \frac{\partial \ln f_0(S_{t+1}|S_t;\theta)}{\partial\theta} \\
+ \quad & \frac{1}{6}[\frac{\partial K^{1,1,1}}{\partial\theta}(h_{111} - E(h_{111})) + K^{1,1,1}(\frac{\partial h_{111}}{\partial\theta} - E(\frac{\partial h_{111}}{\partial\theta})) \\
+ \quad & 3\frac{\partial K^{1,1,2}}{\partial\theta}(h_{112} - E(h_{112})) + 3K^{1,1,2}(\frac{\partial h_{112}}{\partial\theta} - E(\frac{\partial h_{112}}{\partial\theta})) \\
+ \quad & 3\frac{\partial K^{1,2,2}}{\partial\theta}(h_{122} - E(h_{122})) + 3K^{1,2,2}(\frac{\partial h_{122}}{\partial\theta} - E(\frac{\partial h_{122}}{\partial\theta})) \\
+ \quad & \frac{\partial K^{2,2,2}}{\partial\theta}(h_{222} - E(h_{222})) + K^{2,2,2}(\frac{\partial h_{222}}{\partial\theta} - E(\frac{\partial h_{222}}{\partial\theta}))] \\
+ \quad & \frac{1}{24}[\frac{\partial K^{1,1,1,1}}{\partial\theta}(h_{1111} - E(h_{1111})) + K^{1,1,1,1}(\frac{\partial h_{1111}}{\partial\theta} - E(\frac{\partial h_{1111}}{\partial\theta})) \\
+ \quad & 4\frac{\partial K^{1,1,1,2}}{\partial\theta}(h_{1112} - E(h_{1112})) + 4K^{1,1,1,2}(\frac{\partial h_{1112}}{\partial\theta} - E(\frac{\partial h_{1112}}{\partial\theta})) \\
+ \quad & 6\frac{\partial K^{1,1,2,2}}{\partial\theta}(h_{1122} - E(h_{1122})) + 6K^{1,1,2,2}(\frac{\partial h_{1122}}{\partial\theta} - E(\frac{\partial h_{1122}}{\partial\theta})) \\
+ \quad & 4\frac{\partial K^{1,2,2,2}}{\partial\theta}(h_{1222} - E(h_{1222})) + 4K^{1,2,2,2}(\frac{\partial h_{1222}}{\partial\theta} - E(\frac{\partial h_{1222}}{\partial\theta})) \\
+ \quad & \frac{\partial K^{2,2,2,2}}{\partial\theta}(h_{2222} - E(h_{2222})) + K^{2,2,2,2}(\frac{\partial h_{2222}}{\partial\theta} - E(\frac{\partial h_{2222}}{\partial\theta}))]
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial \ln f_0(S_{t+1}|S_t;\theta)}{\partial\theta} = \quad & -\frac{1}{2|\lambda^{i,j}|}\frac{\partial|\lambda^{i,j}|}{\partial\theta} - \frac{1}{2}(S^i - K^i)(S^j - K^j)\frac{\partial\lambda_{i,j}}{\partial\theta} \\
+ \quad & \frac{\partial K^1}{\partial\theta}\lambda_{1,j}(S^j - K^j) + \frac{\partial K^2}{\partial\theta}\lambda_{2,j}(S^j - K^j) \\
= \quad & -\frac{1}{2|\lambda^{i,j}|}\frac{\partial|\lambda^{i,j}|}{\partial\theta} - \frac{1}{2}(S^i - K^i)(S^j - K^j)\frac{\partial\lambda_{i,j}}{\partial\theta} + \frac{\partial K^i}{\partial\theta}h_i
\end{aligned}
$$

We have dropped the subscripts on S for each of notation.

$h_{111} = h_1^3 - 3h_1\lambda_{1,1}$

$h_{112} = h_1^2h_2 - 2h_1\lambda_{1,2} - h_2\lambda_{1,1}$

$h_{122} = h_1h_2^2 - h_1\lambda_{2,2} - 2h_2\lambda_{1,2}$

$h_{222} = h_2^3 - 3h_2\lambda_{2,2}$

$h_{1111} = h_1^4 - 6h_1^2\lambda_{1,1} + 3\lambda_{1,1}^2$

$h_{1112} = h_1^3 h_2 - 3h_1^2\lambda_{1,2} - 3h_1h_2\lambda_{1,1} + 3\lambda_{1,1}\lambda_{1,2}$

$h_{1122} = h_1^2h_2^2 - h_1^2\lambda_{2,2} - 4h_1h_2\lambda_{1,2} - h_2^2\lambda_{1,1} + \lambda_{1,1}\lambda_{2,2} + 2\lambda_{1,2}^2$

$h_{1222} = h_1h_2^3 - 3h_1h_2\lambda_{2,2} - 3h_2^2\lambda_{1,2} + 3\lambda_{1,2}\lambda_{2,2}$

$h_{2222} = h_2^4 - 6h_2^2\lambda_{2,2} + 3\lambda_{2,2}^2$

As in Chapter 2 of McCullagh (1987)[7], the relationship between moments and cumulants are:

$K^{ij} = K^{i,j} + K^iK^j$

$K^{ijk} = K^{i,j,k} + (K^iK^{j,k} + K^jK^{i,k} + K^kK^{i,j}) + K^iK^jK^k = K^{i,j,k} + K^iK^{j,k}[3] + K^iK^jK^k$

$K^{ijkl} = K^{i,j,k,l} + K^iK^{j,k,l}[4] + K^{i,j}K^{k,l}[3] + K^iK^jK^{k,l}[6] + K^iK^jK^kK^l$

where $K^{ij}, K^{ijk}, K^{ijkl}$ are moments, and $K^{i,j}, K^{i,j,k}, K^{i,j,k,l}$ are cumulants.

Moreover,

$$E(S^1 - K^1) = 0$$

$$E(S^2 - K^2) = 0$$

$$\begin{aligned} E[(S^1 - K^1)^2] = & \quad E[(S^1)^2 - 2K^1S^1 + (K^1)^2] \\ = & \quad K^{11} - (K^1)^2 \\ = & \quad (K^{1,1} + K^1K^1) - (K^1)^2 \\ = & \quad K^{1,1} \end{aligned}$$

[7]Details are in Chapter 2: Elementary theory of cumulants, *Tensor Methods in Statistics*, McCullagh P., 1987.

$$
\begin{aligned}
E[(S^1 - K^1)(S^2 - K^2)] = \quad & E[S^1S^2 - K^1S^2 - K^2S^1 + K^1K^2] \\
= \quad & K^{12} - K^1K^2 - K^2K^1 + K^1K^2 \\
= \quad & K^{12} - K^1K^2 \\
= \quad & (K^{1,2} + K^1K^2) - K^1K^2 \\
= \quad & K^{1,2}
\end{aligned}
$$

$$
\begin{aligned}
E[(S^2 - K^2)^2] = \quad & E[(S^2)^2 - 2K^2S^2 + (K^2)^2] \\
= \quad & K^{22} - (K^2)^2 \\
= \quad & (K^{2,2} + K^2K^2) - (K^2)^2 \\
= \quad & K^{2,2}
\end{aligned}
$$

$$
\begin{aligned}
E[(S^1 - K^1)^3] = \quad & E[(S^1)^3 - 3K^1(S^1)^2 + 3(K^1)^2S^1 - (K^1)^3] \\
= \quad & K^{111} - 3K^1K^{11} + 3(K^1)^2K^1 - (K^1)^3 \\
= \quad & (K^{1,1,1} + 3K^1K^{1,1} + (K^1)^3) - 3K^1(K^{1,1} + (K^1)^2) + 2(K^1)^3 \\
= \quad & K^{1,1,1}
\end{aligned}
$$

$$
\begin{aligned}
E[(S^1 - K^1)^2(S^2 - K^2)] = \quad & E[(S^1)^2S^2 - K^2(S^1)^2 - 2K^1S^1S^2 \\
+ \quad & 2K^1K^2S^1 + (K^1)^2S^2 - (K^1)^2K^2] \\
= \quad & K^{112} - K^2K^{11} - 2K^1K^{12} + 2(K^1)^2K^2 \\
= \quad & (K^{1,1,2} + 2K^1K^{1,2} + K^2K^{1,1} + (K^1)^2K^2) \\
- \quad & K^2(K^{1,1} + K^1K^1) - 2K^1(K^{1,2} + K^1K^2) + 2(K^1)^2K^2 \\
= \quad & K^{1,1,2}
\end{aligned}
$$

$$
\begin{aligned}
E[(S^1 - K^1)(S^2 - K^2)^2] = \quad & E[S^1(S^2)^2 - 2K^2S^1S^2 + (K^2)^2S^1 \\
- \quad & K^1(S^2)^2 + 2K^1K^2S^2 - K^1(K^2)^2 \\
= \quad & K^{122} - 2K^2K^{12} - K^1K^{22} + 2K^1(K^2)^2 \\
= \quad & (K^{1,2,2} + K^1K^{2,2} + 2K^2K^{1,2} + K^1(K^2)^2) \\
- \quad & 2K^2(K^{1,2} + K^1K^2) - K^1(K^{2,2} + (K^2)^2) + 2K^1(K^2)^2 \\
= \quad & K^{1,2,2}
\end{aligned}
$$

$$
\begin{aligned}
E[(S^2 - K^2)^3] = \quad & E[(S^2)^3 - 3K^2(S^2)^2 + 3(K^2)^2S^2 - (K^2)^3] \\
= \quad & K^{222} - 3K^2K^{22} + 3(K^2)^2K^2 - (K^2)^3 \\
= \quad & (K^{2,2,2} + 3K^2K^{2,2} + (K^2)^3) - 3K^2(K^{2,2} + (K^2)^2) + 2(K^2)^3 \\
= \quad & K^{2,2,2}
\end{aligned}
$$

$$
\begin{aligned}
E[(S^1 - K^1)^4] = \quad & E[(S^1)^4 - 4K^1(S^1)^3 + 6(K^1)^2(S^1)^2 - 4(K^1)^3S^1 + (K^1)^4] \\
= \quad & K^{1111} - 4K^1K^{111} + 6(K^1)^2K^{11} - 3(K^1)^4 \\
= \quad & (K^{1,1,1,1} + 4K^1K^{1,1,1} + 3(K^{1,1})^2 + 6(K^1)^2K^{1,1} + (K^1)^4) \\
- \quad & 4K^1(K^{1,1,1} + 3K^1K^{1,1} + (K^1)^3) + 6(K^1)^2(K^{1,1} + K^1K^1) - 3(K^1)^4 \\
= \quad & K^{1,1,1,1} + 3(K^{1,1})^2
\end{aligned}
$$

$$
\begin{aligned}
E[(S^1-K^1)^3(S^2-K^2)] = \quad & E[(S^1)^3S^2 - K^1(S^1)^2S^2 - K^2(S^1)^3 + K^1K^2(S^1)^2 \\
- \quad & 2K^1(S^1)^2S^2 + 2(K^1)^2S^1S^2 + 2K^1K^2(S^1)^2 - 2(K^1)^2K^2S^1 \\
+ \quad & (K^1)^2S^1S^2 - (K^1)^3S^2 - (K^1)^2K^2S^1 + (K^1)^3K^2 \\
= \quad & K^{1112} - 3K^1K^{112} - K^2K^{111} \\
+ \quad & 3K^1K^2K^{11} + 3(K^1)^2K^{12} - 3(K^1)^3K^2 \\
= \quad & (K^{1,1,1,2} + 3K^1K^{1,1,2} + K^2K^{1,1,1} + 3K^{1,1}K^{1,2} \\
+ \quad & 3(K^1)^2K^{1,2} + 3K^1K^2K^{1,1} + (K^1)^3K^2) \\
- \quad & 3K^1(K^{1,1,2} + 2K^1K^{1,2} + K^2K^{1,1} + (K^1)^2K^2) \\
- \quad & K^2(K^{1,1,1} + 3K^1K^{1,1} + (K^1)^3) \\
+ \quad & 3K^1K^2(K^{1,1} + K^1K^1) \\
+ \quad & 3(K^1)^2(K^{1,2} + K^1K^2) - 3(K^1)^3K^2 \\
= \quad & K^{1,1,1,2} + 3K^{1,1}K^{1,2}
\end{aligned}
$$

$$
\begin{aligned}
E[(S^1-K^1)^2(S^2-K^2)^2] = {} & E[(S^1)^2(S^2)^2 - 2K^2(S^1)^2S^2 + (K^2)^2(S^1)^2 \\
- {} & 2K^1S^1(S^2)^2 + 4K^1K^2S^1S^2 - 2K^1(K^2)^2S^1 \\
+ {} & (K^1)^2(S^2)^2 - 2(K^1)^2K^2S^2 + (K^1)^2(K^2)^2] \\
= {} & K^{1122} - 2K^2K^{112} - 2K^1K^{122} + (K^2)^2K^{11} \\
+ {} & 4K^1K^2K^{12} + (K^1)^2K^{22} - 3(K^1)^2(K^2)^2 \\
= {} & (K^{1,1,2,2} + 2K^1K^{1,2,2,} + 2K^2K^{1,1,2} + K^{1,1}K^{2,2} + 2(K^{1,2})^2 \\
+ {} & (K^1)^2K^{2,2} + 4K^1K^2K^{1,2} + (K^2)^2K^{1,1} + (K^1)^2(K^2)^2) \\
- {} & 2K^2(K^{1,1,2} + 2K^1K^{1,2} + K^2K^{1,1} + (K^1)^2K^2) \\
- {} & 2K^1(K^{1,2,2} + K^1K^{2,2} + 2K^2K^{1,2} + K^1(K^2)^2) \\
+ {} & (K^2)^2(K^{1,1} + K^1K^1) \\
+ {} & 4K^1K^2(K^{1,2} + K^1K^2) \\
+ {} & (K^1)^2(K^{2,2} + K^2K^2) \\
- {} & 3(K^1)^2(K^2)^2 \\
= {} & K^{1,1,2,2} + K^{1,1}K^{2,2} + 2(K^{1,2})^2
\end{aligned}
$$

$$
\begin{aligned}
E[(S^1 - K^1)(S^2 - K^2)^3] = \quad & E[S^1(S^2)^3 - 2K^2S^1(S^2)^2 + (K^2)^2S^1S^2 - K^1(S^2)^3 \\
+ \quad & 2K^1K^2(S^2)^2 - K^1(K^2)^2S^2 - K^2S^1(S^2)^2 + 2(K^2)^2S^1S^2 \\
- \quad & (K^2)^3S^1 + K^1K^2(S^2)^2 - 2K^1(K^2)^2S^2 + K^1(K^2)^3] \\
= \quad & K^{1222} - 3K^2K^{122} - K^1K^{222} \\
+ \quad & 3(K^2)^2K^{12} + 3K^1K^2K^{22} - 3K^1(K^2)^3 \\
= \quad & (K^{1,2,2,2} + K^1K^{2,2,2} + 3K^2K^{1,2,2} + 3K^{1,2}K^{2,2} \\
+ \quad & 3K^1K^2K^{2,2} + 3(K^2)^2K^{1,2} + K^1(K^2)^3) \\
- \quad & 3K^2(K^{1,2,2} + K^1K^{2,2} + 2K^2K^{1,2} + K^1(K^2)^2) \\
- \quad & K^1(K^{2,2,2} + 3K^2K^{2,2} + (K^2)^3) \\
+ \quad & 3(K^2)^2(K^{1,2} + K^1K^2) \\
+ \quad & 3K^1K^2(K^{2,2} + K^2K^2) \\
- \quad & 3K^1(K^2)^3 \\
= \quad & K^{1,2,2,2} + 3K^{1,2}K^{2,2}
\end{aligned}
$$

$$
\begin{aligned}
E[(S^2 - K^2)^4] = \quad & E[(S^2)^4 - 2K^2(S^2)^3 + (K^2)^2(S^2)^2 - 2K^2(S^2)^3 + 4(K^2)^2(S^2)^2 \\
- \quad & 2(K^2)^3S^2 + (K^2)^2(S^2)^2 - 2(K^2)^3S^2 + (K^2)^4] \\
= \quad & K^{2222} - 4K^2K^{222} + 6(K^2)^2K^{22} - 3(K^2)^4 \\
= \quad & (K^{2,2,2,2} + 4K^2K^{2,2,2} + 3(K^{2,2})^2 + 6(K^2)^2K^{2,2} + (K^2)^4) \\
- \quad & 4K^2(K^{2,2,2} + 3K^2K^{2,2} + (K^2)^3) + 6(K^2)^2(K^{2,2} + K^2K^2) \\
- \quad & 3(K^2)^4 \\
= \quad & K^{2,2,2,2} + 3(K^{2,2})^2
\end{aligned}
$$

$$E(h_i) = 0$$

$$E(\bar{h}_i) = 0$$

$$E(\bar{z}_i) = \bar{z}_i$$

$$E(\lambda_{i,j}) = \lambda_{i,j}$$

$$\begin{aligned}
E(h_1^2) = \quad & E(\lambda_{1,1}^2(S^1 - K^1)^2 + 2\lambda_{1,1}\lambda_{1,2}(S^1 - K^1)(S^2 - K^2) + \lambda_{1,2}^2(S^2 - K^2)^2) \\
= \quad & \lambda_{1,1}^2 K^{1,1} + 2\lambda_{1,1}\lambda_{1,2}K^{1,2} + \lambda_{1,2}^2 K^{2,2} \\
= \quad & \frac{(K^{2,2})^2 K^{1,1} - 2(K^{1,2})^2 K^{2,2} + (K^{1,2})^2 K^{2,2}}{(K^{1,1}K^{2,2} - (K^{1,2})^2)^2} \\
= \quad & \frac{K^{2,2}(K^{1,1}K^{2,2} - (K^{1,2})^2)}{(K^{1,1}K^{2,2} - (K^{1,2})^2)^2} \\
= \quad & \lambda_{1,1}
\end{aligned}$$

$$\begin{aligned}
E(h_1 h_2) = \quad & E[\lambda_{1,1}\lambda_{1,2}(S^1 - K^1)^2 + (\lambda_{1,1}\lambda_{2,2} + \lambda_{1,2}^2)(S^1 - K^1)(S^2 - K^2) + \lambda_{1,2}\lambda_{2,2}(S^2 - K^2)^2] \\
= \quad & \lambda_{1,1}\lambda_{1,2}K^{1,1} + (\lambda_{1,1}\lambda_{2,2} + \lambda_{1,2}^2)K^{1,2} + \lambda_{1,2}\lambda_{2,2}K^{2,2} \\
= \quad & \frac{-K^{2,2}K^{1,2}K^{1,1} + (K^{2,2}K^{1,1} + (K^{1,2})^2)K^{1,2} - K^{1,2}K^{1,1}K^{2,2}}{(K^{1,1}K^{2,2} - (K^{1,2})^2)^2} \\
= \quad & \frac{-K^{1,2}(K^{1,1}K^{2,2} - (K^{1,2})^2)}{(K^{1,1}K^{2,2} - (K^{1,2})^2)^2} \\
= \quad & \lambda_{1,2}
\end{aligned}$$

$$\begin{aligned}
E(h_2^2) = \quad & E(\lambda_{1,2}^2(S_1 - K^1)^2 + 2\lambda_{1,2}\lambda_{2,2}(S^1 - K^1)(S^2 - K^2) + \lambda_{2,2}^2(S^2 - K^2)^2) \\
= \quad & \lambda_{1,2}^2 K^{1,1} + 2\lambda_{1,2}\lambda_{2,2}K^{1,2} + \lambda_{2,2}^2 K^{2,2} \\
= \quad & \frac{(K^{1,2})^2 K^{1,1} - 2(K^{1,2})K^{1,1}K^{1,2} + (K^{1,1})^2 K^{2,2}}{(K^{1,1}K^{2,2} - (K^{1,2})^2)^2} \\
= \quad & \frac{K^{1,1}(K^{1,1}K^{2,2} - (K^{1,2})^2)}{(K^{1,1}K^{2,2} - (K^{1,2})^2)^2} \\
= \quad & \lambda_{2,2}
\end{aligned}$$

$$\begin{aligned}
E(\bar{h}_1 h_1) = \quad & E(\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}(S^1 - K^1)^2 + (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1})(S^1 - K^1)(S^2 - K^2) \\
+ \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}(S^2 - K^2)^2) \\
= \quad & \frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}K^{1,1} + (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1})K^{1,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}K^{2,2}
\end{aligned}$$

$$
\begin{aligned}
E(\bar{h}_1 h_2) = \quad & E(\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}(S^1-K^1)^2 + (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{2,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2})(S^1-K^1)(S^2-K^2) \\
+ \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{2,2}(S^2-K^2)^2) \\
= \quad & \frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}K^{1,1} + (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{2,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2})K^{1,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{2,2}K^{2,2}
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_2 h_1) = \quad & E(\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}(S^1-K^1)^2 + (\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2} + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,1})(S^1-K^1)(S^2-K^2) \\
+ \quad & \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,2}(S^2-K^2)^2) \\
= \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}K^{1,1} + (\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2} + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,1})K^{1,2} + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,2}K^{2,2}
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_2 h_2) = \quad & E(\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}(S^1-K^1)^2 + (\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{2,2} + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,2})(S^1-K^1)(S^2-K^2) \\
+ \quad & \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{2,2}(S^2-K^2)^2) \\
= \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}K^{1,1} + (\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{2,2} + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,2})K^{1,2} + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{2,2}K^{2,2}
\end{aligned}
$$

$$
\begin{aligned}
E(h_{111}) = \quad & E(h_1^3 - 3h_1\lambda_{1,1}) = E(h_1^3) \\
= \quad & E(\lambda_{1,1}^3(S^1-K^1)^3 + 3\lambda_{1,1}^2\lambda_{1,2}(S^1-K^1)^2(S^2-K^2) \\
+ \quad & 3\lambda_{1,1}\lambda_{1,2}^2(S^1-K^1)(S^2-K^2)^2 + \lambda_{1,2}^3(S^2-K^2)^3) \\
= \quad & \lambda_{1,1}^3K^{1,1,1} + 3\lambda_{1,1}^2\lambda_{1,2}K^{1,1,2} + 3\lambda_{1,1}\lambda_{1,2}^2K^{1,2,2} + \lambda_{1,2}^3K^{2,2,2}
\end{aligned}
$$

$$
\begin{aligned}
E(h_{112}) = \quad & E(h_1^2h_2 - 2h_1\lambda_{1,2} - h_2\lambda_{1,1}) = E(h_1^2h_2) \\
= \quad & E(\lambda_{1,1}^2\lambda_{1,2}(S^1-K^1)^3 + 2\lambda_{1,1}\lambda_{1,2}^2(S^1-K^1)^2(S^2-K^2) \\
+ \quad & \lambda_{1,2}^3(S^1-K^1)(S^2-K^2)^2 + \lambda_{1,1}^2\lambda_{2,2}(S^1-K^1)^2(S^2-K^2) \\
+ \quad & 2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}(S^1-K^1)(S^2-K^2)^2 + \lambda_{1,2}^2\lambda_{2,2}(S^2-K^2)^3) \\
= \quad & \lambda_{1,1}^2\lambda_{1,2}K^{1,1,1} + \lambda_{1,1}K^{1,1,2}(2\lambda_{1,2}^2 + \lambda_{1,1}\lambda_{2,2}) \\
+ \quad & \lambda_{1,2}K^{1,2,2}(\lambda_{1,2}^2 + 2\lambda_{1,1}\lambda_{2,2}) + \lambda_{1,2}^2\lambda_{2,2}K^{2,2,2}
\end{aligned}
$$

$$
\begin{aligned}
E(h_{122}) = \quad & E(h_1 h_2^2 - h_1\lambda_{2,2} - 2h_2\lambda_{1,2}) = E(h_1 h_2^2) \\
= \quad & E(\lambda_{1,1}\lambda_{1,2}^2(S^1-K^1)^3 + 2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}(S^1-K^1)^2(S^2-K^2) \\
+ \quad & \lambda_{1,1}\lambda_{2,2}^2(S^1-K^1)(S^2-K^2)^2 + \lambda_{1,2}^3(S^1-K^1)^2(S^2-K^2) \\
+ \quad & 2\lambda_{1,2}^2\lambda_{2,2}(S^1-K^1)(S^2-K^2)^2 + \lambda_{1,2}\lambda_{2,2}^2(S^2-K^2)^3) \\
= \quad & \lambda_{1,1}\lambda_{1,2}^2 K^{1,1,1} + \lambda_{1,2}K^{1,1,2}(\lambda_{1,2}^2 + 2\lambda_{1,1}\lambda_{2,2}) \\
+ \quad & \lambda_{2,2}K^{1,2,2}(2\lambda_{1,2}^2 + \lambda_{1,1}\lambda_{2,2}) + \lambda_{1,2}\lambda_{2,2}^2 K^{2,2,2}
\end{aligned}
$$

$$
\begin{aligned}
E(h_{222}) = \quad & E(h_2^3 - 3h_2\lambda_{2,2}) = E(h_2^3) \\
= \quad & E(\lambda_{1,2}^3(S^1-K^1)^3 + 3\lambda_{1,2}^2\lambda_{2,2}(S^1-K^1)^2(S^2-K^2) \\
+ \quad & 3\lambda_{1,2}\lambda_{2,2}^2(S^1-K^1)(S^2-K^2)^2 + \lambda_{2,2}^3(S^2-K^2)^3) \\
= \quad & \lambda_{1,2}^3 K^{1,1,1} + 3\lambda_{1,2}^2\lambda_{2,2}K^{1,1,2} + 3\lambda_{1,2}\lambda_{2,2}^2 K^{1,2,2} + \lambda_{2,2}^3 K^{2,2,2}
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_1 h_1^2) = \quad & E[\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}^2(S^1-K^1)^3 + \lambda_{1,1}(2\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1})(S^1-K^1)^2(S^2-K^2) \\
+ \quad & \lambda_{1,2}(\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2} + 2\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1})(S^1-K^1)(S^2-K^2)^2 + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^2(S^2-K^2)^3] \\
= \quad & \frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}^2 K^{1,1,1} + \lambda_{1,1}(2\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1})K^{1,1,2} \\
+ \quad & \lambda_{1,2}(\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2} + 2\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1})K^{1,2,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^2 K^{2,2,2}
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_1 h_1 h_2) = \quad & E[\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}\lambda_{1,2}(S^1 - K^1)^3 \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}\lambda_{2,2} + \frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}^2 + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}\lambda_{1,2})(S^1 - K^1)^2(S^2 - K^2) \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}\lambda_{2,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}\lambda_{2,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^2)(S^1 - K^1)(S^2 - K^2)^2 \\
+ \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}\lambda_{2,2}(S^2 - K^2)^3] \\
= \quad & \frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}\lambda_{1,2}K^{1,1,1} \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}\lambda_{2,2} + \frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}^2 + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}\lambda_{1,2})K^{1,1,2} \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}\lambda_{2,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}\lambda_{2,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^2)K^{1,2,2} \\
+ \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}\lambda_{2,2}K^{2,2,2}
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_1 h_2^2) = \quad & E[\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}^2(S^1 - K^1)^3 + (2\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}\lambda_{2,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^2)(S^1 - K^1)^2(S^2 - K^2) \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{2,2}^2 + 2\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}\lambda_{2,2})(S^1 - K^1)(S^2 - K^2)^2 + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{2,2}^2(S^2 - K^2)^3] \\
= \quad & \frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}^2K^{1,1,1} + (2\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}\lambda_{2,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^2)K^{1.1.2} \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{2,2}^2 + 2\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}\lambda_{2,2})K^{1.2.2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{2,2}^2K^{2,2,2}
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_2 h_1^2) = \quad & E[\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}^2(S^1 - K^1)^3 + \lambda_{1,1}(2\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2} + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,1})(S^1 - K^1)^2(S^2 - K^2) \\
+ \quad & \lambda_{1,2}(\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2} + 2\frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,1})(S^1 - K^1)(S^2 - K^2)^2 + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,2}^2(S^2 - K^2)^3] \\
= \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}^2K^{1,1,1} + \lambda_{1,1}(2\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2} + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,1})K^{1,1,2} \\
+ \quad & \lambda_{1,2}(\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2} + 2\frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,1})K^{1,2,2} + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,2}^2K^{2,2,2}
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_2 h_1 h_2) = \quad & E[\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,1}\lambda_{1,2}(S^1 - K^1)^3 \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,1}\lambda_{2,2} + \frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}^2 + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,1}\lambda_{1,2})(S^1 - K^1)^2(S^2 - K^2) \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2} + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,1}\lambda_{2,2} + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}^2)(S^1 - K^1)(S^2 - K^2)^2 \\
+ \quad & \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2}(S^2 - K^2)^3] \\
= \quad & \frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,1}\lambda_{1,2}K^{1,1,1} \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,1}\lambda_{2,2} + \frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}^2 + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,1}\lambda_{1,2})K^{1,1,2} \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2} + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,1}\lambda_{2,2} + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}^2)K^{1,2,2} \\
+ \quad & \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2}K^{2,2,2}
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_2 h_2^2) = \quad & E[\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}^2(S^1 - K^1)^3 + (2\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2} + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}^2)(S^1 - K^1)^2(S^2 - K^2) \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{2,2}^2 + 2\frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2})(S^1 - K^1)(S^2 - K^2)^2 + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{2,2}^2(S^2 - K^2)^3] \\
= \quad & \frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}^2K^{1,1,1} + (2\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2} + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}^2)K^{1,1,2} \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{2,2}^2 + 2\frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2})K^{1,2,2} + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{2,2}^2K^{2,2,2}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial h_{111}}{\partial \theta} = \quad & 3h_1^2\frac{\partial h_1}{\partial \theta} - 3\frac{\partial h_1}{\partial \theta}\lambda_{1,1} - 3h_1\frac{\partial \lambda_{1,1}}{\partial \theta} \\
= \quad & 3\bar{h}_1 h_1^2 - 3\bar{z}_1 h_1^2 - 3\bar{h}_1\lambda_{1,1} + 3\bar{z}_1\lambda_{1,1} - 3h_1\frac{\partial \lambda_{1,1}}{\partial \theta}
\end{aligned}
$$

$$
E(\frac{\partial h_{111}}{\partial \theta}) = 3E(\bar{h}_1 h_1^2) - 3\bar{z}_1 E(h_1^2) + 3\bar{z}_1\lambda_{1,1}
$$

$$
\frac{\partial h_{111}}{\partial \theta} - E(\frac{\partial h_{111}}{\partial \theta}) = \quad 3[\bar{h}_1 h_1^2 - E(\bar{h}_1 h_1^2) - \bar{z}_1(h_1^2 - E(h_1^2)) - \bar{h}_1\lambda_{1,1} - h_1\frac{\partial \lambda_{1,1}}{\partial \theta}]
$$

$$\begin{aligned}
\frac{\partial h_{112}}{\partial\theta} &= 2h_1h_2\frac{\partial h_1}{\partial\theta} + h_1^2\frac{\partial h_2}{\partial\theta} - 2\frac{\partial h_1}{\partial\theta}\lambda_{1,2} \\
&- 2h_1\frac{\partial\lambda_{1,2}}{\partial\theta} - \frac{\partial h_2}{\partial\theta}\lambda_{1,1} - h_2\frac{\partial\lambda_{1,1}}{\partial\theta} \\
&= 2\bar{h}_1h_1h_2 - 2\bar{z}_1h_1h_2 + \bar{h}_2h_1^2 - \bar{z}_2h_1^2 - 2\bar{h}_1\lambda_{1,2} \\
&+ 2\bar{z}_1\lambda_{1,2} - 2h_1\frac{\partial\lambda_{1,2}}{\partial\theta} - \bar{h}_2\lambda_{1,1} + \bar{z}_2\lambda_{1,1} - h_2\frac{\partial\lambda_{1,1}}{\partial\theta}
\end{aligned}$$

$$E(\frac{\partial h_{112}}{\partial\theta}) = 2E(\bar{h}_1h_1h_2) - 2\bar{z}_1E(h_1h_2) + E(\bar{h}_2h_1^2) - \bar{z}_2E(h_1^2) + 2\bar{z}_1\lambda_{1,2} + \bar{z}_2\lambda_{1,1}$$

$$\begin{aligned}
\frac{\partial h_{112}}{\partial\theta} - E(\frac{\partial h_{112}}{\partial\theta}) &= 2(\bar{h}_1h_1h_2 - E(\bar{h}_1h_1h_2)) - 2\bar{z}_1(h_1h_2 - E(h_1h_2)) \\
&+ (\bar{h}_2h_1^2 - E(\bar{h}_2h_1^2)) - \bar{z}_2(h_1^2 - E(h_1^2)) - 2\bar{h}_1\lambda_{1,2} \\
&- 2h_1\frac{\partial\lambda_{1,2}}{\partial\theta} - \bar{h}_2\lambda_{1,1} - h_2\frac{\partial\lambda_{1,1}}{\partial\theta}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial h_{122}}{\partial\theta} &= \frac{\partial h_1}{\partial\theta}h_2^2 + 2h_1h_2\frac{\partial h_2}{\partial\theta} - \frac{\partial h_1}{\partial\theta}\lambda_{2,2} - h_1\frac{\partial\lambda_{2,2}}{\partial\theta} - 2\frac{\partial h_2}{\partial\theta}\lambda_{1,2} - 2h_2\frac{\partial\lambda_{1,2}}{\partial\theta} \\
&= \bar{h}_1h_2^2 - \bar{z}_1h_2^2 + 2\bar{h}_2h_1h_2 - 2\bar{z}_2h_1h_2 - \bar{h}_1\lambda_{2,2} + \bar{z}_1\lambda_{2,2} \\
&- h_1\frac{\partial\lambda_{2,2}}{\partial\theta} - 2\bar{h}_2\lambda_{1,2} + 2\bar{z}_2\lambda_{1,2} - 2h_2\frac{\partial\lambda_{1,2}}{\partial\theta}
\end{aligned}$$

$$E(\frac{\partial h_{122}}{\partial\theta}) = E(\bar{h}_1h_2^2) - \bar{z}_1E(h_2^2) + 2E(\bar{h}_2h_1h_2) - 2\bar{z}_2E(h_1h_2) + \bar{z}_1\lambda_{2,2} + 2\bar{z}_2\lambda_{1,2}$$

$$\begin{aligned}
\frac{\partial h_{122}}{\partial\theta} - E(\frac{\partial h_{122}}{\partial\theta}) &= (\bar{h}_1h_2^2 - E(\bar{h}_1h_2^2)) - \bar{z}_1(h_2^2 - E(h_2^2)) \\
&+ 2(\bar{h}_2h_1h_2 - E(\bar{h}_2h_1h_2)) - 2\bar{z}_2(h_1h_2 - E(h_1h_2)) \\
&- \bar{h}_1\lambda_{2,2} - h_1\frac{\partial\lambda_{2,2}}{\partial\theta} - 2\bar{h}_2\lambda_{1,2} - 2h_2\frac{\partial\lambda_{1,2}}{\partial\theta}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial h_{222}}{\partial\theta} &= 3h_2^2\frac{\partial h_2}{\partial\theta} - 3\frac{\partial h_2}{\partial\theta}\lambda_{2,2} - 3h_2\frac{\partial\lambda_{2,2}}{\partial\theta} \\
&= 3\bar{h}_2h_2^2 - 3\bar{z}_2h_2^2 - 3\bar{h}_2\lambda_{2,2} + 3\bar{z}_2\lambda_{2,2} - 3h_2\frac{\partial\lambda_{2,2}}{\partial\theta}
\end{aligned}$$

$$E(\frac{\partial h_{222}}{\partial\theta}) = \quad 3E(\bar{h}_2 h_2^2) - 3\bar{z}_2 E(h_2^2) + 3\bar{z}_2\lambda_{2,2}$$

$$\frac{\partial h_{222}}{\partial\theta} - E(\frac{\partial h_{222}}{\partial\theta}) = \quad 3[\bar{h}_2 h_2^2 - E(\bar{h}_2 h_2^2) - \bar{z}_2(h_2^2 - E(h_2^2)) - \bar{h}_2\lambda_{2,2} - h_2\frac{\partial\lambda_{2,2}}{\partial\theta}]$$

$$\begin{aligned}
E(h_{1111}) = \quad & E(h_1^4 - 6h_1^2\lambda_{1,1} + 3\lambda_{1,1}^2) \\
= \quad & E(h_1^4) - 6\lambda_{1,1}E(h_1^2) + 3\lambda_{1,1}^2 \\
= \quad & E(h_1^4) - 3\lambda_{1,1}^2 \\
= \quad & E(\lambda_{1,1}^4(S^1 - K^1)^4 + 4\lambda_{1,1}^3\lambda_{1,2}(S^1 - K^1)^3(S^2 - K^2) \\
+ \quad & 6\lambda_{1,1}^2\lambda_{1,2}^2(S^1 - K^1)^2(S^2 - K^2)^2 \\
+ \quad & 4\lambda_{1,1}\lambda_{1,2}^3(S^1 - K^1)(S^2 - K^2)^3 + \lambda_{1,2}^4(S^2 - K^2)^4) - 3\lambda_{1,1}^2 \\
= \quad & \lambda_{1,1}^4[K^{1,1,1,1} + 3(K^{1,1})^2] \\
+ \quad & 4\lambda_{1,1}^3\lambda_{1,2}[K^{1,1,1,2} + 3K^{1,1}K^{1,2}] \\
+ \quad & 6\lambda_{1,1}^2\lambda_{1,2}^2[K^{1,1,2,2} + K^{1,1}K^{2,2} + 2(K^{1,2})^2] \\
+ \quad & 4\lambda_{1,1}\lambda_{1,2}^3[K^{1,2,2,2} + 3K^{1,2}K^{2,2}] \\
+ \quad & \lambda_{1,2}^4[K^{2,2,2,2} + 3(K^{2,2})^2] - 3\lambda_{1,1}^2
\end{aligned}$$

$$
\begin{aligned}
E(h_{1112}) = & \quad E(h_1^3 h_2 - 3h_1^2\lambda_{1,2} - 3h_1h_2\lambda_{1,1} + 3\lambda_{1,1}\lambda_{1,2}) \\
= & \quad E(h_1^3 h_2) - 3\lambda_{1,2}E(h_1^2) - 3\lambda_{1,1}E(h_1h_2) + 3\lambda_{1,1}\lambda_{1,2} \\
= & \quad E(h_1^3 h_2) - 3\lambda_{1,1}\lambda_{1,2} \\
= & \quad E(\lambda_{1,1}^3\lambda_{1,2}(S^1 - K^1)^4 + \lambda_{1,1}^2(3\lambda_{1,2}^2 + \lambda_{1,1}\lambda_{2,2})(S^1 - K^1)^3(S^2 - K^2) \\
+ & \quad 3\lambda_{1,1}\lambda_{1,2}(\lambda_{1,2}^2 + \lambda_{1,1}\lambda_{2,2})(S^1 - K^1)^2(S^2 - K^2)^2 \\
+ & \quad \lambda_{1,2}^2(\lambda_{1,2}^2 + 3\lambda_{1,1}\lambda_{2,2})(S^1 - K^1)(S^2 - K^2)^3 + \lambda_{1,2}^3\lambda_{2,2}(S^2 - K^2)^4) - 3\lambda_{1,1}\lambda_{1,2} \\
= & \quad \lambda_{1,1}^3\lambda_{1,2}[K^{1,1,1,1} + 3(K^{1,1})^2] \\
+ & \quad \lambda_{1,1}^2(3\lambda_{1,2}^2 + \lambda_{1,1}\lambda_{2,2})[K^{1,1,1,2} + 3K^{1,1}K^{1,2}] \\
+ & \quad 3\lambda_{1,1}\lambda_{1,2}(\lambda_{1,2}^2 + \lambda_{1,1}\lambda_{2,2})[K^{1,1,2,2} + K^{1,1}K^{2,2} + 2(K^{1,2})^2] \\
+ & \quad \lambda_{1,2}^2(\lambda_{1,2}^2 + 3\lambda_{1,1}\lambda_{2,2})[K^{1,2,2,2} + 3K^{1,2}K^{2,2}] \\
+ & \quad \lambda_{1,2}^3\lambda_{2,2}[K^{2,2,2,2} + 3(K^{2,2})^2] - 3\lambda_{1,1}\lambda_{1,2}
\end{aligned}
$$

$$
\begin{aligned}
E(h_{1122}) = \quad & E(h_1^2 h_2^2 - h_1^2 \lambda_{2,2} - 4h_1 h_2 \lambda_{1,2} - h_2^2 \lambda_{1,1} + \lambda_{1,1}\lambda_{2,2} + 2\lambda_{1,2}^2) \\
= \quad & E(h_1^2 h_2^2) - E(h_1^2)\lambda_{2,2} - 4E(h_1 h_2)\lambda_{1,2} - E(h_2^2)\lambda_{1,1} + \lambda_{1,1}\lambda_{2,2} + 2\lambda_{1,2}^2 \\
= \quad & E(h_1^2 h_2^2) - \lambda_{1,1}\lambda_{2,2} - 2\lambda_{1,2}^2 \\
= \quad & E(\lambda_{1,1}^2 \lambda_{1,2}^2 (S^1 - K^1)^4 + 2\lambda_{1,1}\lambda_{1,2}(\lambda_{1,1}\lambda_{2,2} + \lambda_{1,2}^2)(S^1 - K^1)^3 (S^2 - K^2) \\
+ \quad & (\lambda_{1,1}^2 \lambda_{2,2}^2 + 4\lambda_{1,1}\lambda_{1,2}^2 \lambda_{2,2} + \lambda_{1,2}^4)(S^1 - K^1)^2 (S^2 - K^2)^2 \\
+ \quad & 2\lambda_{1,2}\lambda_{2,2}(\lambda_{1,1}\lambda_{2,2} + \lambda_{1,2}^2)(S^1 - K^1)(S^2 - K^2)^3 \\
+ \quad & \lambda_{1,2}^2 \lambda_{2,2}^2 (S^2 - K^2)^4) - \lambda_{1,1}\lambda_{2,2} - 2\lambda_{1,2}^2 \\
= \quad & \lambda_{1,1}^2 \lambda_{1,2}^2 [K^{1,1,1,1} + 3(K^{1,1})^2] \\
+ \quad & 2\lambda_{1,1}\lambda_{1,2}(\lambda_{1,1}\lambda_{2,2} + \lambda_{1,2}^2)[K^{1,1,1,2} + 3K^{1,1}K^{1,2}] \\
+ \quad & (\lambda_{1,1}^2 \lambda_{2,2}^2 + 4\lambda_{1,1}\lambda_{1,2}^2 \lambda_{2,2} + \lambda_{1,2}^4)[K^{1,1,2,2} + K^{1,1}K^{2,2} + 2(K^{1,2})^2] \\
+ \quad & 2\lambda_{1,2}\lambda_{2,2}(\lambda_{1,1}\lambda_{2,2} + \lambda_{1,2}^2)[K^{1,2,2,2} + 3K^{1,2}K^{2,2}] \\
+ \quad & \lambda_{1,2}^2 \lambda_{2,2}^2 [K^{2,2,2,2} + 3(K^{2,2})^2] - \lambda_{1,1}\lambda_{2,2} - 2\lambda_{1,2}^2
\end{aligned}
$$

$$
\begin{aligned}
E(h_{1222}) = \quad & E(h_1 h_2^3 - 3h_1 h_2 \lambda_{2,2} - 3h_2^2 \lambda_{1,2} + 3\lambda_{1,2}\lambda_{2,2}) \\
= \quad & E(h_1 h_2^3) - 3E(h_1 h_2)\lambda_{2,2} - 3E(h_2^2)\lambda_{1,2} + 3\lambda_{1,2}\lambda_{2,2} \\
= \quad & E(h_1 h_2^3) - 3\lambda_{1,2}\lambda_{2,2} \\
= \quad & E(\lambda_{1,1}\lambda_{1,2}^3 (S^1 - K^1)^4 + \lambda_{1,2}^2 (3\lambda_{1,1}\lambda_{2,2} + \lambda_{1,2}^2)(S^1 - K^1)^3 (S^2 - K^2) \\
+ \quad & 3\lambda_{1,2}\lambda_{2,2}(\lambda_{1,1}\lambda_{2,2} + \lambda_{1,2}^2)(S^1 - K^1)^2 (S^2 - K^2)^2 \\
+ \quad & \lambda_{2,2}^2 (\lambda_{1,1}\lambda_{2,2} + 3\lambda_{1,2}^2)(S^1 - K^1)(S^2 - K^2)^3 \\
+ \quad & \lambda_{1,2}\lambda_{2,2}^3 (S^2 - K^2)^4) - 3\lambda_{1,2}\lambda_{2,2} \\
= \quad & \lambda_{1,1}\lambda_{1,2}^3 [K^{1,1,1,1} + 3(K^{1,1})^2] \\
+ \quad & \lambda_{1,2}^2 (3\lambda_{1,1}\lambda_{2,2} + \lambda_{1,2}^2)[K^{1,1,1,2} + 3K^{1,1}K^{1,2}] \\
+ \quad & 3\lambda_{1,2}\lambda_{2,2}(\lambda_{1,1}\lambda_{2,2} + \lambda_{1,2}^2)[K^{1,1,2,2} + K^{1,1}K^{2,2} + 2(K^{1,2})^2] \\
+ \quad & \lambda_{2,2}^2 (\lambda_{1,1}\lambda_{2,2} + 3\lambda_{1,2}^2)[K^{1,2,2,2} + 3K^{1,2}K^{2,2}] \\
+ \quad & \lambda_{1,2}\lambda_{2,2}^3 [K^{2,2,2,2} + 3(K^{2,2})^2] - 3\lambda_{1,2}\lambda_{2,2}
\end{aligned}
$$

$$
\begin{aligned}
E(h_{2222}) = \quad & E(h_2^4 - 6h_2^2 \lambda_{2,2} + 3\lambda_{2,2}^2) \\
= \quad & E(h_2^4) - 6E(h_2^2)\lambda_{2,2} + 3\lambda_{2,2}^2 \\
= \quad & E(h_2^4) - 3\lambda_{2,2}^2 \\
= \quad & E(\lambda_{1,2}^4 (S^1 - K^1)^4 + 4\lambda_{1,2}^3 \lambda_{2,2}(S^1 - K^1)^3 (S^2 - K^2) \\
+ \quad & 6\lambda_{1,2}^2 \lambda_{2,2}^2 (S^1 - K^1)^2 (S^2 - K^2)^2 \\
+ \quad & 4\lambda_{1,2}\lambda_{2,2}^3 (S^1 - K^1)(S^2 - K^2)^3 + \lambda_{2,2}^4 (S^2 - K^2)^4) - 3\lambda_{2,2}^2 \\
= \quad & \lambda_{1,2}^4 [K^{1,1,1,1} + 3(K^{1,1})^2] \\
+ \quad & 4\lambda_{1,2}^3 \lambda_{2,2}[K^{1,1,1,2} + 3K^{1,1}K^{1,2}] \\
+ \quad & 6\lambda_{1,2}^2 \lambda_{2,2}^2 [K^{1,1,2,2} + K^{1,1}K^{2,2} + 2(K^{1,2})^2] \\
+ \quad & 4\lambda_{1,2}\lambda_{2,2}^3 [K^{1,2,2,2} + 3K^{1,2}K^{2,2}] \\
+ \quad & \lambda_{2,2}^4 [K^{2,2,2,2} + 3(K^{2,2})^2] - 3\lambda_{2,2}^2
\end{aligned}
$$

$$\begin{aligned}
\frac{\partial h_{1111}}{\partial\theta} &= 4\frac{\partial h_1}{\partial\theta}h_1^3 - 12\frac{\partial h_1}{\partial\theta}h_1\lambda_{1,1} - 6h_1^2\frac{\partial\lambda_{1,1}}{\partial\theta} + 6\lambda_{1,1}\frac{\partial\lambda_{1,1}}{\partial\theta}\\
&= 4\bar{h}_1h_1^3 - 4\bar{z}_1h_1^3 - 12\bar{h}_1h_1\lambda_{1,1} + 12\bar{z}_1h_1\lambda_{1,1} - 6h_1^2\frac{\partial\lambda_{1,1}}{\partial\theta} + 6\lambda_{1,1}\frac{\partial\lambda_{1,1}}{\partial\theta}
\end{aligned}$$

$$E(\frac{\partial h_{1111}}{\partial\theta}) = 4E(\bar{h}_1h_1^3) - 4\bar{z}_1E(h_1^3) - 12E(\bar{h}_1h_1)\lambda_{1,1} - 6E(h_1^2)\frac{\partial\lambda_{1,1}}{\partial\theta} + 6\lambda_{1,1}\frac{\partial\lambda_{1,1}}{\partial\theta}$$

$$\begin{aligned}
\frac{\partial h_{1111}}{\partial\theta} - E(\frac{\partial h_{1111}}{\partial\theta}) = & \; 4[\bar{h}_1h_1^3 - E(\bar{h}_1h_1^3)] - 4\bar{z}_1[h_1^3 - E(h_1^3)]\\
- & \; 12[\bar{h}_1h_1 - E(\bar{h}_1h_1)]\lambda_{1,1} + 12\bar{z}_1h_1\lambda_{1,1}\\
- & \; 6[h_1^2 - E(h_1^2)]\frac{\partial\lambda_{1,1}}{\partial\theta}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial h_{1112}}{\partial\theta} = & \; 3\frac{\partial h_1}{\partial\theta}h_1^2h_2 + \frac{\partial h_2}{\partial\theta}h_1^3 - 6h_1\frac{\partial h_1}{\partial\theta}\lambda_{1,2} - 3h_1^2\frac{\partial\lambda_{1,2}}{\partial\theta} - 3\frac{\partial h_1}{\partial\theta}h_2\lambda_{1,1}\\
- & \; 3h_1\frac{\partial h_2}{\partial\theta}\lambda_{1,1} - 3h_1h_2\frac{\partial\lambda_{1,1}}{\partial\theta} + 3\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2} + 3\lambda_{1,1}\frac{\partial\lambda_{1,2}}{\partial\theta}\\
= & \; 3\bar{h}_1h_1^2h_2 - 3\bar{z}_1h_1^2h_2 + \bar{h}_2h_1^3 - \bar{z}_2h_1^3 - 6\bar{h}_1h_1\lambda_{1,2} + 6\bar{z}_1h_1\lambda_{1,2}\\
- & \; 3h_1^2\frac{\partial\lambda_{1,2}}{\partial\theta} - 3\bar{h}_1h_2\lambda_{1,1} + 3\bar{z}_1h_2\lambda_{1,1} - 3\bar{h}_2h_1\lambda_{1,1} + 3\bar{z}_2h_1\lambda_{1,1}\\
- & \; 3h_1h_2\frac{\partial\lambda_{1,1}}{\partial\theta} + 3\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2} + 3\lambda_{1,1}\frac{\partial\lambda_{1,2}}{\partial\theta}
\end{aligned}$$

$$\begin{aligned}
E(\frac{\partial h_{1112}}{\partial\theta}) = & \; 3E(\bar{h}_1h_1^2h_2) - 3\bar{z}_1E(h_1^2h_2) + E(\bar{h}_2h_1^3) - \bar{z}_2E(h_1^3) - 6E(\bar{h}_1h_1)\lambda_{1,2}\\
- & \; 3E(h_1^2)\frac{\partial\lambda_{1,2}}{\partial\theta} - 3E(\bar{h}_1h_2)\lambda_{1,1} - 3E(\bar{h}_2h_1)\lambda_{1,1} - 3E(h_1h_2)\frac{\partial\lambda_{1,1}}{\partial\theta}\\
+ & \; 3\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2} + 3\lambda_{1,1}\frac{\partial\lambda_{1,2}}{\partial\theta}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial h_{1112}}{\partial\theta} - E(\frac{\partial h_{1112}}{\partial\theta}) = & \; 3[\bar{h}_1h_1^2h_2 - E(\bar{h}_1h_1^2h_2)] - 3\bar{z}_1[h_1^2h_2 - E(h_1^2h_2)] + [\bar{h}_2h_1^3\\
- & \; E(\bar{h}_2h_1^3)] - \bar{z}_2[h_1^3 - E(h_1^3)] - 6[\bar{h}_1h_1 - E(\bar{h}_1h_1)]\lambda_{1,2}\\
+ & \; 6\bar{z}_1h_1\lambda_{1,2} - 3[h_1^2 - E(h_1^2)]\frac{\partial\lambda_{1,2}}{\partial\theta} - 3[\bar{h}_1h_2 - E(\bar{h}_1h_2)]\lambda_{1,1}\\
+ & \; 3\bar{z}_1h_2\lambda_{1,1} - 3[\bar{h}_2h_1 - E(\bar{h}_2h_1)]\lambda_{1,1}\\
+ & \; 3\bar{z}_2h_1\lambda_{1,1} - 3[h_1h_2 - E(h_1h_2)]\frac{\partial\lambda_{1,1}}{\partial\theta}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial h_{1122}}{\partial\theta} = & \; 2\frac{\partial h_1}{\partial\theta}h_1h_2^2 + 2\frac{\partial h_2}{\partial\theta}h_1^2h_2 - 2\frac{\partial h_1}{\partial\theta}h_1\lambda_{2,2} - h_1^2\frac{\partial\lambda_{2,2}}{\partial\theta} - 4\frac{\partial h_1}{\partial\theta}h_2\lambda_{1,2} \\
- & \; 4h_1\frac{\partial h_2}{\partial\theta}\lambda_{1,2} - 4h_1h_2\frac{\partial\lambda_{1,2}}{\partial\theta} - 2h_2\frac{\partial h_2}{\partial\theta}\lambda_{1,1} - h_2^2\frac{\partial\lambda_{1,1}}{\partial\theta} + \frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{2,2} \\
+ & \; \lambda_{1,1}\frac{\partial\lambda_{2,2}}{\partial\theta} + 4\lambda_{1,2}\frac{\partial\lambda_{1,2}}{\partial\theta} \\
= & \; 2\bar{h}_1h_1h_2^2 - 2\bar{z}_1h_1h_2^2 + 2\bar{h}_2h_1^2h_2 - 2\bar{z}_2h_1^2h_2 - 2\bar{h}_1h_1\lambda_{2,2} + 2\bar{z}_1h_1\lambda_{2,2} \\
- & \; h_1^2\frac{\partial\lambda_{2,2}}{\partial\theta} - 4\bar{h}_1h_2\lambda_{1,2} + 4\bar{z}_1h_2\lambda_{1,2} - 4\bar{h}_2h_1\lambda_{1,2} + 4\bar{z}_2h_1\lambda_{1,2} - 4h_1h_2\frac{\partial\lambda_{1,2}}{\partial\theta} \\
- & \; 2\bar{h}_2h_2\lambda_{1,1} + 2\bar{z}_2h_2\lambda_{1,1} - h_2^2\frac{\partial\lambda_{1,1}}{\partial\theta} + \frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{2,2} + \lambda_{1,1}\frac{\partial\lambda_{2,2}}{\partial\theta} + 4\lambda_{1,2}\frac{\partial\lambda_{1,2}}{\partial\theta}
\end{aligned}$$

$$\begin{aligned}
E(\frac{\partial h_{1122}}{\partial\theta}) = & \; 2E(\bar{h}_1h_1h_2^2) - 2\bar{z}_1E(h_1h_2^2) + 2E(\bar{h}_2h_1^2h_2) - 2\bar{z}_2E(h_1^2h_2) - 2E(\bar{h}_1h_1)\lambda_{2,2} \\
- & \; E(h_1^2)\frac{\partial\lambda_{2,2}}{\partial\theta} - 4E(\bar{h}_1h_2)\lambda_{1,2} - 4E(\bar{h}_2h_1)\lambda_{1,2} - 4E(h_1h_2)\frac{\partial\lambda_{1,2}}{\partial\theta} \\
- & \; 2E(\bar{h}_2h_2)\lambda_{1,1} - E(h_2^2)\frac{\partial\lambda_{1,1}}{\partial\theta} + \frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{2,2} + \lambda_{1,1}\frac{\partial\lambda_{2,2}}{\partial\theta} + 4\lambda_{1,2}\frac{\partial\lambda_{1,2}}{\partial\theta}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial h_{1122}}{\partial\theta} - E(\frac{\partial h_{1122}}{\partial\theta}) = & \; 2[\bar{h}_1h_1h_2^2 - E(\bar{h}_1h_1h_2^2)] - 2\bar{z}_1[h_1h_2^2 - E(h_1h_2^2)] \\
+ & \; 2[\bar{h}_2h_1^2h_2 - E(\bar{h}_2h_1^2h_2)] - 2\bar{z}_2[h_1^2h_2 - E(h_1^2h_2)] \\
- & \; 2[\bar{h}_1h_1 - E(\bar{h}_1h_1)]\lambda_{2,2} + 2\bar{z}_1h_1\lambda_{2,2} - [h_1^2 - E(h_1^2)]\frac{\partial\lambda_{2,2}}{\partial\theta} \\
- & \; 4[\bar{h}_1h_2 - E(\bar{h}_1h_2)]\lambda_{1,2} + 4\bar{z}_1h_2\lambda_{1,2} - 4[\bar{h}_2h_1 - E(\bar{h}_2h_1)]\lambda_{1,2} \\
+ & \; 4\bar{z}_2h_1\lambda_{1,2} - 4[h_1h_2 - E(h_1h_2)]\frac{\partial\lambda_{1,2}}{\partial\theta} - 2[\bar{h}_2h_2 - E(\bar{h}_2h_2)]\lambda_{1,1} \\
+ & \; 2\bar{z}_2h_2\lambda_{1,1} - [h_2^2 - E(h_2^2)]\frac{\partial\lambda_{1,1}}{\partial\theta}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial h_{1222}}{\partial\theta} = & \; \frac{\partial h_1}{\partial\theta}h_2^3 + 3\frac{\partial h_2}{\partial\theta}h_1h_2^2 - 3\frac{\partial h_1}{\partial\theta}h_2\lambda_{2,2} - 3\frac{\partial h_2}{\partial\theta}h_1\lambda_{2,2} - 3h_1h_2\frac{\partial\lambda_{2,2}}{\partial\theta} \\
- & \; 6\frac{\partial h_2}{\partial\theta}h_2\lambda_{1,2} - 3h_2^2\frac{\partial\lambda_{1,2}}{\partial\theta} + 3\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{2,2} + 3\lambda_{1,2}\frac{\partial\lambda_{2,2}}{\partial\theta} \\
= & \; \bar{h}_1h_2^3 - \bar{z}_1h_2^3 + 3\bar{h}_2h_1h_2^2 - 3\bar{z}_2h_1h_2^2 - 3\bar{h}_1h_2\lambda_{2,2} + 3\bar{z}_1h_2\lambda_{2,2} \\
- & \; 3\bar{h}_2h_1\lambda_{2,2} + 3\bar{z}_2h_1\lambda_{2,2} - 3h_1h_2\frac{\partial\lambda_{2,2}}{\partial\theta} - 6\bar{h}_2h_2\lambda_{1,2} + 6\bar{z}_2h_2\lambda_{1,2} \\
- & \; 3h_2^2\frac{\partial\lambda_{1,2}}{\partial\theta} + 3\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{2,2} + 3\lambda_{1,2}\frac{\partial\lambda_{2,2}}{\partial\theta}
\end{aligned}$$

$$E(\frac{\partial h_{1222}}{\partial \theta}) = E(\bar{h}_1 h_2^3) - \bar{z}_1 E(h_2^3) + 3E(\bar{h}_2 h_1 h_2^2) - 3\bar{z}_2 E(h_1 h_2^2) - 3E(\bar{h}_1 h_2)\lambda_{2,2}$$
$$- 3E(\bar{h}_2 h_1)\lambda_{2,2} - 3E(h_1 h_2)\frac{\partial \lambda_{2,2}}{\partial \theta} - 6E(\bar{h}_2 h_2)\lambda_{1,2}$$
$$- 3E(h_2^2)\frac{\partial \lambda_{1,2}}{\partial \theta} + 3\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{2,2} + 3\lambda_{1,2}\frac{\partial \lambda_{2,2}}{\partial \theta}$$

$$\frac{\partial h_{1222}}{\partial \theta} - E(\frac{\partial h_{1222}}{\partial \theta}) = [\bar{h}_1 h_2^3 - E(\bar{h}_1 h_2^3)] - \bar{z}_1[h_2^3 - E(h_2^3)] + 3[\bar{h}_2 h_1 h_2^2 - E(\bar{h}_2 h_1 h_2^2)]$$
$$- 3\bar{z}_2[h_1 h_2^2 - E(h_1 h_2^2)] - 3[\bar{h}_1 h_2 - E(\bar{h}_1 h_2)]\lambda_{2,2} + 3\bar{z}_1 h_2 \lambda_{2,2}$$
$$- 3[\bar{h}_2 h_1 - E(\bar{h}_2 h_1)]\lambda_{2,2} + 3\bar{z}_2 h_1 \lambda_{2,2} - 3[h_1 h_2 - E(h_1 h_2)]\frac{\partial \lambda_{2,2}}{\partial \theta}$$
$$- 6[\bar{h}_2 h_2 - E(\bar{h}_2 h_2)]\lambda_{1,2} + 6\bar{z}_2 h_2 \lambda_{1,2} - 3[h_2^2 - E(h_2^2)]\frac{\partial \lambda_{1,2}}{\partial \theta}$$

$$\frac{\partial h_{2222}}{\partial \theta} = 4\frac{\partial h_2}{\partial \theta}h_2^3 - 12\frac{\partial h_2}{\partial \theta}h_2\lambda_{2,2} - 6h_2^2\frac{\partial \lambda_{2,2}}{\partial \theta} + 6\lambda_{2,2}\frac{\partial \lambda_{2,2}}{\partial \theta}$$
$$= 4\bar{h}_2 h_2^3 - 4\bar{z}_2 h_2^3 - 12\bar{h}_2 h_2 \lambda_{2,2} + 12\bar{z}_2 h_2 \lambda_{2,2} - 6h_2^2\frac{\partial \lambda_{2,2}}{\partial \theta} + 6\lambda_{2,2}\frac{\partial \lambda_{2,2}}{\partial \theta}$$

$$E(\frac{\partial h_{2222}}{\partial \theta}) = 4E(\bar{h}_2 h_2^3) - 4\bar{z}_2 E(h_2^3) - 12E(\bar{h}_2 h_2)\lambda_{2,2} - 6E(h_2^2)\frac{\partial \lambda_{2,2}}{\partial \theta} + 6\lambda_{2,2}\frac{\partial \lambda_{2,2}}{\partial \theta}$$

$$\frac{\partial h_{2222}}{\partial \theta} - E(\frac{\partial h_{2222}}{\partial \theta}) = 4[\bar{h}_2 h_2^3 - E(\bar{h}_2 h_2^3)] - 4\bar{z}_2[h_2^3 - E(h_2^3)]$$
$$- 12[\bar{h}_2 h_2 - E(\bar{h}_2 h_2)]\lambda_{2,2} + 12\bar{z}_2 h_2 \lambda_{2,2}$$
$$- 6[h_2^2 - E(h_2^2)]\frac{\partial \lambda_{2,2}}{\partial \theta}$$

$$
\begin{aligned}
E(\bar{h}_1 h_1^3) = \quad & E[\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}^3(S^1-K^1)^4 + (3\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}^2\lambda_{1,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}^3)(S^1-K^1)^3(S^2-K^2) \\
+ \quad & 3(\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}\lambda_{1,2}^2 + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}^2\lambda_{1,2})(S^1-K^1)^2(S^2-K^2)^2 \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}^3 + 3\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}\lambda_{1,2}^2)(S^1-K^1)(S^2-K^2)^3 + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^3(S^2-K^2)^4] \\
= \quad & \frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}^3[K^{1,1,1,1} + 3(K^{1,1})^2] \\
+ \quad & (3\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}^2\lambda_{1,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}^3)[K^{1,1,1,2} + 3K^{1,1}K^{1,2}] \\
+ \quad & 3(\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}\lambda_{1,2}^2 + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}^2\lambda_{1,2})[K^{1,1,2,2} + K^{1,1}K^{2,2} + 2(K^{1,2})^2] \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}^3 + 3\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}\lambda_{1,2}^2)[K^{1,2,2,2} + 3K^{1,2}K^{2,2}] \\
+ \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^3[K^{2,2,2,2} + 3(K^{2,2})^2]
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_1 h_1^2 h_2) = \quad & E[\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}^2\lambda_{1,2}(S^1-K^1)^4 \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}(2\lambda_{1,1}\lambda_{1,2}^2 + \lambda_{1,1}^2\lambda_{2,2}) + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}^2\lambda_{1,2})(S^1-K^1)^3(S^2-K^2) \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}(\lambda_{1,2}^3 + 2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}) + \frac{\partial\lambda_{1,2}}{\partial\theta}(2\lambda_{1,1}\lambda_{1,2}^2 + \lambda_{1,1}^2\lambda_{2,2}))(S^1-K^1)^2(S^2-K^2)^2 \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}^2\lambda_{2,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}(\lambda_{1,2}^3 + 2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}))(S^1-K^1)(S^2-K^2)^3 \\
+ \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^2\lambda_{2,2}(S^2-K^2)^4] \\
= \quad & \frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}^2\lambda_{1,2}[K^{1,1,1,1} + 3(K^{1,1})^2] \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}(2\lambda_{1,1}\lambda_{1,2}^2 + \lambda_{1,1}^2\lambda_{2,2}) + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}^2\lambda_{1,2})[K^{1,1,1,2} + 3K^{1,1}K^{1,2}] \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}(\lambda_{1,2}^3 + 2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}) + \frac{\partial\lambda_{1,2}}{\partial\theta}(2\lambda_{1,1}\lambda_{1,2}^2 + \lambda_{1,1}^2\lambda_{2,2}))[K^{1,1,2,2} \\
+ \quad & K^{1,1}K^{2,2} + 2(K^{1,2})^2] \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}^2\lambda_{2,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}(\lambda_{1,2}^3 + 2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}))[K^{1,2,2,2} + 3K^{1,2}K^{2,2}] \\
+ \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^2\lambda_{2,2}[K^{2,2,2,2} + 3(K^{2,2})^2]
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_1 h_1 h_2^2) = \quad & E[\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}\lambda_{1,2}^2(S^1-K^1)^4 \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}(\lambda_{1,2}^3+2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2})+\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}\lambda_{1,2}^2)(S^1-K^1)^3(S^2-K^2) \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}(2\lambda_{1,2}^2\lambda_{2,2}+\lambda_{1,1}\lambda_{2,2}^2)+\frac{\partial\lambda_{1,2}}{\partial\theta}(\lambda_{1,2}^3+2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}))(S^1-K^1)^2(S^1-K^1)^2 \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}\lambda_{2,2}^2+\frac{\partial\lambda_{1,2}}{\partial\theta}(2\lambda_{1,2}^2\lambda_{2,2}+\lambda_{1,1}\lambda_{2,2}^2))(S^1-K^1)(S^2-K^2)^3 \\
+ \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}\lambda_{2,2}^2(S^2-K^2)^4] \\
= \quad & \frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,1}\lambda_{1,2}^2[K^{1,1,1,1}+3(K^{1,1})^2] \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}(\lambda_{1,2}^3+2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2})+\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}\lambda_{1,2}^2)[K^{1,1,1,2}+3K^{1,1}K^{1,2}] \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}(2\lambda_{1,2}^2\lambda_{2,2}+\lambda_{1,1}\lambda_{2,2}^2)+\frac{\partial\lambda_{1,2}}{\partial\theta}(\lambda_{1,2}^3+2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}))[K^{1,1,2,2} \\
+ \quad & K^{1,1}K^{2,2}+2(K^{1,2})^2] \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}\lambda_{2,2}^2+\frac{\partial\lambda_{1,2}}{\partial\theta}(2\lambda_{1,2}^2\lambda_{2,2}+\lambda_{1,1}\lambda_{2,2}^2))[K^{1,2,2,2}+3K^{1,2}K^{2,2}] \\
+ \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}\lambda_{2,2}^2[K^{2,2,2,2}+3(K^{2,2})^2]
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_1 h_2^3) = \quad & E[\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}^3(S^1-K^1)^4 \\
+ \quad & (3\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}^2\lambda_{2,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^3)(S^1-K^1)^3(S^2-K^2) \\
+ \quad & 3(\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}\lambda_{2,2}^2 + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^2\lambda_{2,2})(S^1-K^1)^2(S^2-K^2)^2 \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{2,2}^3 + 3\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}\lambda_{2,2}^2)(S^1-K^1)(S^2-K^2)^3 \\
+ \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{2,2}^3(S^2-K^2)^4] \\
= \quad & \frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}^3[K^{1,1,1,1} + 3(K^{1,1})^2] \\
+ \quad & (3\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}^2\lambda_{2,2} + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^3)[K^{1,1,1,2} + 3K^{1,1}K^{1,2}] \\
+ \quad & 3(\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{1,2}\lambda_{2,2}^2 + \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^2\lambda_{2,2})[K^{1,1,2,2} + K^{1,1}K^{2,2} + 2(K^{1,2})^2] \\
+ \quad & (\frac{\partial\lambda_{1,1}}{\partial\theta}\lambda_{2,2}^3 + 3\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}\lambda_{2,2}^2)[K^{1,2,2,2} + 3K^{1,2}K^{2,2}] \\
+ \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{2,2}^3[K^{2,2,2,2} + 3(K^{2,2})^2]
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_2 h_1^3) = \quad & E[\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}^3(S^1-K^1)^4 + (3\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}^2\lambda_{1,2} + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,1}^3)(S^1-K^1)^3(S^2-K^2) \\
+ \quad & 3(\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}\lambda_{1,2}^2 + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,1}^2\lambda_{1,2})(S^1-K^1)^2(S^2-K^2)^2 \\
+ \quad & (\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^3 + 3\frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,1}\lambda_{1,2}^2)(S^1-K^1)(S^2-K^2)^3 + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,2}^3(S^2-K^2)^4] \\
= \quad & \frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}^3[K^{1,1,1,1} + 3(K^{1,1})^2] \\
+ \quad & (3\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}^2\lambda_{1,2} + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,1}^3)[K^{1,1,1,2} + 3K^{1,1}K^{1,2}] \\
+ \quad & 3(\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,1}\lambda_{1,2}^2 + \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,1}^2\lambda_{1,2})[K^{1,1,2,2} + K^{1,1}K^{2,2} + 2(K^{1,2})^2] \\
+ \quad & (\frac{\partial\lambda_{1,2}}{\partial\theta}\lambda_{1,2}^3 + 3\frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,1}\lambda_{1,2}^2)[K^{1,2,2,2} + 3K^{1,2}K^{2,2}] \\
+ \quad & \frac{\partial\lambda_{2,2}}{\partial\theta}\lambda_{1,2}^3[K^{2,2,2,2} + 3(K^{2,2})^2]
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_2 h_1^2 h_2) = \quad & E[\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,1}^2\lambda_{1,2}(S^1 - K^1)^4 \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}(2\lambda_{1,1}\lambda_{1,2}^2 + \lambda_{1,1}^2\lambda_{2,2}) + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,1}^2\lambda_{1,2})(S^1 - K^1)^3(S^2 - K^2) \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}(\lambda_{1,2}^3 + 2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}) + \frac{\partial \lambda_{2,2}}{\partial \theta}(2\lambda_{1,1}\lambda_{1,2}^2 + \lambda_{1,1}^2\lambda_{2,2}))(S^1 - K^1)^2(S^2 - K^2)^2 \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}^2\lambda_{2,2} + \frac{\partial \lambda_{2,2}}{\partial \theta}(\lambda_{1,2}^3 + 2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}))(S^1 - K^1)(S^2 - K^2)^3 \\
+ \quad & \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}^2\lambda_{2,2}(S^2 - K^2)^4] \\
= \quad & \frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,1}^2\lambda_{1,2}[K^{1,1,1,1} + 3(K^{1,1})^2] \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}(2\lambda_{1,1}\lambda_{1,2}^2 + \lambda_{1,1}^2\lambda_{2,2}) + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,1}^2\lambda_{1,2})[K^{1,1,1,2} + 3K^{1,1}K^{1,2}] \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}(\lambda_{1,2}^3 + 2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}) + \frac{\partial \lambda_{2,2}}{\partial \theta}(2\lambda_{1,1}\lambda_{1,2}^2 + \lambda_{1,1}^2\lambda_{2,2}))[K^{1,1,2,2} \\
+ \quad & K^{1,1}K^{2,2} + 2(K^{1,2})^2] \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}^2\lambda_{2,2} + \frac{\partial \lambda_{2,2}}{\partial \theta}(\lambda_{1,2}^3 + 2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}))[K^{1,2,2,2} + 3K^{1,2}K^{2,2}] \\
+ \quad & \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}^2\lambda_{2,2}[K^{2,2,2,2} + 3(K^{2,2})^2]
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_2 h_1 h_2^2) = \quad & E[\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,1}\lambda_{1,2}^2(S^1 - K^1)^4 \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}(\lambda_{1,2}^3 + 2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}) + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,1}\lambda_{1,2}^2)(S^1 - K^1)^3(S^2 - K^2) \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}(2\lambda_{1,2}^2\lambda_{2,2} + \lambda_{1,1}\lambda_{2,2}^2) + \frac{\partial \lambda_{2,2}}{\partial \theta}(\lambda_{1,2}^3 + 2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}))(S^1 - K^1)^2(S^2 - K^2)^2 \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2}^2 + \frac{\partial \lambda_{2,2}}{\partial \theta}(2\lambda_{1,2}^2\lambda_{2,2} + \lambda_{1,1}\lambda_{2,2}^2))(S^1 - K^1)(S^2 - K^2)^3 \\
+ \quad & \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2}^2(S^2 - K^2)^4] \\
= \quad & \frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,1}\lambda_{1,2}^2[K^{1,1,1,1} + 3(K^{1,1})^2] \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}(\lambda_{1,2}^3 + 2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}) + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,1}\lambda_{1,2}^2)[K^{1,1,1,2} + 3K^{1,1}K^{1,2}] \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}(2\lambda_{1,2}^2\lambda_{2,2} + \lambda_{1,1}\lambda_{2,2}^2) + \frac{\partial \lambda_{2,2}}{\partial \theta}(\lambda_{1,2}^3 + 2\lambda_{1,1}\lambda_{1,2}\lambda_{2,2}))[K^{1,1,2,2} \\
+ \quad & K^{1,1}K^{2,2} + 2(K^{1,2})^2] \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2}^2 + \frac{\partial \lambda_{2,2}}{\partial \theta}(2\lambda_{1,2}^2\lambda_{2,2} + \lambda_{1,1}\lambda_{2,2}^2))[K^{1,2,2,2} + 3K^{1,2}K^{2,2}] \\
+ \quad & \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2}^2[K^{2,2,2,2} + 3(K^{2,2})^2]
\end{aligned}
$$

$$
\begin{aligned}
E(\bar{h}_2 h_2^3) = \quad & E[\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}^3 (S^1 - K^1)^4 \\
+ \quad & (3\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}^2\lambda_{2,2} + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}^3)(S^1 - K^1)^3 (S^2 - K^2) \\
+ \quad & 3(\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2}^2 + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}^2\lambda_{2,2})(S^1 - K^1)^2 (S^2 - K^2)^2 \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{2,2}^3 + 3\frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2}^2)(S^1 - K^1)(S^2 - K^2)^3 \\
+ \quad & \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{2,2}^3 (S^2 - K^2)^4] \\
= \quad & \frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}^3 [K^{1,1,1,1} + 3(K^{1,1})^2] \\
+ \quad & (3\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}^2\lambda_{2,2} + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}^3)[K^{1,1,1,2} + 3K^{1,1}K^{1,2}] \\
+ \quad & 3(\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2}^2 + \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}^2\lambda_{2,2})[K^{1,1,2,2} + K^{1,1}K^{2,2} + 2(K^{1,2})^2] \\
+ \quad & (\frac{\partial \lambda_{1,2}}{\partial \theta}\lambda_{2,2}^3 + 3\frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{1,2}\lambda_{2,2}^2)[K^{1,2,2,2} + 3K^{1,2}K^{2,2}] \\
+ \quad & \frac{\partial \lambda_{2,2}}{\partial \theta}\lambda_{2,2}^3 [K^{2,2,2,2} + 3(K^{2,2})^2]
\end{aligned}
$$

Appendix 2: Unconditional Moments for the Asset Return and Volatility Processes

Let the conditional moment generating function (MGF) be:

$$M^c = E(e^{tS}|c)$$

And the unconditional MGF is:

$$M^u = E(e^{tS})$$

$$M^u = E(M^c) = E[E(e^{tS}|c)]$$

The logarithmic MGF is:

$$\varphi(S;t) = \ln(M^u) = \ln(E(M^c))$$

The unconditional first moment for the asset returns can be derived as:

$$\frac{\partial\varphi(S;t)}{\partial t_1} = (M^u)^{-1}\frac{\partial M^u}{\partial t_1} = (E(M^c))^{-1}\frac{\partial E(M^c)}{\partial t_1} = (E(M^c))^{-1}E(\frac{\partial M^c}{\partial t_1})$$

Hence the unconditional mean of returns is:

$$K_u^1 = \frac{\partial\varphi(S;t)}{\partial t_1}|_{t=0} = E(K^1) = K^1 = \mu\tau$$

where K^1 is the conditional mean of asset return, K_u^1 is the unconditional mean.

$$\begin{aligned}\frac{\partial^2\varphi(S;t)}{\partial t_1^2} &= -(M^u)^{-2}(\frac{\partial M^u}{\partial t_1})^2 + (M^u)^{-1}\frac{\partial^2 M^u}{\partial t_1^2} \\ &= -(E(M^c))^{-2}[E(\frac{\partial M^c}{\partial t_1})]^2 + (E(M^c))^{-1}E(\frac{\partial^2 M^c}{\partial t_1^2}) \\ &= -(E(M^c))^{-2}[E(E(S_1 e^{tS}|c))]^2 + (E(M^c))^{-1}E(E(S_1^2 e^{tS}|c))\end{aligned}$$

Hence the unconditional second cumulant of return process is:

$$
\begin{aligned}
K_u^{1,1} &= \frac{\partial^2\varphi(S;t)}{\partial t_1^2}|_{t=0} = -[E(E(S_1|c))]^2 + E[E(S_1^2|c)] \\
&= -(K^1)^2 + E[K^{11}] = -(K^1)^2 + E[K^{1,1} + (K^1)^2] \\
&= E[K^{1,1}] \\
&= E[\alpha\tau + \alpha\frac{1-e^{\beta\tau}}{\beta e^{\beta\tau}} + \frac{e^{\beta\tau}-1}{\beta e^{\beta\tau}}v_t] \\
&= \alpha\tau + \alpha\frac{1-e^{\beta\tau}}{\beta e^{\beta\tau}} + \frac{e^{\beta\tau}-1}{\beta e^{\beta\tau}}\alpha \\
&= \alpha\tau
\end{aligned}
$$

Note K^{11} is the conditional second moment hence $K^{11} = K^{1,1} + (K^1)^2$, where $K^{1,1}$ is the conditional second cumulant and K^1 is the conditional first cumulant. $E(v_t) = \alpha$.

$$
\begin{aligned}
\frac{\partial^3\varphi(S;t)}{\partial t_1^3} &= 2(M^u)^{-3}(\frac{\partial M^u}{\partial t_1})^3 - 3(M^u)^{-2}(\frac{\partial M^u}{\partial t_1})(\frac{\partial^2 M^u}{\partial t_1^2}) + (M^u)^{-1}\frac{\partial^3 M^u}{\partial t_1^3} \\
&= 2(E(M^c))^{-3}[E(\frac{\partial M^c}{\partial t_1})]^3 - 3(E(M^c))^{-2}(E(\frac{\partial M^c}{\partial t_1}))(E(\frac{\partial^2 M^c}{\partial t_1^2})) \\
&+(E(M^c))^{-1}E(\frac{\partial^3 M^c}{\partial t_1^3})
\end{aligned}
$$

Hence the unconditional third cumulant of returns is:

$$
\begin{aligned}
K_u^{1,1,1} &= \frac{\partial^3\varphi(S;t)}{\partial t_1^3}|_{t=0} = 2(K^1)^3 - 3(K^1)[E(K^{11})] + [E(K^{111})] \\
&= 2(K^1)^3 - 3(K^1)[E(K^{1,1} + K^1K^1)] + [E(K^{1,1,1} + 3K^1K^{1,1} + K^1K^1K^1)] \\
&= E[K^{1,1,1}] \\
&= E[\frac{3e^{-\beta\tau}}{\beta^2}(\alpha(2+\beta\tau - e^{\beta\tau}(2-\beta\tau)) - (1 - e^{\beta\tau} + \beta\tau)v_t)\rho\sigma] \\
&= \frac{3e^{-\beta\tau}}{\beta^2}(\alpha(2+\beta\tau - e^{\beta\tau}(2-\beta\tau)) - (1 - e^{\beta\tau} + \beta\tau)\alpha)\rho\sigma \\
&= \frac{3\rho\sigma(\beta\alpha\tau + \alpha e^{-\beta\tau} - \alpha)}{\beta^2}
\end{aligned}
$$

$$\begin{aligned}\frac{\partial^4\varphi(S;t)}{\partial t_1^4} &= -6(M^u)^{-4}(\frac{\partial M^u}{\partial t_1})^4 + 12(M^u)^{-3}(\frac{\partial M^u}{\partial t_1})^2(\frac{\partial^2 M^u}{\partial t_1^2}) \\ &-3(M^u)^{-2}(\frac{\partial^2 M^u}{\partial t_1^2})^2 - 4(M^u)^{-2}(\frac{\partial M^u}{\partial t_1})(\frac{\partial^3 M^u}{\partial t_1^3}) + (M^u)^{-1}\frac{\partial^4 M^u}{\partial t_1^4} \\ &= -6(E(M^c))^{-4}[E(\frac{\partial M^c}{\partial t_1})]^4 + 12(E(M^c))^{-3}[E(\frac{\partial M^c}{\partial t_1})]^2 E(\frac{\partial^2 M^u}{\partial t_1^2}) \\ &-3(E(M^c))^{-2}(E(\frac{\partial^2 M^c}{\partial t_1^2}))^2 - 4(E(M^c))^{-2}(E(\frac{\partial M^c}{\partial t_1}))E(\frac{\partial^3 M^c}{\partial t_1^3}) \\ &+(E(M^c))^{-1}E(\frac{\partial^4 M^c}{\partial t_1^4})\end{aligned}$$

Hence the unconditional fourth cumulant of returns is:

$$\begin{aligned}K_u^{1,1,1,1} &= \frac{\partial^4\varphi(S;t)}{\partial t_1^4}|_{t=0} \\ &= -6(K^1)^4 + 12(K^1)^2[E(K^{11})] - 3[E(K^{11})]^2 - 4(K^1)E(K^{111}) + [E(K^{1111})] \\ &= -6(K^1)^4 + 12(K^1)^2[E(K^{1,1} + K^1K^1)] - 3[E(K^{1,1} + K^1K^1)]^2 \\ &-4(K^1)E(K^{1,1,1} + 3K^1K^{1,1} + K^1K^1K^1) \\ &+[E(K^{1,1,1,1} + 4K^1K^{1,1,1} + 3K^{1,1}K^{1,1} + 6K^1K^1K^{1,1} + K^1K^1K^1K^1)] \\ &= E[K^{1,1,1,1}] + 3E[K^{1,1}K^{1,1}] - 3[E(K^{1,1})]^2 \\ &= \frac{3\sigma^2(-\alpha e^{-2\beta\tau} + 4\alpha e^{-\beta\tau} + 8\alpha\beta\tau\rho^2 e^{-\beta\tau} + 16\alpha\rho^2 e^{-\beta\tau})}{2\beta^3} \\ &+\frac{3\sigma^2(-16\alpha\rho^2 - 3\alpha + 8\alpha\beta\tau\rho^2 + 2\alpha\beta\tau)}{2\beta^3}\end{aligned}$$

Given the relationship between moments and cumulants, we are able to obtain the unconditional moments for the asset return process:

$$E(x) = K_u^1 = \mu\tau$$

$$Var(x) = K_u^{1,1} = \alpha\tau$$

$$E[(x - E(x))^3] = K_u^{1,1,1}$$

$$E[(x-E(x))^4] = K_u^{1,1,1,1} + 3(K_u^{1,1})^2$$

Similarly, the unconditional first moment for the volatilities can be derived as:

$$\frac{\partial\varphi(S;t)}{\partial t_2} = (M^u)^{-1}\frac{\partial M^u}{\partial t_2} = (E(M^c))^{-1}\frac{\partial E(M^c)}{\partial t_2} = (E(M^c))^{-1}E(\frac{\partial M^c}{\partial t_2})$$

Hence the unconditional mean of volatility is:

$$K_u^2 = \frac{\partial\varphi(S;t)}{\partial t_2}|_{t=0} = E(K^2) = E(\alpha(1-e^{-\beta\tau}) + v_t e^{-\beta\tau}) = \alpha$$

where K^2 is the conditional mean of volatilities, K_u^2 is the unconditional mean.

$$\begin{aligned}\frac{\partial^2\varphi(S;t)}{\partial t_2^2} &= -(M^u)^{-2}(\frac{\partial M^u}{\partial t_2})^2 + (M^u)^{-1}\frac{\partial^2 M^u}{\partial t_2^2} \\ &= -(E(M^c))^{-2}[E(\frac{\partial M^c}{\partial t_2})]^2 + (E(M^c))^{-1}E(\frac{\partial^2 M^c}{\partial t_2^2}) \\ &= -(E(M^c))^{-2}[E(E(S_2 e^{tS}|c))]^2 + (E(M^c))^{-1}E(E(S_2^2 e^{tS}|c))\end{aligned}$$

Hence the unconditional second cumulant is:

$$
\begin{aligned}
K_u^{2,2} &= \frac{\partial^2 \varphi(S;t)}{\partial t_2^2}|_{t=0} = -[E(E(S_2|c))]^2 + E[E(S_2^2|c)] \\
&= -\alpha^2 + E[K^{22}] = -\alpha^2 + E[K^{2,2} + (K^2)^2] \\
&= -\alpha^2 + E[\frac{\alpha\sigma^2}{2\beta}(1-e^{-\beta\tau})^2 + \frac{v_t\sigma^2}{\beta}e^{-\beta\tau}(1-e^{-\beta\tau})] \\
&+E[\alpha^2(1-e^{-\beta\tau})^2 + 2\alpha v_t e^{-\beta\tau}(1-e^{-\beta\tau}) + v_t^2 e^{-2\beta\tau}] \\
&= -\alpha^2 + \frac{\alpha\sigma^2}{2\beta}(1-e^{-\beta\tau})^2 + \frac{\alpha\sigma^2}{\beta}e^{-\beta\tau}(1-e^{-\beta\tau}) \\
&+\alpha^2(1-e^{-\beta\tau})^2 + 2\alpha^2 e^{-\beta\tau}(1-e^{-\beta\tau}) + e^{-2\beta\tau}E(v_t^2) \\
&= -\alpha^2 + \frac{\alpha\sigma^2}{2\beta}(1-e^{-\beta\tau})(1-e^{-\beta\tau}+2e^{-\beta\tau}) \\
&+\alpha^2(1-e^{-\beta\tau})(1-e^{-\beta\tau}+2e^{-\beta\tau}) + e^{-2\beta\tau}(E(v_t)^2 + var(v_t)) \\
&= -\alpha^2 + (1-e^{-\beta\tau})(1+e^{-\beta\tau})(\frac{\alpha\sigma^2}{2\beta} + \alpha^2) \\
&+\alpha^2 e^{-2\beta\tau} + var(v_t)e^{-2\beta\tau} \\
&= (1-e^{-2\beta\tau})\frac{\alpha\sigma^2}{2\beta} + var(v_t)e^{-2\beta\tau} \\
&= (1-e^{-2\beta\tau})\frac{\alpha\sigma^2}{2\beta} + \frac{\alpha\sigma^2}{2\beta}e^{-2\beta\tau} \\
&= \frac{\alpha\sigma^2}{2\beta}
\end{aligned}
$$

Given the relationship between moments and cumulants, the unconditional variance of volatilities is calculated as:

$$Var(v) = K_u^{2,2} = \frac{\alpha\sigma^2}{2\beta}$$

Following the procedure, the unconditional third moment of volatilities is obtained as:

$$E[(v - E(v))^3] = \frac{\alpha(2\alpha^2\beta^2 + 3\alpha\beta\sigma^2 + \sigma^4)}{2\beta^2}$$

and the unconditional fourth moment of volatilities is calculated as:

$$E[(v - E(v))^4] = \frac{\alpha(4\alpha^3\beta^3 + 12\alpha^2\beta^2\sigma^2 + 11\alpha\beta\sigma^4 + 3\sigma^6)}{4\beta^3}$$

Plugging the values $\alpha = 0.867, \beta = 0.269, \rho = -0.5, \sigma = 0.613, \mu = 0.059$ and $\tau = 1/252\ (1/12)$ into the formula, the unconditional four moments of both asset return and volatility processes can be calculated.

Appendix 3:

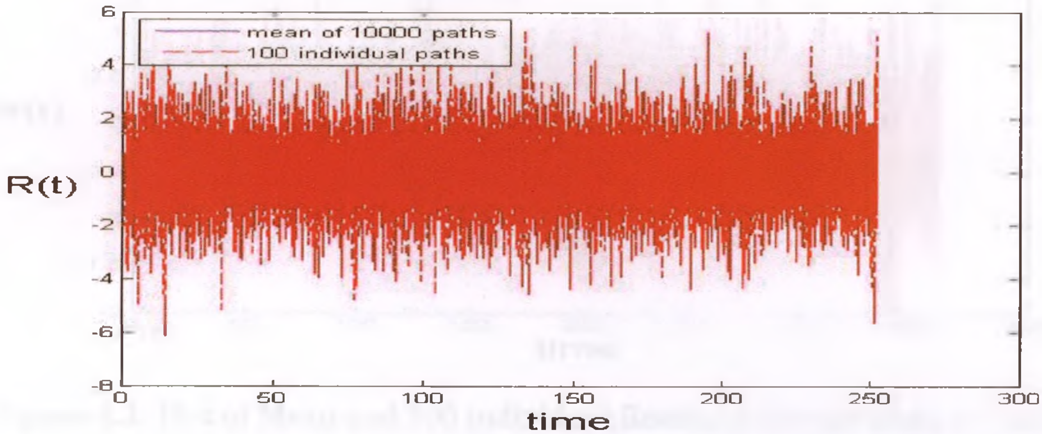

Figure 4.1: Plot of Mean and 100 individual Return processes when $\tau = 1/252$

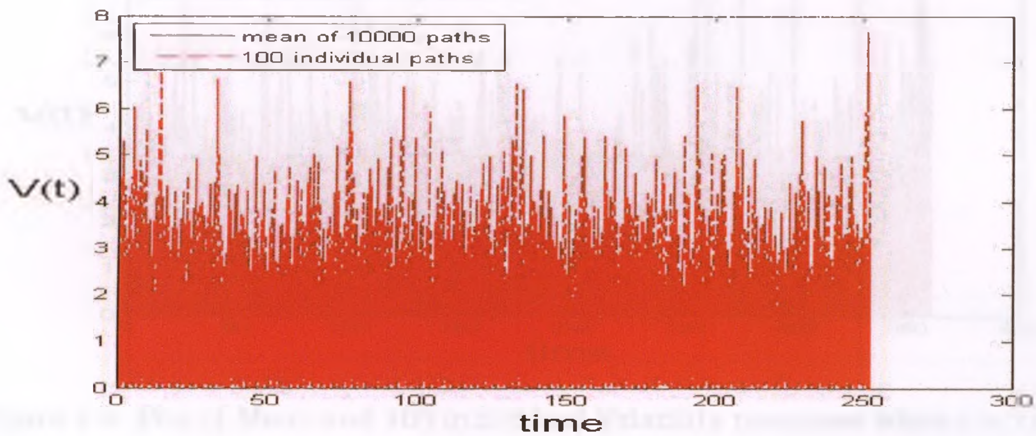

Figure 4.2: Plot of Mean and 100 individual Volatility processes when $\tau = 1/252$

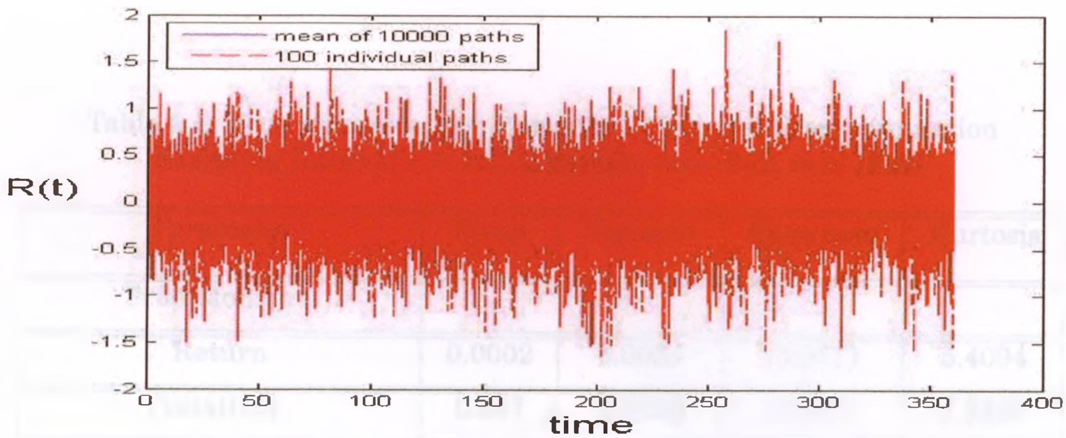

Figure 4.3: Plot of Mean and 100 individual Return processes when $\tau = 1/12$

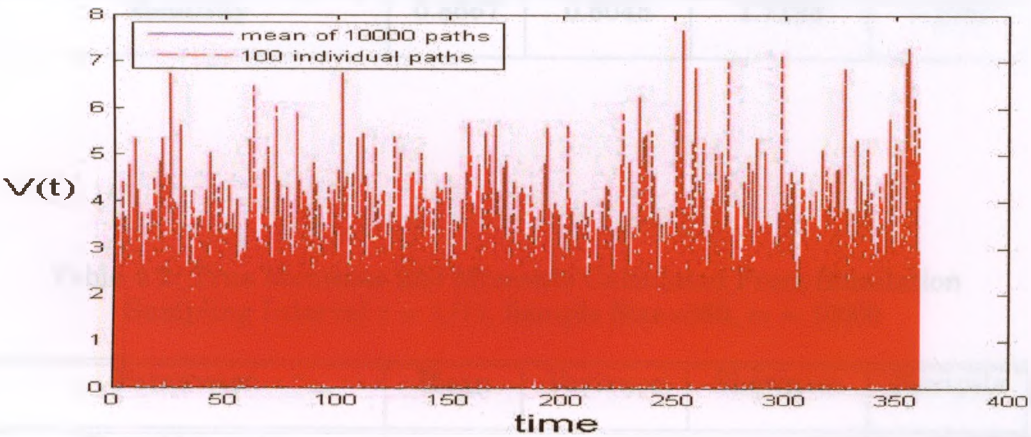

Figure 4.4: Plot of Mean and 100 individual Volatility processes when $\tau = 1/12$

Table 4.1: True Moments and Moments Calculated From Simulation
Sampling Interval $\tau = 1/252$, Sample Size=252, $m = 10000$

Parameter	Mean	Variance	Skewness	Kurtosis
True Moments				
Return	0.0002	0.0034	-0.0311	5.4094
Volatility	0.867	0.6056	1.7951	7.8336
Moments from Simulation				
Return	0.0002	0.0034	-0.0300	5.4140
Volatility	0.8667	0.6048	1.7193	7.1907

Table 4.2: True Moments and Moments Calculated From Simulation
Sampling Interval $\tau = 1/12$, Sample Size=360, $m = 10000$

Parameter	Mean	Variance	Skewness	Kurtosis
True Moments				
Return	0.0049	0.0722	-0.1415	3.0657
Volatility	0.867	0.6056	1.7951	7.8336
Moments from Simulation				
Return	0.0043	0.0737	-0.1106	3.6726
Volatility	0.8665	0.6051	1.7321	7.2612

Table 4.3: Monte Carlo Results with Observed Volatility
Sampling Interval $\tau = \frac{1}{252}$, Sample Size=252, $m = 10000$

Parameter	True Value	Estimate	Median	Std. Dev.	75 Percentile
α	0.867	0.8600	0.8570	0.0193	0.8735
β	0.269	0.2619	0.2590	0.0167	0.2781
ρ	-0.5	-0.4886	-0.4900	0.0116	-0.4700
σ	0.613	0.6148	0.6030	0.0283	0.6323
μ	0.059	-0.0588	0.0490	0.6382	0.0756

Table 4.4: Monte Carlo Results with Observed Volatility
Sampling Interval $\tau = \frac{1}{12}$, Sample Size=360, $m = 10000$

Parameter	True Value	Estimate	Median	Std. Dev.	75 Percentile
α	0.867	0.8872	0.9274	0.5462	1.1828
β	0.269	0.3296	0.3291	0.1996	0.4747
ρ	-0.5	-0.4772	-0.5038	0.0842	-0.4709
σ	0.613	0.6015	0.6244	0.1039	0.6568
μ	0.059	-0.0154	-0.0219	0.1512	0.0648

4.7 Bibliography

AIT-SAHALIA, Y. (2002): "Maximum Likelihood Estimation of Discretely Sampled Diffusion: A Closed-Form Approximation Approach," *Econometrica*, 70(1), 223–62.

BAILEY, W., AND R. M. STULZ (1989): "The Pricing of Stock Index Options in a General Equilibrium Model," *The Journal of Financial and Quantitative Analysis*, 24(1), 1–12.

BROADIE, M., AND O. KAYA (2006): "Exact Simulation of Stochastic Volatility and Other Affine Jump Diffusion Processes," *Operations Research*, 54(2), 217–231.

BROTO, C., AND E. RUIZ (2004): "Estimation Methods for Stochastic Volatility Models:A Survey," *Journal of Economic Surveys*, 18(5), 613–649.

CHACKO, G., AND L. M. VICEIRA (2003): "Spectral GMM Estimation of Continuous-Time Processes," *Journal of Econometrics*, 116, 259–92.

CHAN, K. C., G. A. KAROLYI, F. A. LONGSTAFF, AND A. B. SANDERS (1992): "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *The Journal of Finance*, 47(3), 1209–27.

COX, J. C., J. E. INGERSOLL, AND S. ROSS (1985): "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53, 363–384.

DUFFIE, D., J. PAN, AND K. SINGLETON (2000): "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68(6), 1343–76.

DUFFIE, D., AND K. J. SINGLETON (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–52.

ERAKER, B. (1998): "MCMC Analysis of Diffusion Models with Application to Finance," Working papers, Graduate School of Business, University of Chicago.

GALLANT, A. R., AND J. R. LONG (1997): "Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Square," *Biometrika*, 85, 125–41.

GALLANT, A. R., AND G. TAUCHEN (1996): "Which Moments to Match?," *Econometric Theory*, 12, 657–81.

HANSEN, L. P., AND J. SCHEINKMAN (1996): "Back to the Future: Generalized Moment Implications for Continuous Time Markov Process," *Econometrica*, 63, 767–804.

HESTON, S. L. (1993): "A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327–44.

HULL, J., AND A. WHITE (1987): "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance*, 42, 281–300.

JACQUIER, E., N. G. POLSON, AND P.ROSSI (1994): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371–417.

JIANG, G., AND J. KNIGHT (2002): "Estimation of Continuous-Time Processes via the Empirical Characteristic Function," *Journal of Business and Economics Statistics*, 21(2), 198–212.

JIANG, G., AND J. KNIGHT (2010): "ECF Estimation of Markov Models Where the Transition Density is Unknown," *Econometrics Journal*, 13, 245–70.

JOHNSON, H., AND D. SHANNO (1987): "Option Pricing When the Variance Is Changing," *Journal of Financial and Quantitative Analysis*, 22, 143–52.

KIM, S., N. SHEPHARD, AND S. CHIB (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *The Review of Economic Studies*, 65(3), 361–393.

MCCULLAGH, P. (1987): "Tensor Methods in Statistics," *Tensor Methods in Statistics*, Chapman and Hall/CRC; 1st edition.

SCOTT, L. (1987): "Option Pricing When the Variance Changes Randomly: Theory, Estimators, and Applications," *Journal of Financial and Quantitative Analysis*, 22, 419–38.

SINGLETON, K. J. (2001): "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics*, 102, 111–41.

VAN HAASTRECHT, A., AND A. PELSSER (2008): "Efficient, Almost Exact Simulation of the Heston Stochastic Volatility Model," Working paper.

WIGGINS, J. B. (1987): "Option Values Under Stochastic Volatility: Theory and Empirical Estimates," *Journal of Financial Economics*, 19, 351–72.

Chapter 5

Estimation of Continuous-Time Stochastic Volatility Model Using Realized and Implied Volatilities

5.1 Introduction

Continuous-time stochastic volatility (SV) model has figured prominently in the econometrics and finance literature. As it is able to reproduce the main features of financial time series, it has dominated the option pricing literature since the mid-1980s. However, in general, the estimation of such models is complicated since volatility is latent hence has to be integrated out from the model. The estimation is even more difficult because aside from a few special cases, the transition density functions do not have a closed form expression. For the affine diffusion and affine jump diffusion processes, although the transition density functions are unknown, the corresponding conditional characteristic functions (CCF) can be derived explicitly. Consequently, the empirical characteristic function (ECF) approach can be applied. Because of the one-to-one correspondence between the distribution function (DF) and the characteristic function

(CF), the ECF contains the same amount of information as the empirical DF, hence can achieve the asymptotic efficiency as the maximum likelihood (ML) function. As discussed in Jiang and Knight (2010), by choosing the optimal weight function appropriately and approximating it via an Edgeworth/Gram-Charlier expansion, the ECF method can also ensure the consistency, hence they named the procedure as the consistent approximate maximum likelihood (hereafter C-AMLE). For the affine continuous-time SV model, when volatility is treated as observed, the C-AMLE approach is convenient and the estimation process is simplified as it circumvents the need to integrate out the unobservable variables. In Chapter 4, we investigated the estimation performance of the C-AMLE procedure via a Monte Carlo study for the affine continuous-time SV model when the underlying volatility is observed. The implementation was straightforward. Using simulated asset returns and volatilities at both daily and monthly frequencies, the Monte Carlo evidence showed that the C-AMLE method did a good job at recovering the true parameters. In this chapter, we extend our study by examining the estimation of the affine continuous-time SV model via an empirical study. In reality, the underlying volatility is unobserved, but a volatility proxy can be employed.

Volatility measures have been widely employed in modern academic and financial market practitioner literatures over the last two decades. On one hand, with the availability of high frequency intra-day transactions, the realized volatility, motivated by the theory of quadratic variation of financial asset price process, is able to be computed. On the other hand, the forward looking market based implied volatility can be inferred from option prices. Both volatility measures are model-free because the calculations do not require any particular model. Generally realized volatility is constructed by summing the squared asset returns over very small time intervals in a particular trading period, and is used as an approximation for the quadratic variation of the log price process. As Andersen, Bollerslev, and Diebold (2003), and Bandorff-Nielsen and Shephard (2002) showed, this type of volatility measure provides very accurate ex-post observations of the actual unobservable volatility. On the contrary, the implied volatility is the ex-ante risk-neutral expectation of future market volatility. The construction of this measure is different from the traditional method of implied volatility which depends on the Black-Scholes pricing formula. As Britten-Jones and Neuberger (2000) demonstrated, calculation of implied volatility does not depend on any particular option pricing model since the risk-neutral integrated return variance is only determined by current options prices.

The development of volatility proxies has opened a new door in the estimation of the continuous-time SV models. For example, Bandorff-Nielsen and Shephard (2002) used the quasi-maximum likelihood (QML) estimation method based on the time series of realized volatility to estimate the parameters of continuous-time SV models. Bollerslev and Zhou (2002) constructed daily realized volatility from the high-frequency five-minute foreign exchange rates, and applied a GMM method to estimate the affine diffusion allowing no instantaneous correlation between asset return and volatility processes. Bollerslev, Gibson and Zhou (2010) employed a GMM with the sample moments of realized volatility and implied volatility to estimate the volatility risk premium or risk aversion, the estimation was straightforward and easy to implement. Bollerslev and Zhou (2006) studied the relationships between return and realized volatility, return and model-free implied volatility based on the affine continuous-time SV model. Corradi and Distaso (2006) used the realized volatility and employed both GMM and the simulated method of moments (SMM) for the eigenfunction SV models. Ait-Sahalia and Kimmel (2007) studied the maximum likelihood estimation of SV models including the affine continuous-time SV (the Heston) model. By proposing closed form approximations to the true unknown likelihood function of the joint observations of the underlying asset price and model free implied volatility, they found using an implied volatility proxy did not have adverse consequences, and resulted in a large computational efficiency gain. Phillips and Yu (2009) used realized volatility and proposed a two-stage approach to the estimation of diffusion processes. In their study, they isolated the parameters in the diffusion function from those in the drift function. In the first stage, making use of the central limit theory for realized volatility, they estimated the diffusion function parameters by running a nonlinear least squares regression of the standardized realized volatility. In the second stage, the maximum approximate log-likelihood function was applied to obtain the estimates of parameters in the drift function. Todorov (2009) employed a GMM approach for general continuous-time SV models containing price jumps using realized volatility.

Motivated by the accuracy of these volatility proxies and computational convenience by using them in the estimation, in this study, we use both realized volatility and model free implied volatility and employ the C-AMLE as well as the QML procedures for the affine continuous-time SV model. Our empirical analysis is based on returns and volatilities of both the S & P 500 and Dow Jones Industrial Average Indexes. Specifically, when realized volatility is used as a proxy in the model, we construct both daily and monthly returns and realized volatilities for these two markets and use them in our estimation. Moreover, we also estimate the affine continuous-time SV model using monthly returns, realized volatilities, and model-free implied volatilities of both equity markets. The rest of this paper is organized as follows. In section 5.2 we briefly introduce the affine continuous-time model and a necessary model transformation. We also list the estimation methods employed in this study. In the next section, we discuss two volatility proxies, namely realized volatility and model free implied volatility. Section 5.4 reports the empirical results. A brief conclusion follows.

5.2 Affine Continuous-Time SV Model and Model Transformation

5.2.1 Model Specification

The affine continuous-time SV model consists of two stochastic differential equations (SDE):

$$dx_t = \mu dt + \sqrt{V_t} dB_{1t} \tag{5.1a}$$

$$dV_t = \beta(\alpha - V_t)dt + \sigma\sqrt{V_t} dB_{2t} \tag{5.1b}$$

$$dB_{1t} dB_{2t} = \rho dt \tag{5.1c}$$

where x_t, V_t are state variables, B_{1t}, B_{2t} are two different standard Brownian Motions, and $\alpha, \beta, \rho, \sigma, \mu$ are five parameters.

The affine continuous-time SV model was proposed by Heston (1993) (hence is also known as the Heston model). As we discussed in Chapter 4, the Heston model can capture many stylized facts of financial time series, moreover, it provides a closed-form solution for the European option prices and the CCF of the asset return, hence has been widely used in both econometrics and empirical finance literature. But in general volatility is treated as latent in such model. In this chapter, we employ two types of volatility measures, namely realized volatility and model-free implied volatility, in our estimation. However, it is noted that the model above is constructed under the physical measure P, while the model-free implied volatility represents the risk-neutral expectation of future market volatility, it is only well defined under the equivalent Martingale measure Q. Some transformation is necessary.

Define a standard Brownian motion under measure Q, B_{1t}^Q, as

$$B_{1t}^Q = B_{1t} + \int_0^t \frac{\mu-\mu^*}{\sqrt{V_s}} ds$$

where the probability measure Q is defined by its Radon-Nikodym derivative with respect to P,

$$\frac{dQ}{dP}\Big|_{\mathcal{F}_\tau} = e^{-\int_0^\tau \omega_t dB_{1t} - \frac{1}{2}\int_0^\tau \omega_t^2 dt}$$

with $\omega_t = \frac{\mu-\mu^*}{\sqrt{V_t}}$.

The term $\int_0^t \frac{\mu-\mu^*}{\sqrt{V_s}} ds$ is known as the market price of risk. Differentiating and rearranging yields:

$$dB_{1t}^Q = dB_{1t} + \frac{\mu-\mu^*}{\sqrt{V_t}} dt$$

Therefore

$$\begin{aligned} dx_t = & \quad \mu dt + \sqrt{V_t} dB_{1t} \\ = & \quad \mu dt + \sqrt{V_t}(dB_{1t}^Q - \frac{\mu - \mu^*}{\sqrt{V_t}} dt) \\ = & \quad (\mu - \sqrt{V_t}\frac{\mu - \mu^*}{\sqrt{V_t}})dt + \sqrt{V_t} dB_{1t}^Q \\ = & \quad \mu^* dt + \sqrt{V_t} dB_{1t}^Q \end{aligned}$$

Define a Q-Brownian motion

$$B_{2t}^{Q} = B_{2t} + \omega_t t$$

Specify the price of risk $\omega_t = \sqrt{V_t}$,

$$\begin{aligned}
dV_t = & \quad \beta(\alpha - V_t)dt + \sigma\sqrt{V_t}dB_{2t} \\
= & \quad \beta(\alpha - V_t)dt + \sigma\sqrt{V_t}dB_{2t}^{Q} - \sigma\sqrt{V_t}\omega_t dt \\
= & \quad (\beta\alpha - \beta V_t - \sigma V_t)dt + \sigma\sqrt{V_t}dB_{2t}^{Q} \\
= & \quad (\beta + \sigma)(\frac{\beta\alpha}{\beta + \sigma} - V_t)dt + \sigma\sqrt{V_t}dB_{2t}^{Q} \\
= & \quad \beta^*(\alpha^* - V_t)dt + \sigma\sqrt{V_t}dB_{2t}^{Q}
\end{aligned}$$

Hence the "adjusted", or "risk-neutralized" affine continuous-time SV model is written as:

$$dx_t = \mu^* dt + \sqrt{V_t}dB_{1t}^{Q} \tag{5.2a}$$

$$dV_t = \beta^*(\alpha^* - V_t)dt + \sigma\sqrt{V_t}dB_{2t}^{Q} \tag{5.2b}$$

$$dB_{1t}^{Q}dB_{2t}^{Q} = \rho dt \tag{5.2c}$$

where μ^* refers to the risk-neutral interest rate.[1] The "adjusted" long-run mean parameter $\alpha^* = \frac{\beta\alpha}{\beta+\sigma}$, and mean-reversion parameter $\beta^* = \beta + \sigma$. The parameter ρ again controls the skewness of the distribution. $\rho > 0$ induces a right-skewed distribution, and $\rho < 0$ induces a left-skewed distribution. The functional form of the model is invariant with respect to switching from the physical measure P to the equivalent Martingale measure Q. $\mu^*, \alpha^*, \beta^*, \rho, \sigma$ are time and state homogenous coefficients which we need to estimate.

[1]Usually one uses r for the risk-neutral interest rate, we use μ^* to compare it with the original asset return parameter μ.

5.2.2 C-AMLE And QML Estimation Methods by Using Volatility Proxies

In Chapter 4, we discussed the C-AMLE procedure and examined the estimation of this procedure for the affine continuous-time SV model when the underlying volatility was observed via a Monte Carlo study, the evidence showed that this approach did a good job at recovering the true parameters. In this chapter, we extend our investigation via an empirical study in which a volatility proxy is employed. The system of the C-AMLE estimation equations is following:

$$\frac{1}{T}\sum_{t=1}^{T-1}\{\frac{\partial K^i}{\partial\theta}h_i - \frac{1}{2}\frac{\partial\lambda_{i,j}}{\partial\theta}[(S^i_{t+1} - K^i)(S^j_{t+1} - K^j) - K^{i,j}]$$
$$+\frac{1}{6}[\frac{\partial K^{i,j,k}}{\partial\theta}(h_{ijk} - E(h_{ijk}|S_t)) + 3K^{i,j,k}((\bar{h}_i h_j h_k - E(\bar{h}_i h_j h_k|S_t))$$
$$-\bar{z}_i(h_j h_k - E(h_j h_k|S_t)) - \bar{h}_i\lambda_{j,k} - h_i\frac{\partial\lambda_{j,k}}{\partial\theta})]$$
$$+\frac{1}{24}[\frac{\partial K^{i,j,k,l}}{\partial\theta}(h_{ijkl} - E(h_{ijkl}|S_t)) + K^{i,j,k,l}(4(\bar{h}_i h_j h_k h_l - E(\bar{h}_i h_j h_k h_l|S_t))$$
$$-4\bar{z}_i(h_j h_k h_l - E(h_j h_k h_l|S_t)) - 12(\bar{h}_i h_j\lambda_{k,l} - E(\bar{h}_i h_j\lambda_{k,l}|S_t))$$
$$+12\bar{z}_i h_j\lambda_{k,l} - 6(h_i h_j - E(h_i h_j|S_t))\frac{\partial\lambda_{k,l}}{\partial\theta})]\} = 0 \quad (5.3)$$

The vector θ consists of five parameters, consequently there are five equations. $S_{t+1} = \{x_{t+1}, V_{t+1}\}$ is the vector of state variables. Plugging the observations of asset return together with those of a volatility proxy into these equations, the solution yields the estimates for $\mu^*, \alpha^*, \beta^*, \rho$, and σ [2].

[2]The detailed discussion about these estimation equations is provides in the Appendix one in Chapter 4.

As a special case, if the logarithmic transition density function is approximated as a log-normal distribution, then the approach is essentially the quasi-maximum likelihood (QML) method. The system equations are simplified as:

$$\frac{1}{T}\sum_{t=1}^{T-1}\{\frac{\partial K^i}{\partial \theta}h_i - \frac{1}{2}\frac{\partial \lambda_{i,j}}{\partial \theta}[(S^i_{t+1} - K^i)(S^j_{t+1} - K^j) - K^{i,j}]\} = 0 \tag{5.4}$$

5.3 Volatility Proxies

The underlying volatility entering the affine continuous-time SV model is not directly observed, while some volatility proxies have been developed and employed for estimating and forecasting purpose in econometrics and empirical finance literature over the last two decades. In this study, we employ both realized volatility and model-free implied volatility. As we discussed in Chapter 3, realized volatility is a consistent, efficient estimator for the latent underlying volatility. We employed this volatility proxy in the estimation and forecasting of the discrete-time SV models, empirical evidence showed that the estimates were stable, and forecasting performances of different models were easily compared. In this chapter, we extend our study of SV model by using realized volatility in the estimation of the affine continuous-time SV model. In addition, we employ another popular volatility proxy, model-free implied volatility, in our estimation. The estimation performance by using these different volatility proxies hence can be examined.

5.3.1 Realized Volatility Construction

In a trading period t, the transaction prices $P_{t,d}, d = 1, \ldots, D$ are observed, where D is the number of observations in the period t. The asset return can be computed by $r_{t,d} = log(P_{t,d}) - log(P_{t,d-1})$. A standard formula for realized volatility in the trading period t is:

$$RV_t = \sum_{d=0}^{D} r_{t,d}^2$$

We have discussed realized volatility, its distribution in Chapter 3. In the same chapter, we considered different approaches to construct daily realized volatility. For instance, we considered an approach to deal with the closed-part effect when we constructed daily realized volatility of equity markets; we also considered an approach to take into account the first-order autocorrelation of high-frequency asset returns[3]. Our evidence showed that realized volatilities constructed from these different approaches yielded similar estimation and forecasting results. In this chapter, we employ the standard approach above. As we mentioned, we construct realized volatilities of S & P 500 index as well as those of Dow Jones Industrial Average indexes at both daily and monthly frequencies. The daily realized volatilities are constructed by summing five-minute squared asset returns within each trading day. Specifically, as equity markets open from 9:30am to 4:00pm, in total we have 78 five-minute returns in each trading day, the daily realized volatilities are thus constructed by:

$$RV_t = \sum_{d=1}^{78} r_{t,d}^2$$

Similarly, there are approximately 22 trading days in each month, the monthly realized volatilities are constructed by summing the squared daily returns:

$$RV_t = \sum_{d=1}^{22} r_{t,d}^2$$

[3]Detailed discussion is provided in Chapter 3.

5.3.2 Model-Free Implied Volatility

In contrast to realized volatility which affords the ex-post observations of the true latent volatility, the implied volatility provides the ex-ante risk neutral expectation of future financial market volatility. The traditional implied volatility is derived based on the Black-Scholes model (hence BSIV). Although the BSIV has been commonly used to forecast volatility for a long time, it is well recognized that the constant volatility assumption of the BSIV violates the reality. Instead, Breeden and Litzerberger (1978) showed that the entire risk-neutral distribution and model-free variance of the underlying asset could be extracted from options prices. Further, in Britten-Jones and Neuberger (2000), they first derived model-free implied volatility under the diffusion assumption. As demonstrated in Britten-Jones and Neuberger (2000), the risk-neutral integrated return variance is entirely determined by current option prices. Consequently the calculation of implied volatility should not depend on any particular option pricing formula.

Generally, as shown in Britten-Jones and Neuberger (2000), the model-free implied volatility is calculated by:

$$MFIV_{t,t+\Delta} = 2\int_{0}^{\infty} \frac{C(t+\Delta, K) - C(t, K)}{K^2} dK \tag{5.5}$$

where $C(t, K)$ refers to the price of a European call option maturing at time t with strike price K.

Britten-Jones and Neuberger (2000) showed that the model-free implied volatility is equivalent to the risk-neutral expectation of the integrated volatility of future financial market:

$$MFIV_{t,t+\Delta} = E^{*}(V_{t,t+\Delta}|F_t) \tag{5.6}$$

where E^{*} refers to the expectation under the equivalent Martingale measure Q.

Britten-Jones and Neuberger (2000) illustrated their method assuming that the underlying price path is continuous. Further, Jiang and Tian (2005) extended Britten-Jones and Neuberger (2000)'s study by considering asset price processes with jumps and showed that the computation of model-free implied volatility in this case was still valid. The concept of model-free implied volatility has been widely accepted and applied in modern empirical finance. For example, the Chicago Board Options Exchange (CBOE) started to provide the new VIX index for the S & P 500 volatility since September 23, 2003[4] and VXD index for the Dow Jones Industrial Average volatility since March 2, 2007. These two implied volatilities are both model-free and calculated on the concept of fair value of future variance which is theoretically identical to the approach developed in Britten-Jones and Neuberger (2000). In our study, we construct the model-free implied volatilities of both the S & P 500 and the Dow Jones Industrial Average indexes based on the new VIX and VXD indexes provided by the CBOE. We name the measure from these indexes as MFIV1. As the new VIX index is square root of the risk neutral expectation of S & P 500 variance over the next 30 calender days and quoted on an annualized variance basis, we first take square then de-annualize hence obtain the MFIV1 for the S & P 500 index. The MFIV1 for the Dow Jones Industrial Average index is calculated

[4]The old VIX index provided by the CBOE until September 21, 2003 was based on index options for the S & P 100, and relied on the Black-Scholes formula. It is now labeled VXO to be distinguished from the new VIX.

from the same procedure since the VXD is calculated using the same methodology as the VIX. We compare the first four moments, the extreme values, the percentiles, and the ACF values of our MFIV1 of S & P 500 index with those calculated in Bollerslev and Zhou (2006), all values are very similar.

In their recent study, Jiang and Tian (2007) refined the computation of the model-free implied volatility. They examined the new VIX provided by the CBOE. As they showed, the truncation errors induced by the limited availability of strike prices, the discretization error induced by numerical integration, and expansion error induced by the Taylor series expansion of the log function used in the CBOE procedure resulted in either underestimation or overestimation of the true volatility. To fix the problem, they proposed a smooth method for extracting model-free implied volatility from option prices. In their study, the interpolation was first implemented between listed strike prices to construct a smooth function that exactly fitted the known implied volatilities, then the extrapolation procedure was implemented outside the range of listed strike prices to construct an extension of the implied volatility function in the two tails of the strike price distribution. As they demonstrated, their method ensured that the constructed implied volatility function was smooth over the entire range of strike prices, and the constructed implied volatilities were accurate and robust. To investigate whether using implied volatility introduced by Jiang and Tian (2007) will improve the estimation of continuous-time SV model, we also construct the model-free implied volatility of S & P 500 index based on Jiang and Tian(2007)'s method and name it as MFIV2.

5.4 Empirical Application

Our empirical analysis is based on returns and volatilities for the S & P 500 and Dow Jones Industrial Average indexes at both daily and monthly frequencies. The daily returns and realized volatilities for the S & P 500 index span the period from March 4, 2003 through September 24, 2008 with 1402 daily observations. Those for the Dow Jones Industrial Average indexes cover the period from April 1, 2003 through September 19, 2008 with the sample size $T = 1378$. The monthly returns, realized volatilities, and MFIV1 for the S & P 500 index span the period from January 1996 through December 2009 with $T = 168$. Those for the Dow Jones Industrial Average indexes cover the period from October 1997 through December 2009 with sample size $T = 147$. As we mentioned we also construct MFIV2 of S & P 500 index based on Jiang and Tian (2007)'s implied volatilities, the data is available from January 1996 through May 2004 with sample size $T = 101$. The daily realized volatilities for both equity markets are constructed from the five-minute high frequency data which are kindly provided by Dinghai Xu. The monthly realized volatilities of the S & P 500 index are constructed from the daily closing prices provided by the CBOE, and those of the Dow Jones Industrial Average index are constructed from the daily closing prices provided by Yahoo finance. The monthly returns for these markets are obtained from Wharton Research Data Services (WRDS). Both VIX and VXD indexes are downloaded from the CBOE web. The MFIV2 of the S & P 500 index are constructed from implied volatilities kindly provided by George Jiang.

5.4.1 Descriptive Statistics

The summary statistics of daily return, squared return, realized volatility and logarithmic realized volatility series for the S & P 500 index are reported in Table 5.1. Those for the Dow Jones Industrial Average are reported in Table 5.2. In each table, the top panel shows the first four moments of these series. The middle panel reports the minimum, the 5th, 25th, 50th, 75th, 95th percentiles, and the maximum values. And the bottom panel displays the autocorrelation functions up to lag 10. The mean value of S & P 500 returns is 0.0252 with standard deviation 0.9189. The value of skewness is -0.2466 and that of kurtosis is 5.9066. The negative skewness value along with the large kurtosis value (much greater than 3) suggest that the return series are negatively skewed and have fatter tails comparing with a normally distributed random variable. The range of the returns is 9.0314. The ACF values of the return series are mostly very small in absolute value (less than 0.1) suggesting that there is no significant serial correlation, however, the ACF values of the squared daily returns are all over 0.1. This finding is consistent with the stylized fact that the return series do not exhibit serial correlation while the squared or absolute return series display pronounced serial correlation. The mean value of S & P 500 daily realized volatilities is 0.7566 with standard deviation 1.3150. The positive value of skewness along with large value of kurtosis indicate that daily realized volatility series do not follow a normal distribution, instead, they are positively skewed with fat tails. The range is 22.1412. The ACF values are all large with slow decay suggesting a strong serial correlation between volatilities. This finding is consistent with the stylized fact that volatility series are clustering and persistent hence predictable. On the other hand, when we take logarithm of the realized volatility series, the value of skewness decreases significantly towards zero, moreover, the kurtosis value decreases dramatically towards 3, suggesting that the distribution of logarithmic realized volatility is close to a normal distribution, which is a common finding in the literature. The ACF values of the logarithmic realized volatilities are all very large, suggesting a pronounced serial correlation. The moments, quantiles, and ACF values of the Dow Jones Industrial Average in Table 2 display the similar patterns.

Figure 5.1-5.2 plot the S & P 500 daily return and realized volatility series, respectively. Overall, the return series display a stationary property, all the values are fluctuating around the long-run mean. We notice that the return values are high in absolute value in early 2003 then they are relatively smooth from 2004 through 2006, after mid-2007, the returns series again have high absolute values, and in September 2008, the absolute values of returns reach the peak. The relatively large absolute return values in 2003, 2007-2008 correspond the stock market downturn from 2002, the second U.S. bear market of the 21st century as a result of subprime mortgage lending crisis, respectively. The observation above is confirmed by the plot the daily realized volatility series. The volatilities are relatively small from mid-2003 to mid-2007, then they are very high after mid-2007 and experience an unusually high value in 2008 corresponding the acute crisis during that period. Figure 5.3-5.4 plot the Dow Jones Industrial Average daily return and volatility series, respectively. These plots display similar patterns as those of S & P 500 index which is not surprising since the two equity indexes are related.

Table 5.3-5.5 report the summary statistics of monthly returns, realized volatilities, model-free implied volatilities and logarithmic volatilities for both equity markets. Same as Table 5.1 and 5.2, the top panels show the first four moments, middle panels report the quantiles and extreme values, and the bottom panels display the ACF values. In each table, the second column reports results for the return process, the third to fourth columns report those for monthly realized volatilities and logarithmic realized volatilities. And the fifth to sixth columns report results for monthly MFIVs and logarithmic MFIVs. All three tables display similar patterns. From the examination of these tables, we find that the monthly return series are not normally distributed. Same as daily returns, they are all negatively skewed with fat tails, in addition, the standard deviation values are very large. The range values of monthly returns are between 24 to 27. The ACF values of monthly returns are very small and non-significant suggesting that the series are not serial correlated. These observations along with those from daily returns suggest that the return series are negatively skewed with fat tails and at low frequency (such as daily, monthly) they are not serial correlated. However the range values of monthly returns are much larger than those of daily returns. The mean values of monthly realized volatilities of both equity markets are all over 30, and the values of skewness and kurtosis are much larger than zero and 3, respectively, indicating that same as the daily realized volatilities, the monthly realized volatilities are distributed with positive skewness and high kurtosis. Also same as the daily realized volatilities, the monthly realized volatilities exhibit a pronounced serial correlation. Again, the values of skewness and kurtosis for the logarithmic monthly realized volatilities decrease dramatically towards zero and 3 respectively, suggesting that they are approximately a Gaussian process. The mean values of MFIVs are over 40, much larger than those of realized volatilities. Same as realized volatilities, the values of skewness and kurtosis of MFIVs are all greater than zero and 3 respectively suggesting that the distributions of MFIVs are positively skewed with fat tails. Also ACF values are all very large suggesting that there is significant serial correlation between the MFIVs. The logarithmic MFIVs display similar patterns as the logarithmic realized volatilities. Comparing the monthly realized volatilities and MFIVs, we find MFIVs have higher mean values. The common features of both series are those of positive skewness and fat tails. The ACF values of both series are high and decay very slowly, suggesting that both proxies are able to capture the stylized facts of true volatilities, such as non-normal distribution, volatility persistency.

Figure 5.5, 5.7, 5.9 plot monthly returns for both equity indexes. Figure 5.6, 5.8, 5.10 report the plots monthly realized volatilities along with monthly model-free implied volatilities in which the dash line represents realized volatilities and the solid line is the model-free implied volatilities . For both indexes, the lowest returns occurred in August 1998, 2002, and mid-2008 indicating that the Asian finance crisis, the stock market downturn of 2002, and the subprime mortgage lending crisis which significantly affected the equity indexes. Overall the monthly return series display a stationarity property, the returns take either positive or negative values but all fluctuate around the long run mean. The above finding is confirmed by the plots of the volatility series. From Figure 5.6, 5.8, 5.10, we notice that for both volatility series, the most volatile periods are 1998, 2002, and 2008. In 2008, both volatilities reached the unusual high value corresponding the greatest equity loss since 1931. Comparing the plots of these two volatility series, we find although in general the values of MFIVs are higher than those of realized volatilities, they display similar patterns. They are both able to capture the volatile periods for both equity indexes, indicating that both volatility proxies are accurate estimators for the latent underlying volatility.

As discussed in Jiang and Tian (2007), the VIX index provided by the CBOE either underestimates or overestimates the true volatility, to fix the problem, they proposed a smooth method. In Figure 5.11, we plot both MFIV1 from the VIX and MFIV2 from Jiang and Tian (2007)'s volatilities during January 1996 through May 2004. The dash line is MFIV1, and the solid line represents MFIV2. We find the values of both model-free implied volatilities are almost identical during the volatile periods, however, the values of MFIV2 are always higher than those of MFIV1 when volatilities have low values, suggesting that Jiang and Tian (2007)'s approach fixed the problem of underestimation significantly.

5.4.2 Empirical Results

Table 5.6 reports the parameter estimates and their asymptotic standard errors using daily realized volatilities. The top panel reports the results of S & P 500 index and the bottom panel shows the results of Dow Jones Industrial Average indexes. In each panel results from both the QML and the C-AMLE approaches are reported. Overall, the two sets of data provide similar parameter estimates except that the signs of the estimated ρ, μ^* from the C-AMLE procedure are positive for S & P 500 index while they are negative for Dow Jones Industrial Average indexes. Most estimates are statistically significant regardless of the estimation method except those of parameter μ^*. We notice that the values of estimated α^*, β^*, σ, and μ^* from the C-AMLE method are much larger than those from the QML approach. Based on these estimates, we are able to calculate the first four unconditional moments of returns and realized volatilities by plugging the estimates into the analytical moments derived from the model [5]. Table 5.7 reports the moments for both equity indexes. Comparing

[5] the formulas are provided in Chapter 4.

these moments with those calculated from empirical data which are reported at the top panel of Table 5.1-5.2 (the second column reports the moments of return series, and the fourth column reports moments of realized volatilities), we find both QML and C-AMLE estimates can capture the asymmetric behavior of both returns and realized volatilities of Dow Jones Industrial Average indexes, however, the value of skewness of S & P 500 index returns is positive based on the C-AMLE estimates while it is negative based on the QML estimates. On the other hand, the signs of skewness and kurtosis of realized volatilities are always positive, and the magnitude of kurtosis is larger than 3, consistent with those calculated from realized volatility series. However, we also find some of the moments especially the fourth moments of realized volatilities are different from those calculated from their data counterparts, the possible reason is the realized volatilities are unusually high in 2008, the high kurtosis value induced by those unusual high values is hard to be captured. Comparing the moments from two estimation methods, we find for S & P 500 index, the C-AMLE approach outperforms the QML method at capturing the distribution of returns since the mean and kurtosis values from the former are much closer to their data counterparts. For Dow Jones Industrial Average, the third and fourth moments of returns from the C-AMLE estimates are closer to their data counterparts, while the mean value is negative while that from real data is positive.

Table 5.8 reports the estimates and their asymptotic standard errors from S & P 500 monthly returns and realized volatilities, and Table 5.9 reports results using S & P 500 monthly returns and MFIV1 based on the VIX index. Most estimates have relatively small standard errors regardless of the estimation method and volatility proxy, suggesting that the estimates are stable. The signs of all estimates are consistent across different methods and volatility proxies. The estimated correlation coefficient parameter ρ are always negative, implying the negative skewed return distribution which is typically observed in financial markets. From examination of Table 5.8, we find the estimates of $\alpha^*, \beta^*, \sigma$ from the C-AMLE approach are higher than those from the QML method, whereas the estimates of ρ, μ^* are smaller in absolute value. From Table 5.9, we notice that all the estimates from the C-AMLE method are greater than those from the QML in absolute value. Table 5.10 shows the moments based on estimates. Overall, the moments based on estimates from monthly data are closer to their empirical counterparts comparing with those based on estimates from daily data. In particular, when MFIV1 is used as a volatility proxy, most moments of MFIV1 based on the C-AMLE estimates are close to those computed from real data, and the moments of returns from the C-AMLE method are also closer to moments from data comparing with those using realized volatilities, suggesting that the C-AMLE method outperforms the QML procedure when MFIV1 is employed in the estimation. It is noticed that using monthly realized volatilities, the QML method overestimates the mean parameter of returns and underestimates the mean parameter of volatilities, while the C-AMLE procedure overestimates the mean parameter of volatilities, it provides an accurate estimate of return mean parameter.

Table 5.11-5.12 report the results using S & P 500 realized volatilities and MFIV2, respectively. When realized volatility is employed, the signs of estimates are consistent across different estimation methods, however when MFIV2 is used, the estimated ρ is positive while the estimated μ^* is negative when the QML is applied. When realized volatility is used, the estimates of α^* are different across different approaches, while when MFIV2 is used, the estimates of this parameter are very similar. We examine the moments reported in Table 5.13, and find the moments of returns and realized volatilities from the QML are much closer to their empirical counterpart than those from the C-AMLE, indicating that the QML approach fits data better than the C-AMLE when realized volatilities are applied. On the contrary, when MFIV2 is employed as a volatility proxy, the C-AMLE procedure provides more stable and accurate estimates than the QML. For example, the negative estimated ρ from the C-AMLE suggests that the negatively skewed distribution of returns is captured. When we compare the moments from these methods with those from real data, we find almost all moments of return series and those of MFIV2 based on the C-AMLE estimates are close to their empirical counterparts suggesting the C-AMLE approach outperforms the QML method when MFIV2 is employed.

Table 5.14-5.15 report the results using Dow Jones Industrial Average indexes realized volatilities and MFIV1, respectively. And Table 5.16 displays the moments calculated from both the QML and the C-AMLE approaches. From examination of Table 5.14, 5.15, we find most estimates from the QML method have relatively large standard errors, whereas the estimates from the C-AMLE approach have much smaller standard errors, suggesting the C-AMLE provides more stable estimates. Comparing moments in Table 5.16 with those in Table 5.5, we find when using realized volatilities, the moments of returns and realized volatilities from the C-AMLE procedure are closer to those calculated from real data. This is also the case when MFIV1 series are used except that C-AMLE overestimates the long run mean of MFIV1 while QML underestimates it. These findings suggest that the C-AMLE approach fits Dow Jones Industrial Average indexes monthly data better than the QML.

In summary, we find when daily realized volatilities are used in the estimation, in general neither the QML nor the C-AMLE approach does a reasonable job. However, when using monthly data, the estimates especially those from the C-AMLE procedure are mostly stable and are able to capture the negative skewness and fat tails of return distribution as well the positive skewness and fat tails of volatility distributions. One possible reason is the daily data spans the period from about April 2003 through September 2008, during which the U.S. equity markets experienced unusual bear market of the 21st century as a result of subprime mortgage lending crisis. The dynamic properties of return and volatility processes during these extreme volatile period are difficult to capture. On the other hand, the monthly data span from 1990s through 2009, during which the U.S. equity markets experienced volatile but also relatively longer period of smoothness. We notice that the moments calculated from the C-AMLE estimates are not perfectly but relatively close to those computed from real data. For the S & P 500 index, we use both MFIV1 based on the VIX provided by the CBOE and MFIV2 based on volatilities from Jiang and Tian (2007)'s method, the results are similar, suggesting that Jiang and Tian (2007)'s smooth method does not improve the estimation of the affine continuous-time SV model significantly.

5.5 Conclusion and Extension

In this chapter, we investigated the estimation of the affine continuous-time SV model using two different volatility proxies via an empirical study. We applied both the QML and the C-AMLE approaches and our empirical analysis was based on daily and monthly data of S & P 500 and Dow Jones Industrial Average indexes. The evidence showed that using daily realized volatility, neither method did a good job. However, when using monthly data, the estimation improved. Especially when we employed the model-free implied volatilities, the estimates were stable, and the moments calculated from the estimates were close to the moments of real time series. In summary, we found the C-AMLE method outperformed the QML approach, and using model-free implied volatilities yielded more stable estimates than employing realized volatilities.

When the volatility state variable is unobserved, other estimation methods, such as the generalized method of moments (GMM) can be applied. In our future study, to further investigate the estimation performance of the C-AMLE procedure, we will extend our research by comparing the estimates of the C-AMLE using volatility proxies with those applying the GMM method when volatility is unobserved.

5.6 Appendix

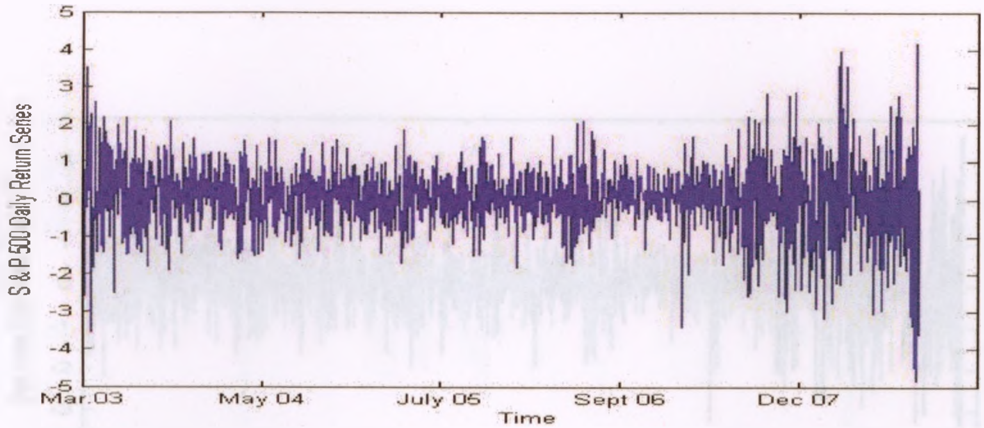

Figure 5.1: Plot of S & P 500 Daily Return Series

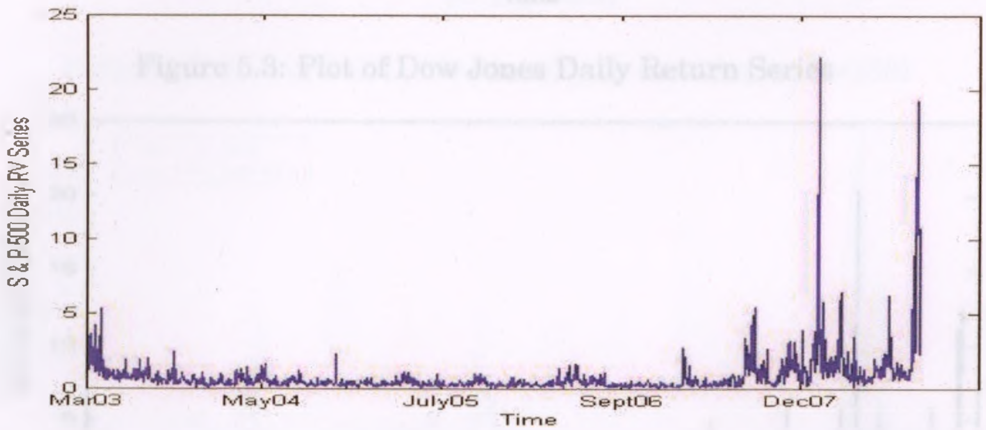

Figure 5.2: Plot of S & P 500 Daily Realized Volatility Series

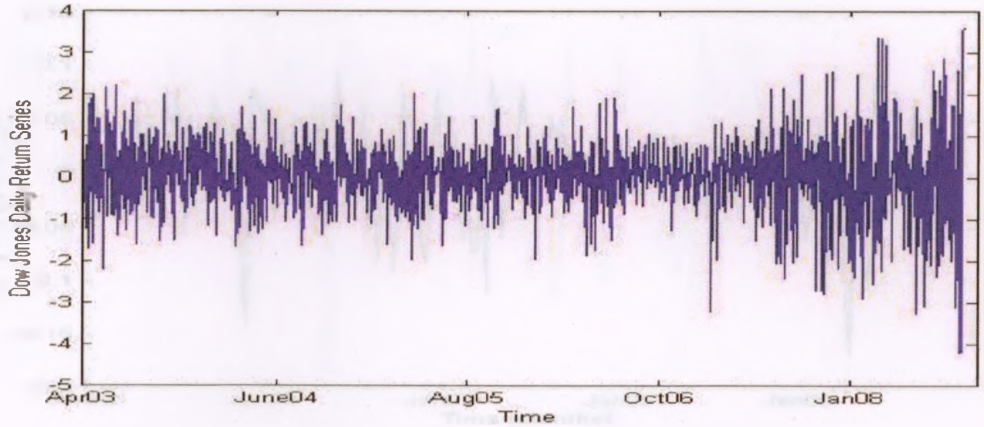

Figure 5.3: Plot of Dow Jones Daily Return Series

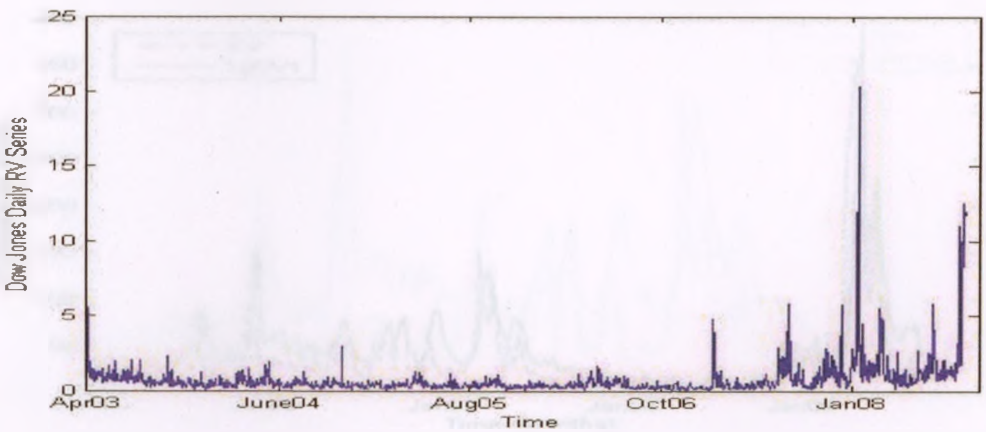

Figure 5.4: Plot of Dow Jones Daily Realized Volatility Series

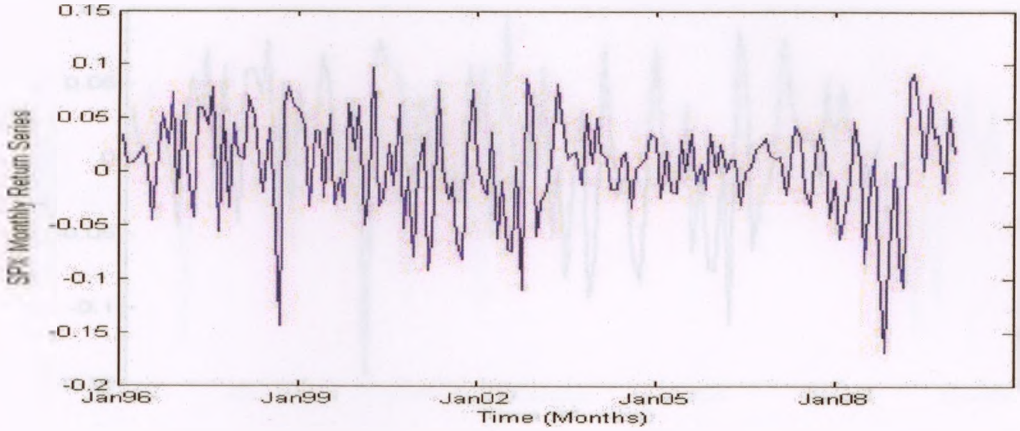

Figure 5.5: Plot of S & P 500 Monthly Return Series (T=168)

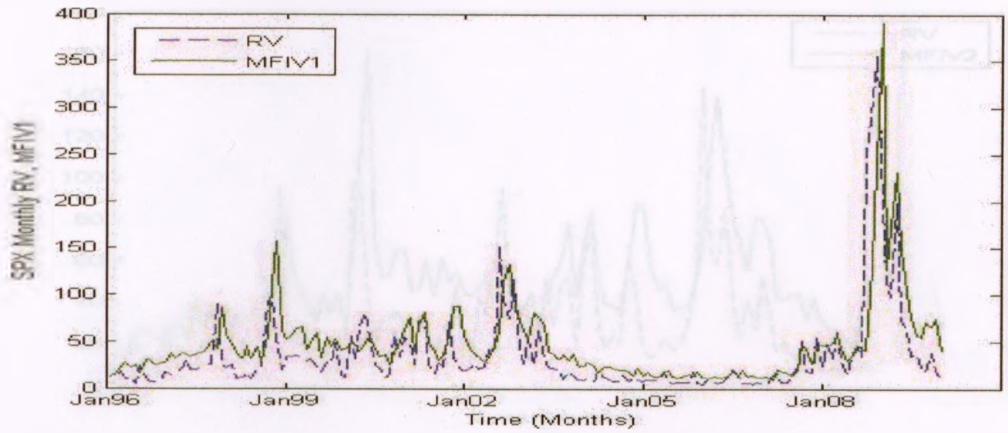

Figure 5.6: Plot of S & P 500 Monthly RV and MFIV1 Series (T=168)

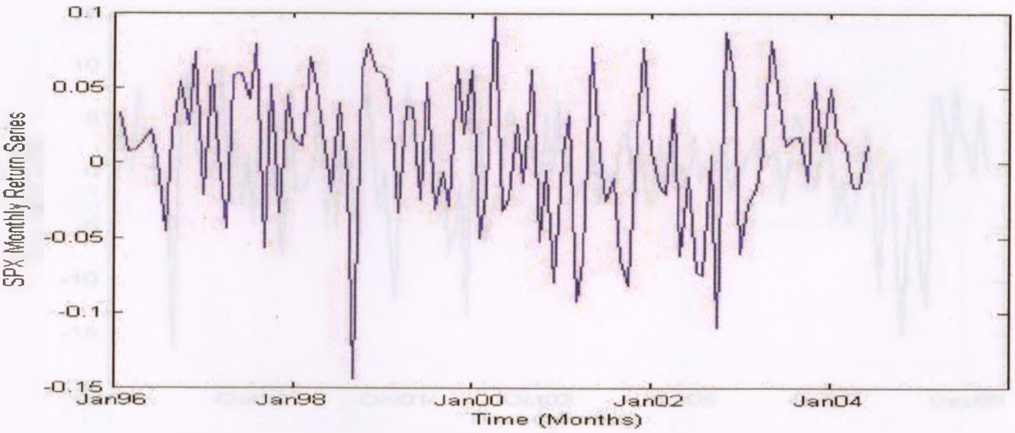

Figure 5.7: Plot of S & P 500 Monthly Return Series (T=101)

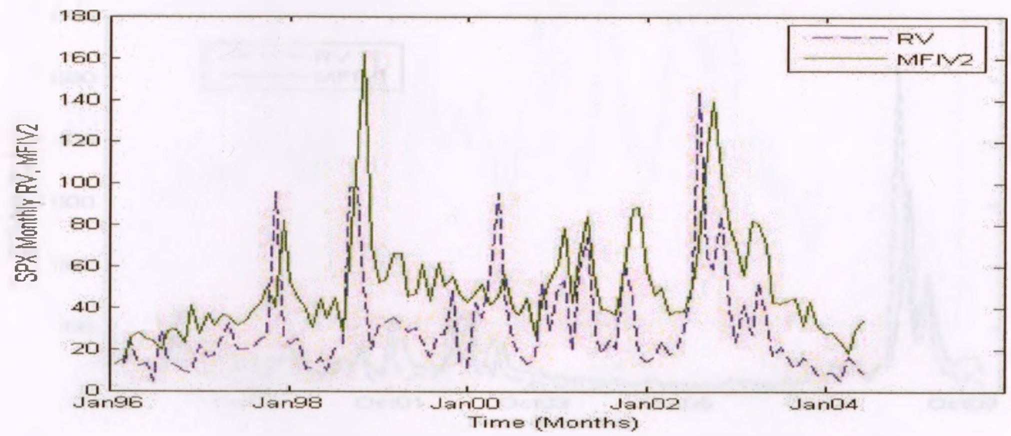

Figure 5.8: Plot of S & P 500 Monthly RV and MFIV2 Series (T=101)

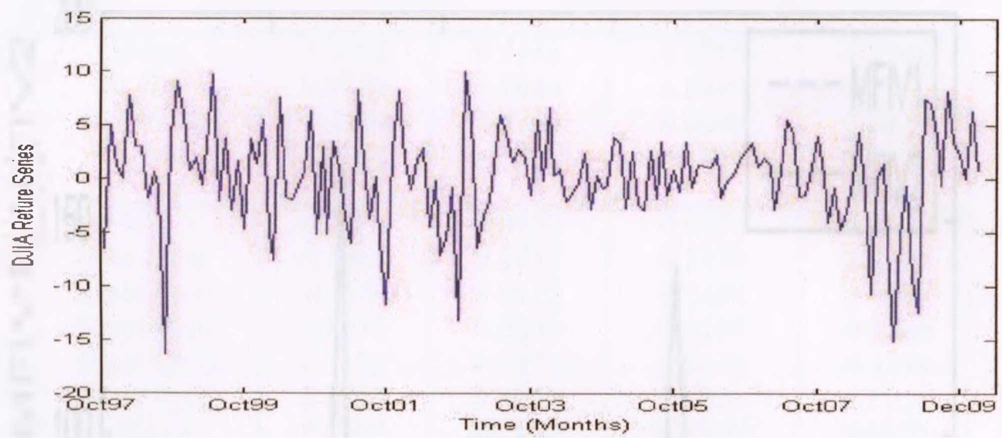

Figure 5.9: Plot of Dow Jones Monthly Return Series

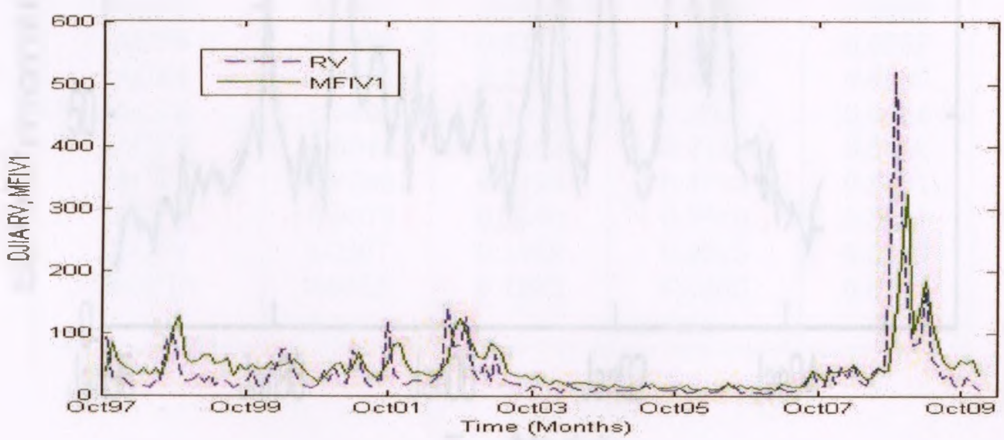

Figure 5.10: Plot of Dow Jones Monthly RV and MFIV1 Series

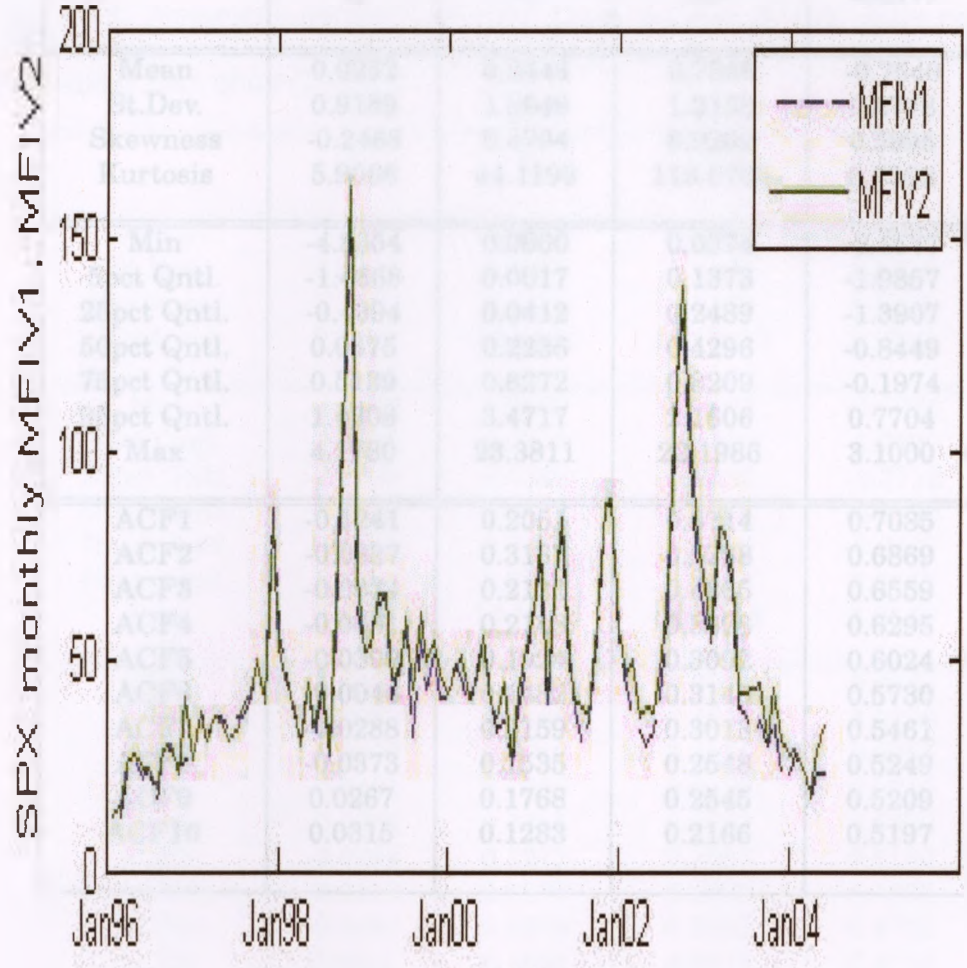

Figure 5.11: Plot of S & P 500 Monthly MFIV1, MFIV2 (T=101)

Table 5.1: Summary statistics of S& P 500 daily return, squared return, RV, and ln(RV) during the period of March 4th 2003 through September 24th 2008 (T=1402)

	x_t	x_t^2	RV	ln(RV)
Mean	0.0252	0.8444	0.7566	-0.7540
St.Dev.	0.9189	1.8649	1.3150	0.8778
Skewness	-0.2466	5.4794	8.9202	0.5995
Kurtosis	5.9066	44.1199	113.5753	3.7302
Min	-4.8354	0.0000	0.0574	-2.8577
5pct Qntl.	-1.4668	0.0017	0.1373	-1.9857
25pct Qntl.	-0.4394	0.0412	0.2489	-1.3907
50pct Qntl.	0.0875	0.2236	0.4296	-0.8449
75pct Qntl.	0.5139	0.8272	0.8209	-0.1974
95pct Qntl.	1.4308	3.4717	2.1606	0.7704
Max	4.1780	23.3811	22.1986	3.1000
ACF1	-0.1241	0.2051	0.6714	0.7085
ACF2	-0.0327	0.3137	0.5238	0.6869
ACF3	-0.0074	0.2121	0.4865	0.6559
ACF4	-0.0081	0.2148	0.3628	0.6295
ACF5	-0.0309	0.1924	0.3092	0.6024
ACF6	-0.0046	0.1482	0.3149	0.5730
ACF7	-0.0288	0.2159	0.3013	0.5461
ACF8	-0.0373	0.1535	0.2548	0.5249
ACF9	0.0267	0.1768	0.2545	0.5209
ACF10	0.0315	0.1283	0.2166	0.5197

Table 5.2: Summary statistics of Dow Jones Indexes (DJIA) daily return, squared return, RV, and ln(RV) during the period of April 1st 2003 through September 19th 2008 (T=1378)

	x_t	x_t^2	RV	ln(RV)
Mean	0.0257	0.7415	0.7040	-0.7708
St.Dev.	0.8610	1.5398	1.1151	0.8316
Skewness	-0.2320	5.0372	8.5459	0.5373
Kurtosis	5.3408	38.7601	109.7932	3.9035
Min	-4.2258	0.0000	0.0463	-3.0726
5pct Qntl.	-1.4178	0.0009	0.1410	-1.9590
25pct Qntl.	-0.4039	0.0388	0.2538	-1.3712
50pct Qntl.	0.0512	0.2022	0.4386	-0.8243
75pct Qntl.	0.4881	0.7475	0.7790	-0.2497
95pct Qntl.	1.3858	3.2713	1.9057	0.6449
Max	3.5893	17.8574	20.3753	3.0143
ACF1	-0.0951	0.1509	0.6050	0.6645
ACF2	-0.0279	0.2406	0.4610	0.6320
ACF3	-0.0007	0.1762	0.4056	0.5955
ACF4	-0.0101	0.1988	0.3109	0.5710
ACF5	-0.0363	0.1365	0.2731	0.5370
ACF6	-0.0224	0.1242	0.2983	0.5130
ACF7	-0.0113	0.2172	0.2880	0.4826
ACF8	-0.0361	0.1639	0.2552	0.4725
ACF9	0.0344	0.1868	0.2512	0.4714
ACF10	0.0269	0.1301	0.2080	0.4644

Table 5.3: Summary statistics of monthly return, RV, MFIV1 for S & P 500 index during the period of January 1996 through December 2009 (T=168)

	x_t	RV	ln(RV)	MFIV1	ln(MFIV1)
Mean	0.4838	34.8632	3.0650	46.8196	3.5755
St.Dev.	4.6566	48.4525	0.9225	44.9078	0.7063
Skewness	-0.6924	4.0240	0.4928	3.9884	0.3303
Kurtosis	3.8871	22.4278	3.1691	25.7486	3.2441
Min	-16.8454	3.5791	1.2801	8.8580	2.1813
5pct Qntl.	-8.1680	5.4221	1.6905	11.8611	2.4731
25pct Qntl.	1.9766	10.8142	2.3807	22.1001	3.0956
50pct Qntl.	0.9956	20.5937	3.0249	37.2064	3.6165
75pct Qntl.	3.5847	34.2758	3.5344	54.9559	4.0065
95pct Qntl.	7.6954	99.6325	4.6002	128.3962	4.8551
Max	9.7434	356.2634	5.8757	391.1350	5.9691
ACF1	0.1234	0.7585	0.7601	0.7472	0.8767
ACF2	-0.0279	0.5587	0.6519	0.5647	0.7786
ACF3	0.0985	0.3936	0.5921	0.4893	0.7083
ACF4	0.0831	0.3101	0.5198	0.4115	0.6742
ACF5	0.0420	0.2507	0.4955	0.3132	0.6331
ACF6	-0.0478	0.2034	0.4556	0.2159	0.5719
ACF7	0.0555	0.0959	0.3779	0.1691	0.5385
ACF8	0.0975	0.0895	0.3944	0.1585	0.5148
ACF9	-0.0119	0.0349	0.3475	0.1586	0.4958
ACF10	-0.0075	0.0477	0.2895	0.1633	0.4833

Table 5.4: Summary statistics of monthly return, RV, MFIV2 volatilities for S & P 500 index during the period of January 1996 through May 2004 (T=101)

	x_t	RV	ln(RV)	MFIV2	ln(MFIV2)
Mean	0.7364	30.2988	3.1733	49.5819	3.8005
St.Dev.	4.8223	23.8968	0.6771	24.9146	0.4486
Skewness	-0.4892	2.0887	0.2610	1.8729	0.1831
Kurtosis	2.9788	8.2242	2.8787	7.7506	3.5631
Min	-14.4500	4.6814	1.5436	12.8677	2.5547
5pct Qntl.	-9.2027	7.7864	2.0523	23.5181	3.1577
25pct Qntl.	-2.2050	15.2509	2.7246	34.5176	3.5414
50pct Qntl.	1.0200	23.3539	3.1508	43.1947	3.7657
75pct Qntl.	5.0125	34.8895	3.5522	57.8429	4.0577
95pct Qntl.	8.1440	97.4140	4.5790	95.9318	4.5599
Max	9.7400	143.9771	4.9697	161.4515	5.0867
ACF1	-0.0061	0.4745	0.5697	0.6936	0.7232
ACF2	-0.0449	0.1987	0.3697	0.3915	0.5219
ACF3	0.0593	0.1560	0.2817	0.2139	0.3701
ACF4	-0.0829	-0.0243	0.1317	0.1654	0.3074
ACF5	0.0926	-0.0033	0.1376	0.1576	0.2862
ACF6	0.0741	0.0528	0.1280	0.1184	0.2246
ACF7	0.0910	-0.0771	0.0157	0.0642	0.1620
ACF8	0.0683	-0.0046	0.0589	0.0705	0.1737
ACF9	0.0578	0.0218	0.0828	0.0919	0.1856
ACF10	0.1029	0.0328	-0.0197	0.1738	0.1999

Table 5.5: Summary statistics of monthly return and volatilities for Dow Jones Indexes (DJIA) during the period of October 1997 through December 2009 (T=147)

	x_t	RV	ln(RV)	MFIV1	ln(MFIV1)
Mean	0.1850	35.1168	3.0360	44.4316	3.5142
St.Dev.	4.7276	55.9198	0.9406	40.1199	0.7419
Skewness	-0.7627	5.5570	0.5403	3.4240	0.1076
Kurtosis	4.3800	43.0839	3.4074	20.5580	2.7640
Min	-16.4073	3.2604	1.1818	7.8732	2.0635
5pct Qntl.	-7.9446	5.0522	1.6194	10.2946	2.3316
25pct Qntl.	-1.9571	11.3759	2.4315	18.6377	2.9252
50pct Qntl.	0.6963	20.2984	3.0105	37.5240	3.6250
75pct Qntl.	2.9594	36.8427	3.6066	52.3964	3.9588
95pct Qntl.	7.5783	1117.6689	4.7675	115.9765	4.7534
Max	10.0792	519.4188	6.2527	324.4800	5.7822
ACF1	0.0847	0.6587	0.7547	0.7523	0.8906
ACF2	-0.0971	0.3646	0.6321	0.5485	0.8012
ACF3	0.0552	0.2415	0.5568	0.4739	0.7356
ACF4	0.1294	0.2008	0.4953	0.4071	0.7042
ACF5	-0.0101	0.2046	0.4675	0.3149	0.6635
ACF6	-0.0814	0.1044	0.4201	0.2174	0.6073
ACF7	0.0450	0.0499	0.3582	0.1649	0.5760
ACF8	0.0340	0.0285	0.3482	0.1457	0.5411
ACF9	-0.0636	0.0133	0.3159	0.1456	0.5261
ACF10	-0.1261	0.0082	0.2748	0.1535	0.5149

Table 5.6: QML and C-AMLE Estimates using Daily RV

	α^*	β^*	ρ	σ	μ^*
S & P 500 Daily RV					
QML	1.0981 (0.1934)	0.1269 (0.0312)	-0.0820 (0.0383)	0.7266 (0.0245)	0.0064 (0.0258)
C-AMLE	1.9831 (0.2639)	0.2333 (0.0608)	0.0551 (0.0239)	0.9569 (0.0370)	0.0644 (0.0443)
Dow Jones Industrial Average Daily RV					
QML	0.9299 (0.3162)	0.1838 (0.0276)	-0.1092 (0.0837)	0.7689 (0.2398)	0.0068 (0.5972)
C-AMLE	1.8735 (0.5361)	0.3486 (0.0876)	-0.1697 (0.0913)	1.1222 (0.0489)	-0.0329 (0.0112)

Table 5.7: Model Moments Calculated From Parameter Estimates Using Daily RV

Method	Process	Mean	Std. Dev.	Skewness	Kurtosis
S & P 500 Daily RV					
QML	Return	0.0064	1.0479	-0.1001	8.4583
	RV	1.0981	1.5114	5.3159	25.812
C-AMLE	Return	0.0644	1.4083	0.0520	5.7562
	RV	1.9831	1.9727	6.0212	24.0219
Dow Jones Industrial Average Daily RV					
QML	Return	0.0068	0.9643	-0.1441	7.9024
	RV	0.9299	1.2229	5.3511	25.1806
C-AMLE	Return	-0.0329	1.3688	-0.1864	5.6018
	RV	1.8735	1.8396	6.0755	24.0838

Table 5.8: QML and C-AMLE Estimates using RV for S& P 500 during the period of January 1996 through December 2009 (monthly, T=168)

	α^*	β^*	ρ	σ	μ^*
QML	20.2404 (7.9640)	0.1148 (0.0544)	-0.4380 (0.0801)	5.2964 (0.4616)	0.8338 (0.0285)
C-AMLE	97.7521 (31.4516)	0.2250 (0.1321)	-0.1901 (0.1263)	6.1978 (0.6632)	0.4236 (0.2351)

Table 5.9: QML and C-AMLE Estimates using MFIV1 (based on VIX) for S& P 500 during the period of January 1996 through December 2009 (monthly, T=168)

	α^*	β^*	ρ	σ	μ^*
QML	24.3402 (15.6586)	0.0405 (0.1741)	-0.2875 (0.1809)	3.0006 (0.5347)	0.5978 (0.8096)
C-AMLE	46.1472 (5.0789)	0.3721 (0.1117)	-0.4573 (0.0586)	4.1397 (0.9088)	0.7436 (0.1169)

Table 5.10: Model Moments Calculated From Different Parameter Estimates (S & P 500 monthly(T=168))

	Method	Process	Mean	Std. Dev.	Skewness	Kurtosis
RV	QML	Return RV	0.6887 29.5615	5.4370 28.0469	-0.8764 1.8975	5.9669 8.4009
RV	C-AMLE	Return RV	0.4574 75.2412	8.6742 73.7350	-0.4668 1.9600	5.9541 8.7622
MFIV1	QML	Return MFIV1	0.5978 24.3402	4.9336 52.1085	-0.6245 4.2817	16.8404 30.4992
MFIV1	C-AMLE	Return MFIV1	0.7436 46.1472	6.7932 32.5984	-0.7003 3.4128	4.4568 30.0340

Table 5.11: QML and C-AMLE Estimates using RV for S& P 500 during the period of January 1996 through May 2004 (monthly, T=101)

	α^*	β^*	ρ	σ	μ^*
QML	29.5615 (15.3617)	0.3795 (0.1344)	-0.4507 (0.3459)	4.4941 (6.6201)	0.6887 (0.7867)
C-AMLE	75.2412 (24.2639)	0.1011 (0.1608)	-0.4825 (0.2239)	3.8224 (0.6370)	0.4574 (0.3443)

Table 5.12: QML and C-AMLE Estimates using MFIV2 for S& P 500 during the period of January 1996 through May 2004 (monthly, T=101)

	α^*	β^*	ρ	σ	μ^*
QML	44.6809 (13.1177)	0.2374 (0.0508)	0.0511 (0.0977)	2.5207 (3.2616)	-0.0298 (0.5450)
C-AMLE	46.3175 (5.0789)	0.5833 (0.1117)	-0.0989 (0.0586)	2.3797 (0.9088)	0.8156 (0.1169)

Table 5.13: Model Moments Calculated From Different Parameter Estimates (S & P 500 monthly, T=101)

	Method	Process	Mean	Std. Dev.	Skewness	Kurtosis
RV	QML	Return RV	0.6887 29.5615	5.4370 28.0469	-0.8764 1.8975	5.9669 8.4009
RV	C-AMLE	Return RV	0.4574 75.2412	8.6742 73.7350	-0.4668 1.9600	5.9541 8.7622
MFIV2	QML	Return MFIV2	-0.0298 44.6809	6.6844 24.4527	0.0401 1.0945	3.8327 4.7971
MFIV2	C-AMLE	Return MFIV2	0.8156 46.3175	6.8057 24.9946	-0.4043 1.6475	3.3705 3.6288

Table 5.14: QML and C-AMLE Estimates using RV of Dow Jones Index during the period of October 1997 through December 2009 (monthly)

	α^*	β^*	ρ	σ	μ^*
QML	5.0671 (10.1261)	0.0761 (0.0270)	-0.4306 (3.9530)	5.4121 (0.1393)	0.6773 (5.3986)
C-AMLE	31.8560 (5.7683)	0.3556 (0.1238)	-0.1552 (0.1091)	5.2224 (0.1701)	0.3719 (0.2432)

Table 5.15: QML and C-AMLE Estimates using MFIV1 (based on VXD) of Dow Jones Index during the period of October 1997 through December 2009 (monthly)

	α^*	β^*	ρ	σ	μ^*
QML	22.3913 (23.0044)	0.0441 (0.0160)	-0.2520 (5.8030)	2.8382 (0.2286)	0.3832 (6.5625)
C-AMLE	69.6347 (4.3312)	0.1010 (0.0892)	-0.1656 (0.0589)	3.1983 (0.8081)	0.2066 (0.1298)

Table 5.16: Model Moments Calculated From Different Parameter Estimates (Dow Jones Index monthly)

	Method	Process	Mean	Std. Dev.	Skewness	Kurtosis
RV	QML	Return RV	0.6773 5.0671	2.2510 31.2276	-2.4442 12.3256	118.5587 230.8813
RV	C-AMLE	Return RV	0.3719 31.8560	5.6441 34.9518	-0.3900 2.1944	6.3302 10.2228
MFIV1	QML	Return MFIV1	0.3832 22.3913	4.7319 45.2219	-0.4669 4.0392	15.2152 27.4731
MFIV1	C-AMLE	Return MFIV1	0.2066 69.6347	8.3447 59.3822	-0.7664 3.8360	5.1177 25.5049

5.7 Bibliography

AIT-SAHALIA, Y., AND R. KIMMEL (2007): "Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Financial Economics*, 83, 413–452.

BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2002): "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society*, B64, 253–280.

BLACK, E., AND M. SCHOLES (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–59.

BOLLERSLEV, T., M. GIBSON, AND H. ZHOU (2010): "Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities," *Journal of Econometrics*, forthcoming, doi: 10.1016/j.jeconom.2010.03.033.

BOLLERSLEV, T., AND H. ZHOU (2002): "Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility," *Journal of Econometrics*, 109, 33–65.

——— (2006): "Volatility Puzzles: A Simple Framework for Gauging Return-Volatility Regressions," *Journal of Econometrics*, 131(1-2), 123–50.

BREEDEN, D. T., AND R. H. LITZENBERGER (1978): "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business*, 51, 621–51.

BRITTEN-JONES, M., AND A. NEUBERGER (2000): "Option Prices, Implied Price Processes, and Stochastic Volatility," *Journal of Finance*, 55, 839–866.

CORRADI, V., AND W. DISTASO (2006): "Semi-Parametric Comparison of Stochastic Volatility Models Using Realized Measures," *Review of Economic Studies*, 73, 635–67.

HESTON, S. L. (1993): "A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327–44.

JIANG, G., AND J. KNIGHT (2010): "ECF Estimation of Markov Models Where the Transition Density is Unknown," *Econometrics Journal*, 13, 245–70.

JIANG, G. J., AND Y. S. TIAN (2005): "The Model-Free Implied Volatility and Its Information Content," *The Review of Financial Studies*, 18(4), 1305–42.

——— (2007): "Extracting Model-Free Volatility from Option Prices: An Examination of the VIX Index," *The Journal of Derivatives*, Spring 2007.

PHILLIPS, P. C. B., AND J. YU (2009): "A Two-Stage Realized Volatility Approach to Estimation of Diffusion Processes with Discrete Data," *Journal of Econometrics*, 150, 139–150.

TODOROV, V. (2009): "Estimation of Continuous-Time Stochastic Volatility Models with Jumps Using High-Frequency Data," *Journal of Econometrics*, 148, 131–48.

Chapter 6

Conclusion

The first essay investigates the estimation of discrete-time SV models when volatility is observed. We study the statistical properties of SV models with lagged inter-temporal and contemporaneous dependencies. The explicit expressions of moments and cross moments of returns as well as those of lag-lead correlations between returns and volatilities are different between two models. Both models deserve attention in the literature. Treating volatility as an observable variable, we apply both FIML and 3SLS approaches and show the estimation is straightforward and computationally easy. We undertake two Monte Carlo experiments to examine the estimation performance of these two traditional methods. The results suggest that if the underlying volatility is observed, both FIML and 3SLS approaches do a reasonable job at recovering the true parameters. The two models both fit data well. On the other hand, if the underlying volatility is unobserved, consequently a volatility proxy is employed in the estimation, we should be very careful in choosing an appropriate volatility proxy such that the measurement error does not spread too much, in this case, using volatility proxy then apply traditional methods in SV models are able to provide good estimation performance.

Realized volatility has been shown to be a consistent, unbiased and highly efficient estimator of the true return volatility, motivated by the accuracy of realized volatility, the second essay focuses on investigating the estimation and forecasting of discrete-time SV models using realized volatility. Using daily realized volatility constructed from high frequency data and applying both FIML and 3SLS approaches, we find the estimators from both methods can produce good finite sample properties. We then examine the one-day-ahead volatility point forecasts. Four different models are considered to examine whether allowing asymmetric relationships between return and volatilities, or modeling the long-memory behavior of volatility would result in an improvement in forecast accuracy. The results indicate that allowing asymmetric behavior and leptokurtosis does not seem to improve point forecasts, while modeling long-memory behavior seems do.

The third essay investigates the estimation performance of the C-AMLE method via a Monte Carlo experiment to the affine continuous-time SV model with volatility observed. We generate volatility process based on its conditional density function and return process applying an efficient approximation of the exact scheme at both daily and monthly frequencies. Evidences show that our simulation of both return and volatility processes are accurate. The results suggest that the C-AMLE does a good job at recovering the true parameters.

The fourth essay examines the estimation of the affine continuous-time SV model via empirical application. The empirical analysis is based on the data of S & P 500 index and Dow Jones Industrial Average indexes. Both realized volatility and model-free implied volatility are employed as a volatility proxy, and the C-AMLE as well as the QML are applied in the estimation. The evidence shows that using daily realized volatility, neither method does a good job. However, when the model-free implied volatility is employed, the estimates are stable, and the moments calculated from the estimates are close to the moments of real time series. In general, the C-AMLE approach outperforms the QML method. When the volatility is unobserved, other estimation methods, such as the generalized method of moments (GMM) can be applied. In the future, we will extend our research of investigating the estimation of the affine continuous-time SV model with volatility observed by comparing the estimates from the C-AMLE and QML using volatility proxies with those from the GMM approach when volatility is unobserved.