MODELING AND FORECASTING LONG RANGE DEPENDENCE IN VOLATILITY

Nan Qu

A Thesis submitted for the degree of Doctor of Philosophy

Department of Econometrics and Business Statistics

Monash University

November 2010

Contents

List of Figures

List of Tables

Abstract

This thesis conducts three exercises on volatility modeling of financial assets. We are essentially interested in the estimation and forecasting of daily volatility, a measure of the strength of price movements over daily intervals. Two of the exercises are in the realm of high frequency data: modeling and forecasting realized volatility which is constructed from intra-day returns. The other exercise is concerned with discrete stochastic volatility modeling using daily returns. The main focus of each exercise is to represent the high degree of volatility persistence, which is an important stylized fact of daily volatility.

In the first exercise, daily realized volatility of the Yen/USD exchange rate is modeled through an autoregressive and moving-average fractionally integrated (ARFIMA) process. We differ from previous studies by averaging across a set of ARFIMA and ARMA models with different orders of autoregressive and moving-average polynomials. The vehicle used to execute this averaging exercise is Bayesian model averaging, through which part of the uncertainty introduced by model selection is integrated out. We examine the practical usefulness of our method by conducting a rolling-sample estimation, and the results indicate the weighted average forecast out-performs that of a single model at long-term horizons by providing smaller mean squared forecast errors.

The second exercise is concerned with Bayesian estimation of a long memory stochastic volatility (SV) model. We use a high-order moving-average process to approximate the fractional integration specified for the latent log volatility. As such, the long memory SV model can be expressed in a state-space form, which facilitates the implementation of

Markov chain Monte Carlo (MCMC) simulation when parameters and latent volatility are estimated. We update the set of memory parameter and volatility of volatility parameter in one block in the MCMC algorithm, by using the hessian matrix. A Monte Carlo study indicates in general, when the posterior mean is treated as a point estimator of parameters, our Bayesian method compares well with classical methods. Furthermore, the Bayesian estimator tends to outperform the popular frequency quasi maximum likelihood estimator, according to the root mean square error criterion, with small and medium sample size. An empirical analysis of the daily Yen/USD exchange rate spanning 26 years is conducted, and the degree of persistency in volatility is found to be consistent with that from the first exercise when high frequency data are used.

In the third exercise, we look at the long memory property from a different angle. There has been a large literature using specifications other than fractional integration to mimic the long memory property in time series analysis, although there are few applications to realized volatility. In this exercise, regime switching models are fitted to daily realized volatility of the JPY/USD exchange rate from 1996 to 2009. Both in-sample fit and out-of-sample forecasting are used to compare across the three types of models, including ARFIMA, regime switching and sum of short memory processes. An extensive recursive estimation over one year suggests that regime switching is superior in capturing the dynamics of the time series examined, and generating more accurate out-of-sample forecasts.

Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any university or equivalent institution, and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Nan Qu

Acknowledgements

I would like to thank Brett Inder for his careful supervision during my period as a graduate student at Monash University, Clayton. Without his encouragement, I could not have been able to continue my thesis. I also would like to thank Param Silvapulle, who introduced me to the study of Econometrics at Master's level; Mervyn Silvapulle for teaching me maximum likelihood estimation; Don Poskitt for teaching me generalized method of moments, and Catherine Forbes for teaching me about state space models. Thanks also are due to Rob Hyndman and Robert Brooks for responding to my email queries; and to Kathryn Cornwell and Barry Goss for helping me to improve English and taking the pains to correct my written English. In addition, I want to express my appreciation to Lizbeth Moon for her excellent support in providing study facilities, and to Audrey Fernando for her help with accommodation.

I would also like to thank Farah Yasmeen, Rong Zhang and Bing Li, my fellow graduate students, with whom I have shared so many joyful and painful moments through the completion of my thesis.

This thesis is dedicated to my husband Xiao Lei Li, my parents Nan Peiru and Qu Zhucheng, and my brother Qu Bin for their unconditional love and generous support from every perspective.

Chapter 1

Introduction

Modeling and forecasting volatility of asset returns is a crucial aspect of modern financial theory. The literature in volatility modeling is voluminous. There are numerous modeling approaches available based on different assumptions about the underlying data generating process. In general, the popular GARCH and stochastic volatility (SV) models treat volatility as a latent process; this methodology is naturally related to the Mixture of Distributions Hypothesis dating back to Clark (1973). Both GARCH and SV models have been extensively examined theoretically and successfully applied in practice since the 1980's. The latency of volatility introduces some difficulties in estimation, especially for SV models, since one cannot observe the realization of volatility but its approximation. In recent years, attention has shifted to incorporating high-frequency data into volatility modeling following the advent of highly informative intraday data. GARCH and SV models are capable of dealing with low frequency data, such as daily or weekly data, although when facing tick-by-tick data, they are too restrictive to capture the rich dynamic implied by intraday returns. In particular, various market microstructure effects render the straight implementation of either approach cumbersome if not infeasible. One of the popular responses in the recent literature is to construct a consistent volatility estimator, i.e. realized volatility, by cumulating squared intraday returns. According to the well-established quadratic theory and empirical evidence, an appealing feature of realized volatility is that it can be treated as observable rather than latent, and modeled in a sim-

ilar way to economic variables. This allows a wide array of time series analytical methods to be applied to realized volatility, for example, the well-developed ARMA models, or nonlinear specifications.

This thesis conducts three exercises on volatility modeling, with two of them in the framework of realized volatility (Chapters 3 and 5) and the other in the discipline of SV models (Chapter 4). We focus on the long memory property of volatility, which is a well-recognized stylized fact of returns, with consistent evidence obtained from both low-frequency and high-frequency data. A good understanding of volatility persistence is critical in obtaining successful volatility forecasts.

In Chapter 3 we conduct a Bayesian model averaging exercise on daily realized volatility of the Yen/USD exchange rate. Our exercise differs from previous studies in the aspect of model selection. We choose to deal with an *overall averaged* model rather than a single model with predetermined orders of autoregressive and moving average terms. In realized volatility modeling, a common practice is to select a single model from the family of autoregressive fractionally integrated moving-average (ARFIMA) models. While this is computationally convenient, an inappropriate choice of the order of autoregressive and moving-average terms can result in a biased estimate of the memory parameter. To avoid this bias, it is desirable to take account of the uncertainty in model selection.

In terms of methodology, Bayesian model averaging (BMA) is used due to its inherent formality and coherency. The implementation of BMA is straightforward once one obtains Bayesian estimates of each model. This is because the marginal likelihood, the criterion for Bayesian model selection, is a byproduct of estimation for each single model. Posterior model probabilities, a summary of data evidence in support of each model specification, are used to determine the weight allocated to each model. Statistical inference for the memory parameter and out-of-sample predictions are the weighted averages over the corresponding quantities from the individual models considered. Models to be averaged across include both long memory and short memory specifications, namely, ARFIMA and ARMA models with different orders of autoregressive and moving-average terms. It is clear that the overall model nests single models, so a more general specification is dealt with, allowing robust

statistical inference, because uncertainty due to single model selection, is integrated out in a consistent way. With the empirical exercise conducted, we address the question whether additional computational cost of dealing with a number of models rather than a single model, can be justified. This question is of practical relevance and is addressed from two aspects: posterior inference on the memory parameter, and forecasting performance evaluated at different horizons. First, a simulation study, with various data generating processes, is considered, and second comparisons of forecasting performance are made between BMA and single models. Both aspects support the usefulness of Bayesian model averaging.

Our second exercise on realized volatility is conducted in Chapter 5, where we experiment with a nonlinear time-series specification, regime-switching model, to capture the high degree of volatility persistence. The conventional method of modeling long memory property is via fractional integration: e.g., fractional integrated GARCH (Baillie, Bollerslev and Mikkelsen (1996)), long memory SV models (Harvey (1998) and Breidt, Crato and deLima (1998)), and the ARFIMA specifications used in realized volatility (Andersen, Bollerslev, Diebold and Labys (2001, 2003)). We propose to use alternative specifications of modeling realized volatility, motivated by the following observation. There is a large literature in applying nonlinear models to mimic the long memory property in GARCH and SV models, and in some applications both in-sample fit and out-of-sample forecasting appear promising. Most of these applications are on daily squared returns or daily absolute returns, which were popular volatility proxies prior to the introduction of realized volatility. Commonly used alternative models include structural breaks, regime switching and sum of short memory processes. However, there have been few applications of these alternatives applied to realized volatility. To our knowledge, one exception is Barndorff-Nielsen and Shephard (2002a) who propose a multi-factor SV specification to capture the long memory property in realized volatility. We argue that since realized volatility is regarded as a superior proxy for latent volatility compared with squared or absolute daily returns, we expect these promising results obtained from low frequency data to hold when high frequency data are examined.

Our empirical exercise on daily realized volatility of the Yen/USD from 1996 to 2009

provides clear evidence of the superior performance of regime switching specifications, in terms of in-sample fit and out-of-sample forecasting. We conduct an extensive recursive estimation based on one year rolling samples. Three types of models are considered: ARFIMA, regime switching and multi-factor SV models. There are few comparisons of these alternatives when applied to realized volatility in the literature. Our recursive estimations indicate advantages of regime switching over ARFIMA models, especially according to residual diagnostics and long-term forecasting. We argue this might be due to the flexibility implied by regime switching, by allowing level and innovation variance i.e. volatility of volatility, to switch between regimes. Given a time series spanning over a decade, it might be too restrictive to assume volatility dynamics are fully characterized by a set of constant memory, autoregressive and moving average parameters, as in the specification of fractional integration.

In Chapter 4 we propose a Bayesian estimator of long memory SV models. There is a large literature in the estimation of long memory SV models by classical methods, although few Bayesian estimators have been proposed. Our aim is to compare the performance of the proposed Bayesian estimators with that of classical counterparts. Volatility is treated as genuinely latent in SV models; this assumption is appealing in theory but renders the likelihood evaluation difficult, especially for classical methods. The SV model is one of the most successful areas of application for Bayesian methods, because it allows volatility to be treated as a parameter via data augmentation. Moreover, posterior simulation methods such as Markov chain Monte Carlo (MCMC) simulation facilitate the availability of posterior volatility estimates by integrating out the uncertainty of other parameters. Volatility estimates obtained by the classical method are conditional on the point estimates of other parameters, and this implies additional uncertainty. In comparison, Bayesian estimation is based on a state space form of SV model, with fractional Gaussian noise being approximated by a moving average process of high order. We propose to update the parameters of interest in one block to improve estimation efficiency, because a block updating scheme is helpful in variance reduction. In particular, a hessian matrix, obtained from quasi maximum likelihood estimation, is used to construct the proposal density for the random walk Metropolis-Hastings algorithm when MCMC is implemented.

A Monte Carlo simulation study is conducted in Chapter 4 and the result suggests the performance of the proposed estimator is comparable to that of classical methods. Our estimator outperforms the classical counterpart in terms of root mean square error for most of the parameter settings, with small and medium size samples. We argue that the parameter settings of the Monte Carlo study are of more practical relevance compared with previous studies: this is because variance of innovations to volatility are chosen at values which are empirically realistic. A sensitivity analysis is conducted to examine the impact of parameter values on estimation, the result of which is useful for empirical applications of long memory SV model.

The rest of this thesis is organized as follows: Chapter 2 reviews some important questions in volatility modeling, including the Mixture-of-Distributions Hypothesis, stylized facts of returns with an emphasis on the long memory property, and the basic framework of SV models and realized volatility modeling. The questions reviewed are selective since our aim is to provide the basic background relevant to the chapters which follow. A Bayesian model averaging exercise is conducted on realized volatility in Chapter 3. Chapter 4 proposes a Bayesian estimator of a long memory SV model, while Chapter 5 seeks to use the specification of regime switching to capture the long memory property in realized volatility. Chapter 6 concludes.

Chapter 2

Literature Review

2.1 Introduction

Statistical analysis of asset prices is difficult given the observation that asset prices are non-stationary as manifested by increased variance of prices over time. It is of more practical relevance to examine asset returns, i.e. changes in prices, because much empirical evidence suggests that returns can be modeled by stationary processes. Asset returns are commonly measured and modeled with fixed regular frequency, such as monthly, weekly and daily. For returns over short intervals such as daily, the first moment is not of primary interest in most cases, as within a frictionless continuous time framework the no-arbitrage requirement generally guarantees that the return innovations are of an order of magnitude larger than the mean return. Consequently, the second moment of returns, i.e. return volatility, is usually the focus. Volatility is an important economic variable: for example, the original ARCH model of Engle (1982) provides a tool for measuring the dynamics of inflation uncertainty. Furthermore, most of the important developments and applications in volatility modeling and forecasting have occurred in financial economics.

Our focus in this review is on volatility of daily returns, and return volatility is defined as the standard deviation of returns. Conceptually, volatility measures the speed at which prices are changing, i.e. the intensity of return variation. A small standard deviation implies a low chance of a large price movement, and in particular, a small likelihood

of a large price fall. The risks from two alternative investments can be contrasted by comparing their standard deviations. Indeed volatility is commonly treated as a measure of risk, and risk management is a crucial aspect of modern financial econometrics, as noted by Campbell, Lo and MacKinlay (1997, p. 3):

> "... what distinguishes financial economics is the central role that uncertainty plays in both financial theory and its empirical implementation. ... Indeed in the absence of uncertainty, the problems of financial economics reduce to exercises in basic microeconomics. The very existence of financial economics as a discipline is predicated on uncertainty".

All this highlights the importance of volatility modeling. In particular, according to Shephard (1996), there are two motivations for return volatility models: empirical stylized facts and the pricing of contingent assets. Volatility is a key input to the pricing of derivative securities, for such markets are sometimes characterized as "volatilities are traded". Return volatility is time-varying as conjectured by finance theory and acknowledged in empirical evidence.[1] The challenging aspect of volatility modeling is that it is inherently latent. According to Andersen, Bollerslev, Christoffersen and Diebold (2006): "This departure of finance from standard microeconomics is even more striking once one recognizes that volatility is inherently unobserved, or latent, and evolves stochastically through time".

The search for volatility model specification and selection is always guided by the salient feature of volatility being latent and time-varying. There is a voluminous literature on return volatility modeling, so the review in this chapter cannot be exhaustive. The literature on asset volatility and risk management has grown considerably in the last 20 years: Bollerslev, Chou and Kroner (1992), Bollerslev, Engle and Nelson (1994), Hamilton (1994) and Engle (1995)[2] provide comprehensive surveys on ARCH models; Bates (1995) covers empirical regularities regarding derivative securities and implied volatilities;

[1] The Mixture of Distributions Hypothesis, dating back to Clark (1973), attributes the variability of speculative returns to the intensity of relevant news arrivals. Early evidence on fluctuations in return standard deviations can be found in Merton (1980), and those references given in Taylor (1986, p. 38) on various assets including stocks and futures. It is a well-recognized fact that financial asset return volatility is time-varying, the evidence of which is persistent across assets, time periods and countries. We will discuss more on these in Section 2.2 and 2.3.

[2] The book of Engle (1995) is an extensive collection of GARCH papers.

Taylor (1994), Ghysels, Harvey and Renault (1996) and Shephard (1996, 2005) are excellent surveys on stochastic volatility models; Poon and Granger (2003) and the book length treatment of Knight and Satchell (2007) are concerned with volatility forecasting; Andersen et al. (2006) and Andersen, Bollerslev, Christoffersen and Diebold (2007) are also concerned with volatility forecasting in the role of risk management but with an emphasis on high frequency data. Some recent surveys discuss volatility in a rather general and broad context, for example, Barndorff-Nielsen and Shephard (2007), Shephard and Andersen (2009) and Andersen, Bollerslev and Diebold (2010). The surveys of Andersen and Benzoni (2009) and McAleer and Medeiros (2008) are mainly concerned with realized volatility.

This chapter concerns the basic concepts of volatility modeling, relevant to Chapters 3, 4 and 5. The treatment in this thesis is exclusively univariate, although a multivariate framework, involving time-varying volatility and also time-varying covariance, plays an important role in analyzing returns of a portfolio.[3] The rest of this review is organized as follows: Section 2.2 discusses the Mixture-of-Distributions Hypothesis (MDH) as it provides a convenient theoretical explanation for changing levels of return volatility, which has motivated the formulation of some of the most widely used empirical discrete-time volatility models including ARCH/GARCH and stochastic volatility models. Section 2.3 focuses on stylized facts of daily returns, as a useful volatility model needs to accommodate these empirical facts, especially when a key objective of volatility modeling is to generate accurate volatility forecasts. Section 2.4 and 2.5 review the developments of two popular strands in volatility modeling: the stochastic volatility model and realized volatility. The former is the building block for the long memory stochastic volatility model considered in Chapter 4, and the latter is the focus of Chapters 3 and 5. Section 2.6 concludes.

[3] According to Shephard (1996, p. 41): "Most of macro-economics and finance is about how variables interact, which, for multivariate volatility models, means it is important to capture changing cross-covariance patterns. Multivariate modeling of means is difficult and rather new: constructing multivariate models of covariance is much harder, dogged by extreme problems of lack of parsimony".

2.2 Mixture-of-Distributions Hypothesis

The Mixture-of-Distributions Hypothesis (MDH), pioneered by Clark (1973), is a way to conceptualize the distributional characteristics of asset returns. Let y_t denote financial returns, defined over a unit time interval, for example, daily.[4] Under this assumption, returns are governed by an event time process which can be formulated as

$$y_t = \mu_y s_t + \sigma_y s_t^{1/2} \varepsilon_t \tag{2.1}$$

where s_t is a strictly positive process reflecting intensity of the information flow in the market, μ_y represents the mean response of the variable per news event, σ_y is a scale parameter and ε_t is assumed to be *i.i.d.* $N(0,1)$. Furthermore, s_t is assumed to be unobservable to the econometrician, and σ_y is constrained at unity to ensure the identifiability of the latent process s_t.

It is clear that returns on a given trading day are normally distributed conditioning on s_t since $y_t|s_t \sim N(\mu_y s_t, \sigma_y^2 s_t)$. Consequently, the unconditional distribution of returns is a mixture of normals. According to Praetz (1972), an i.i.d. mixture of a Gaussian term and an inverted Gamma distribution for the variance will produce Student's t distributed returns. Student's t distribution has fatter tails than normal distribution. As such, MDH is compatible with the fat-tailed unconditional distributions found for finance returns. It is a common observation that the normality assumption is violated in the marginal distributions of daily returns.[5] Empirical evidence on heavy tails of the marginal distributions of asset returns dates back to the early 1960s, see Mandelbrot (1963) and Fama (1963). This feature of *leptokurtosis* has important implications for risk management, as thick-tailed distributions indicate frequent occurrence of returns with extreme magnitudes.

When returns are measured over short horizons such as daily or weekly, μ_y is negligible

[4] Suppose closing prices are recorded on each trading day, let p_t be the price recorded on trading day t (ignoring the dividend.) Throughout, the compound return is defined as $y_t = \log(p_t) - \log(p_{t-1})$. See Taylor (1986, Ch. 1) for discussions on other types of return definitions used in the literature. Another commonly used definition, simple return, is defined as $y_t' = (p_t - p_{t-1})/p_{t-1}$. As discussed in Taylor (1986, pp. 12-13), researchers who have studied both the compound return and simple return find that the important conclusions are the same for each type of return.

[5] Returns over a month or more have unconditional distributions much closer to the normal shape than daily returns.

or can be fixed at a constant level. Thus, the driving force of y_t is the *state variable* s_t. The state variable represents the number of news arrivals. Intuitively, days with high information flow imply high likelihood of large price moments, so s_t indicates the strength of return variation. Furthermore, if s_t is assumed to be a positively autocorrelated process, by allowing for more realistic temporal dependencies in the underlying information arrival, the formulation in (2.1) is able to mimic observed volatility clustering in financial markets. This last feature is another important stylized fact of returns, i.e. the tendency for large returns to be followed by future returns of large magnitude, and small returns to be followed by additional small ones.

The Mixture-of-Distributions Hypothesis is the foundation of the most widely used empirical discrete-time volatility models. Andersen et al. (2006, pp. 815-816) provides an excellent intuitive introduction to the Mixture-of-Distributions Hypothesis. Different assumptions on the dynamics of s_t result in various strands of volatility models. For GARCH models, s_t is assumed to be fully determined by information up to time t, while stochastic volatility models assume s_t to be genuinely latent. Recently, Andersen, Bollerslev, Diebold and Labys (2000*a*) and Thomakos and Wang (2003) find that daily returns standardized by realized volatility are closer to normal than the standardized residuals from stochastic volatility models, suggesting realized volatility can estimate s_t more accurately. There is a large literature in seeking an alternative "mixing variable" to represent s_t. Clark (1973) observed that trading volume is highly correlated with return volatility and suggests using volume as a good proxy for s_t. The papers of Epps and Epps (1976), Taylor (1982), Tauchen and Pitts (1983), Andersen (1996) and Andersen and Bollerslev (1997*a*), are among many others that address this issue.

2.3 Stylized Facts of Assets Returns

An important motivation of modeling return volatility is to capture the stylized facts of returns. As noted in Ghysels et al. (1996, p. 7): "The search for model specification and selection is always guided by empirical stylized facts. A model's ability to reproduce such stylized facts is a desirable feature and failure to do so is most often a criterion to dismiss

a specification although one typically does not try to fit or explain all possible empirical regularities at once with a single model". This section will briefly discuss some of the most important stylized facts of asset returns, in particular, the slowly-decaying autocorrelations observed in absolute or squared daily returns, as this so-called long memory property is a major focus in the following chapters. Stylized facts about volatility have been extensively reviewed in the literature. See for instance, Taylor (1986, Ch. 2), Bollerslev et al. (1994) which is under a general set-up of GARCH models. Excellent discussions within the framework of stochastic volatility models can be found in Shephard (1996) and Ghysels et al. (1996). Bates (1995) provides a detailed discussion on stylized facts of volatility observed in derivative markets.

The aspects to be discussed in this section include: leptokurtosis, i.e. thick-tailed distributions of returns and time-varying standard deviation, the dynamics of which feature as volatility clustering and highly persistent shocks to return innovations. All these characteristics are indeed closely related: the time-varying dynamics render an unconditional distribution of returns made up of mixture of normals with varying standard deviation, and the mixing causes the higher kurtosis. We intend to ignore other important stylized facts in this section, mainly because they are not relevant to the following chapters. Ghysels et al. (1996, section 2.2) provides a detailed discussion on other stylized facts, such as volatility comovements, which is useful in examining international comovements of speculative markets, and volatility smiles relevant to option pricing.

2.3.1 High Kurtosis

The evidence on leptokurtosis (heavy-tailed return distributions) was documented by Mandelbrot (1963) and Fama (1963). An explanation of this fact is the varying standard deviation of returns. As noted in Taylor (1986, p. 47), there is no reason to suppose that the number of price changes within each day is identical. Neither is it particularly reasonable to assume that every change in price has the same variance. The literature prior to the conditionally heteroscedastic models draws from fat-tailed distributions such as Paretian or Lévy without addressing the conditionality of volatility. Under the framework

of GARCH and stochastic volatility models, the feature of leptokurtosis fits nicely into the modeling of conditional variance of returns. To illustrate, suppose the mean of daily returns is negligible, and daily return y_t is formulated as

$$y_t = \sigma_t \varepsilon_t \qquad \varepsilon_t \sim i.i.d.\ N(0,1) \tag{2.2}$$

where σ_t denotes the conditional volatility. Gourieroux and Jasiak (2001, p. 118) relate the thickness of tails to the dispersion of volatility, and shows mathematically that the conditionally heteroscedastic models are capable of reconciling leptokurtosis with the normality assumption by distinguishing between the marginal and conditional kurtosis.[6]

GARCH models specify the conditional variance as a function of past squared or absolute returns, which is an important assumption because if the variance mixture is not linked to observed variables, the i.i.d. mixture is indistinguishable from a standard fat-tailed error distribution (see Andersen et al. (2006, p. 828)). However, empirical applications typically find that the time varying volatility is not sufficient to account for all of the mass in the tails in the distributions of daily or weekly returns. Therefore, Student's t distribution is proposed for ε_t in some applications. Leptokurtosis becomes less pronounced under temporal aggregation, as the departure from normality is commonly found to diminish for returns over a month or more.

2.3.2 Volatility Clustering

It is typical for financial time series to exhibit bunching of high and low volatility episodes. Volatility clustering refers to the feature that large absolute returns are more likely to be followed by large absolute returns than by small absolute returns, the evidence of which was first documented by Fama (1965). This fact is also implied by strong positive autocorrelations found in commonly used volatility approximators, such as squared or absolute daily returns. Taylor (1986) argues that the autocorrelation found in absolute and

[6] Intuitively, the unconditional distribution is a mixture of normals, some with small variances that concentrate mass around the mean and some with large variances that put mass in the tail of the distribution. As such, a mixture with fatter tails is formed. See Campbell et al. (1997, p. 478) for more discussion. A student's t distribution has fatter tails than normal as it is a mixture of normal distributions with different variances. However, it is not capable of capturing the volatility clustering.

squared returns can be explained by time-varying volatility, and that any reasonable model for returns must be non-linear because a linear process cannot explain the autocorrelation in absolute returns. Furthermore, the decaying of autocorrelations appears to be slow, which is why numerous papers have proposed the use of a long memory specification for the evolution of σ_t in equation (2.2).

The explanation for volatility clustering is intimately related to MDH, according to which changes in the level of market activity are the cause of the variances changes. Information arrivals are nonuniform through time, and it is reasonable to assume the intensity of information flow is persistent in consecutive trading days. The huge success of the GARCH group of models is largely due to its ability to mimic volatility clustering and to integrate the conditional volatility dynamic with unconditional leptokurtosis. The stochastic volatility model is a capable alternative although its likelihood functions are more difficult to evaluate, an issue which we will discuss more in Section 2.4.[7] Volatility clustering can be identified by a formal test such as the ARCH Lagrange multiplier test proposed by Engle (1982). It is also widely documented that ARCH effects disappear with temporal aggregation; see, for example Diebold (1988) and Drost and Nijman (1993).

2.3.3 Leverage Effect

Leverage effect, a term coined by Black (1976), refers to the observation that negative innovations to financial returns tend to increase volatility more than positive innovations of the same magnitude. This asymmetry is pronounced in stock market data, as shown by Schwert (1989), Nelson (1991), Campbell and Hentschel (1992), Brock, Lakonishok and LeBaron (1992) and Poon and Taylor (1992). The effect is small for interest rate and

[7] Various dynamics of conditional volatility movements can be explained by changing levels of information flow: numerous papers documented increased volatility of financial markets around dividend announcements and macroeconomic data releases; day-of-the-week effects might introduce extra complicity to the dynamics of time-varying volatility, see for example Baillie and Bollerslev (1989) report overnight and weekend market closures and their effect on volatility. When returns are examined on high frequency basis, more complicated dynamics of conditional variance are introduced. A good example is foreign exchange markets, where major currencies are traded 24 hours and different time zones can produce strong seasonal patterns in around the clock trading activity. See Müller, Dacorogna, Olsen, Pictet, Schwarz and Morgenegg (1990), Baillie and Bollerslev (1990), Harvey and Huang (1991), Dacorogna, Müller, Nagler, Olsen and Pictet (1993), Andersen and Bollerslev (1997*b*) and Bollerslev and Ghysels (1996) for detailed discussions on the intradaily patterns of foreign exchange markets.

exchange rate series, and Shephard (1996) notes that it is non-standard for currencies where the asymmetry effects are usually not significant. Taylor (1994) argues there are plausible theories for a negative correlation between price movements and volatility in stock models but none for a nonzero correlation in currency models. See also the discussion in Ghysels et al. (1996, p. 11).

Andersen et al. (2006, p. 803) provide a discussion of various explanations for leverage effect. For example, falling stock prices imply an increased leverage of firms and it is believed that this entails more uncertainty and hence volatility. This asymmetry effect has motivated many refinements to the basic GARCH and stochastic volatility models: EGARCH (Nelson (1991)), GJR (Glosten, Jagannathan and Runkle (1993)) and TGARCH (Zakoian and Rabemananjara (1993)) and asymmetric stochastic volatility models are the most well known among many others. However, leverage effect complicates the form of likelihood functions by linking the first moment to the second moment of the conditional return distribution. Consequently, it is more difficult to undertake parameter estimation and volatility prediction, especially for stochastic volatility models.

2.3.4 Long Memory

At the intuitive level, a characteristic of a long memory time series process is a hyperbolic rate of decay for the autocorrelations instead of an exponential rate as implied by a stationary autoregressive process. To illustrate, a simulated long memory process in Harvey (1998) indicates the slowly-decaying pattern: at a lag as high as 100, the long memory autocorrelation is still 0.14, whereas in the autoregressive case it is only 0.000013. Considerable empirical evidence indicates volatility autocorrelations decay at a rapid rate for shorter lags, but then at a much slower hyperbolic rate at longer lags, which matches closely to the theoretical shape of autocorrelations of long memory processes. The evidence tends to be stronger when volatility is measured with daily and weekly returns recorded over decades, or when measured using high-frequency data over a few years.

2.3.4.1 Empirical Evidence

Early evidence is documented in Ding, Granger and Engle (1993), who observe that for daily S&P500 return series, power transformations of a volatility proxy, absolute daily returns, $|y_t|^d$ have quite high autocorrelations for long lags, and strongest temporal dependence is observed with d close to one. Numerous papers afterwards observed similar patterns for other stock market indices, commodity markets and foreign exchange series; see for example, Baillie et al. (1996) and Andersen and Bollerslev (1997*a*). Many estimates of GARCH models for these series suggest near unit root behavior of the conditional variance process.[8] Estimates of stochastic volatility models show similar persistent patterns in conditional variance. In recent years, consistent evidence has also been observed from realized volatility, constructed from intraday returns. Correlogram plots for logarithmic daily realized volatility series show a distinct hyperbolic decay, e.g. Figure 11.5 in Andersen, Bollerslev, Christoffersen and Diebold (2007, p. 529), where the autocorrelations are significantly positive for all 100 lags when compared with the conventional 95 percent Bartlett confidence intervals. See also Andersen and Benzoni (2009, p. 567) for a review of evidence across different assets.

In other words, volatility tends to change slowly, implying that the effects of shocks take a considerable time to decay. This has important implications in volatility forecasting. Bollerslev and Mikkelsen (1999) for example, observe that when pricing very long-lived financial contracts, the long memory volatility approach can result in materially different prices from those implied by standard GARCH models with exponential decay. The most common approach to modeling long memory is the fractional integrated process, and we provide a brief review of this and other approaches in the following subsection.

[8]This observation motivates the integrated GARCH (IGARCH) model proposed by Engle and Bollerslev (1986). According to Ghysels et al. (1996, p. 9), there has been a debate regarding modeling persistence in the conditional variance process either via a unit root or a long memory process. One of the appealing features of the latter approach is that its inference is not affected by the kind of unit root issues. For example, a likelihood ratio test of stationarity can be constructed by standard theory, see Robinson (1994).

2.3.4.2 Approaches to Capture the Long Memory Property of Volatility

Fractional integration There is a long history in time series analysis using fractional integration to capture slowly-decaying autocorrelations; see Chapter 3 Section 3.2 for detailed discussions on definition and estimation of fractionally integrated processes. The implementation of fractional integration to GARCH models was first proposed by Baillie et al. (1996), as FIGARCH, which adapts autoregressive fractionally integrated moving average (ARFIMA) models developed by Granger and Joyeux (1980) and Hosking (1981) to long-memory models of conditional variance.[9] Another popular long memory GARCH model is the FIEGARCH proposed by Bollerslev and Mikkelsen (1996), which is a generalization of the EGARCH model of Nelson (1991). The FIEGARCH model treats the logarithm of the conditional variance as a distributed lag of past ε_t^2 involving the fractional difference operator. In parallel to the development within the GARCH framework, Breidt et al. (1998) and Harvey (1998) proposed the long memory stochastic volatility model, where log volatility is assumed to follow a fractionally integrated process; see Chapter 4 for detailed discussions on long memory stochastic volatility models. In a continuous-time framework, Comte and Renault (1998) propose a continuous-time stochastic volatility model with discrete-time long-memory implications. The continuous time set-up has the advantage that it can be used for options pricing without excessive computational effort. Although intuitively appealing, estimation of long memory volatility models is not straightforward, especially for long memory stochastic volatility, where volatility is genuinely latent.[10] Breidt et al. (1998) and Harvey (1998) use quasi maximum likelihood methods to estimate long memory stochastic volatility models, based on a frequency domain representation of the likelihood function, which is computationally simple but not efficient. Also, implementation of fractional integration renders the *smoothing* and *filter-*

[9] Andersen et al. (2006, p. 805) examine the impulse effect of a time-t shock on the forecast of the variance h period into the future implied by the GARCH (1,1) model: $\partial\sigma^2_{t+h|t}/\partial\varepsilon_t^2 = k\delta^h$ with $0 < \delta < 1$. Whereas, to have autocorrelations decay at a hyperbolic rate the volatility model is expected to provide an impulse effect as $\partial\sigma^2_{t+h|t}/\partial\varepsilon_t^2 \approx kh^\delta$ for large values of h. FIGARCH model nests the IGARCH model of Engle and Bollerslev (1986), however one weakness of FIGARCH model is that the unconditional variance does not exist.

[10] Shephard (1996, p. 16) argues that initialization might be problematic in FIGARCH model estimation. In particular, distribution function $f(y_t|y_{t-h}, \ldots, y_{t-2}, y_{t-1})$ has to contain a large amount of relevant data due to the highly persistent effect of past observations. Meanwhile, likelihood functions of GARCH models are constructed conditional on y_0. As such, initialization of y_0 is important, although Baillie et al. (1996) argue this is not the case.

ing of volatility computationally more difficult than their short memory counterparts, due to the length of the state vector.[11]

In recent years, with the advent of high frequency data, numerous applications treat realized volatility as an observed proxy for the underlying volatility (see Section 2.5 for more discussions on realized volatility). In this framework, Andersen et al. (2000*a*, 2003) specify a fractionally integrated model to account for the apparent long-memory observed in daily realized volatility.[12]

The estimates of the degree of fractional integration, i.e. the memory parameter, usually take a value in the 0.3-0.48 range, appearing to be stable irrespective of the sampling frequencies of the underlying returns (daily or intra-daily), the sampling periods or asset classes: see Andersen and Bollerslev (1997*a*). This finding is consistent with the assumption that volatility is a stationary but highly persistent process.

Alternative approaches to modeling long memory in volatility There exists an extensive literature in using alternative methodologies to capture slowly-decaying autocorrelations. This remains a very active area of current research: see Andersen et al. (2006, p. 813) and McAleer and Medeiros (2008, pp. 36-37), and the references therein.

In general, two alternatives are available. Firstly, it is possible to approximate the long memory feature by specifying a volatility process via a sum of short memory processes such as first-order autoregressive components: e.g. the multi-factor stochastic volatility model of Chernov, Gallant, Ghysels and Tauchen (2003), [13] the model in Ding and Granger (1996) which is based on the sum of an infinite number of ARCH models, and the component GARCH model in Engle and Lee (1999), as well as the related developments in Gallant,

[11] See Section 2.4 for volatility filtering and smoothing for stochastic volatility models, and Section 4.2.3.1 in Chapter 4 for state space representation of long memory stochastic volatility model.

[12] The logarithmic transformation has a few appealing features when modeling realized volatility: the log specification guarantees positive forecast volatilities. In addition, the unconditional distribution of log realized volatility can be better approximated by Gaussian and is much closer to being homoscedastic.

[13] Andersen et al. (2010, p. 29) discuss one-factor and multi-factor parametric volatility models. Empirical evidence suggests that multi-factor models provide major improvements over the traditional one-factor models. They attribute the advantage of multi-factor models to the fact that the system is allowed to be defined through a sum of different factors, each following a simple dynamic process, rather than a single factor with a more complex specification.

Hsu and Tauchen (1999).[14] In more recent work, Barndorff-Nielsen and Shephard (2001), Bollerslev and Wright (2001), Andersen, Bollerslev and Diebold (2007) and Corsi (2009) apply the same methodology to realized volatility modeling. Secondly, a possible source of long memory in financial volatility is infrequent breaks, as discussed in Hyung, Poon and Granger (2006), or switching between regimes and structural breaks as shown in Diebold and Inoue (2001). Structural breaks in volatility are also considered, for example by Schwert (1989) and Andreou and Ghysels (2002), while connections between long memory and structural breaks are examined by Banerjee and Urga (2005) and Andreou and Ghysels (2009) (see Chapter 5 Section 5.1 for further discussion).

2.4 Stochastic Volatility Models

GARCH and stochastic volatility (SV) models, referred to as the second generation of volatility models by Shephard and Andersen (2009), are parametric models for volatility. In the literature prior to GARCH models, rolling window methods were used to estimate time-varying volatility. For instance, Officer (1973) used a rolling standard deviation of returns measured over some interval of time; Parkinson (1980) and Garman and Klass (1980) used the difference between daily high and low prices as daily volatility estimates. The weakness of these methods is that they are not based on formal statistical models, and the assumption of constant volatility over a rolling window period is counterfactual: see the example discussed in Andersen, Bollerslev, Christoffersen and Diebold (2007, p. 516), which demonstrates the poor forecasting performance of these "historical" volatility measures. The same example also indicates the weakness of J.P. Morgan's RiskMetrics (Morgan (1997)) due to its lack of volatility mean-reversion structure, although RiskMetrics is frequently used by finance practitioners.

[14] The plots of volatility impulse response functions in Andersen et al. (2006, p. 807) indicate the ability of component GARCH model to match the hyperbolic decay rate for the long-memory FIGARCH model. They argue: "if the different components display strong, but varying, degrees of persistence they may combine to produce a volatility dependence structure that is indistinguishable from long memory over even relatively long horizons". As noted in Andersen et al. (2006), there is a long history in time series analysis for using the sum of a few individual short-memory components to mimic a long memory process. Granger (1980) first show that the superposition of an infinite number of stationary autoregressive of order one processes may result in true long-memory process. Tiao and Tsay (1994) provide early results on modeling and forecasting long-run dynamic dependencies in the mean.

GARCH models, which pre-date SV models,[15] have been tremendously successful in empirical work and now become a standard toolbox of finance and econometric modeling. This is because the models are designed to facilitate maximum likelihood estimation. In recent years, SV models have become more popular, essentially because estimation difficulties that previously plagued these models can be alleviated by new approaches, in particular, simulation-based inferences. A further reason for the popularity of SV models is their use in options pricing, where volatility modeling is useful. We will discuss further the differences between the two types of models in the following subsections.

In general, SV models can be cast either in discrete-time or continuous-time. The first centers on the MHD discussed in Section 2.2. The other approach is to incorporate diffusions employed by much of modern finance, where price and volatility processes are modeled jointly through continuous sample diffusions governed by stochastic differential equations.[16] As argued by many researchers, continuous-time SV models are motivated by convenience and intuition rather than by studies of observed prices. The real-life scenario is that prices are invariably constrained to lie on a discrete grid, both in the price and time dimension, even if the underlying process actually evolves continuously within a no-arbitrage setting. (Section 3.4.1 in Chapter 3 and Section 5.3.1 in Chapter 5 describe the raw tick-by-tick data used in this thesis.) In other words, continuously recorded data do not exist as real-time price data are not available at every instant, and it is only feasible to measure return and volatility realizations over discrete time intervals. This is the major motivation of discrete-time modeling. However, the two perspectives are closely related because discrete-time SV models can be treated as approximations (but not exact discretizations) to continuous-time SV models. For example, the initial discrete-time SV model introduced by Taylor (1982) can be thought of as an Euler discretization of the continuous counterpart proposed by Hull and White (1987), with returns observed at equally spaced discrete intervals. Also, the specifications of the most popular discrete-time

[15] GARCH model is usually attributed to Bollerslev (1986), although it was proposed simultaneously by Taylor (1986).

[16] As noted in Andersen et al. (2006, p. 815): "Adopting the rational perspective that asset prices reflect the discounted value of future expected cash flow, such prices should react almost continuously to the myriad of news that arrive on a given trading day". Applications in portfolio and asset pricing that were cast in continuous-time framework can be dated back as early as Merton (1969) and Black and Scholes (1973).

SV models are often rationalized through the discretization of specific continuous-time SV models. Dassios (1995) examines the issue of approximation accuracy; see also Ghysels et al. (1996, section 4) for a rigorous discussion of the relationship between discrete and continuous time SV models.

The remainder of this section focuses on discrete-time SV models by assuming that prices are only observable at discrete and equally spaced points in time. Section 2.5 deals with the continuous-time counterpart, which is the underlying assumption of realized volatility.

2.4.1 Stochastic Volatility Models: Specification

Motivated by MDH, the one-period return is decomposed into an expected conditional mean and an innovation,

$$y_t = \mu_t + \sigma_t \varepsilon_t \qquad \varepsilon_t \sim iid\ (0,1) \tag{2.3}$$

The innovation is expressed as a standardized white noise process scaled by time-varying conditional volatility. It is implied that prices are observed regularly at fixed time intervals, e.g. daily closing prices or end-of-the-week prices. Furthermore, the predictable part μ_t is negligible, which holds empirically over short intervals such as daily or weekly, and is also consistent with a no-arbitrage condition. Accordingly, we simplify (2.3) to

$$y_t = \sigma_t \varepsilon_t \qquad \varepsilon_t \sim iid\ (0,1)$$

There are, of course, an unlimited number of specifications that may be entertained for the conditional volatility process. It is desirable for a model to capture the stylized facts as discussed in Section 2.3. GARCH and SV models have been the two most popular alternatives in the literature. Basically, the two models differ in the assumption of whether volatility is observable or latent: the former assumes σ_t^2 is a deterministic function of past returns; while the latter assumes σ_t^2 is not solely determined by the current and lagged values of y_t, but is also driven by additional noise. Both models feature conditional

heteroscedasticity.[17] However, the notions of conditional volatility σ_t in both models are not exactly identical, as they do not represent the same information set. The assumption of GARCH is relatively strong[18], but has convenient implications as it facilitates likelihood evaluation. Although the SV model is difficult to estimate (see below), it is naturally connected to continuous-time diffusion and this is a strong motivation for its use.

2.4.1.1 Basic Stochastic Volatility Model

The lognormal SV model was first analyzed by Taylor (1982), but was not as popular as GARCH mainly due to estimation difficulties until the influential works by Harvey, Ruiz and Shephard (1994), Jacquier, Polson and Rossi (1994), and Kim, Shephard and Chib (1998). The model is parameterized as

$$\begin{aligned} y_t &= \sigma \exp(h_t/2)\varepsilon_t \qquad \varepsilon_t \sim N(0,1) \\ h_t &= \phi h_{t-1} + \eta_t \qquad \eta_t \sim N(0,\sigma_\eta^2) \end{aligned} \tag{2.4}$$

where the exponential is to ensure positive conditional volatility, and the unconditional mean of h_t is assumed to be zero for the scale parameter σ to be identified. The return y_t is stationary if and only if the conditional volatility is stationary since y_t is the product of volatility and a white noise. All the odd moments of y_t are zero. As shown by Taylor (1986), the kurtosis of y_t is $3\exp(\sigma_\eta^2) \geq 3$. The kurtosis of a normal distribution is 3, so the model is able to capture the stylized fact of high kurtosis. The autoregression of order one (AR(1)) for the state variable is to address volatility clustering. Most empirical applications find the autoregressive coefficient to be close to 1, implying highly persistent volatility dynamic. The state variable h_t is assumed to be normally distributed. As such, the unconditional distribution of the return is a lognormal mixture of normal distributions,

[17] The conditional perspective is exclusively relevant in volatility modeling. As discussed in Section 2.3, volatility clustering is a stylized fact of returns. Therefore, the current level of volatility constitutes critical conditioning information for volatility forecasts, especially in terms of short-term forecasts. Andersen, Bollerslev, Christoffersen and Diebold (2007, p. 514) argue that the conditional perspective is exclusively relevant for most financial risk management purposes. This also explains the weakness of "historical volatility" methods as they lack the ability to incorporate conditionality. The difference between conditional and unconditional moments are crucial for time series forecasting.

[18] Nelson and Foster (1994) argue this assumption is *ad hoc* on both economic and statistical grounds.

corresponding to the assumption postulated by Clark (1973). Shephard (1996) interprets the state variable h_t as a representation of the random and uneven flow of new information into financial markets. See Taylor (1986, 1994) for a detailed discussion on statistical aspects of the model.

For estimation convenience, equation (2.4) can be linearized as[19]

$$\begin{aligned} \log y_t^2 &= \log \sigma^2 + h_t + \log \varepsilon_t^2 \\ h_t &= \phi h_{t-1} + \eta_t \end{aligned} \tag{2.5}$$

The latent log volatility h_t follows an AR(1) process. The first equation indicates $\log y_t^2$ is treated as an AR(1) process with an additive white noise. As such, the autocorrelation function of $\log y_t^2$ is similar to that of an ARMA(1,1), which behaves similarly to the popular GARCH(1,1).[20] This autocorrelation structure is important in capturing the significant autocorrelations at high lags, i.e. the slowly decaying pattern.[21] Taylor (1986) and Harvey (1998) show that for any positive power transformation of $|y_t|$, the autocorrelation function has a similar structure. The distribution of $\log \varepsilon_t^2$ is heavily skewed with a long left-hand tail, being far from normal. This causes some problems in estimation: when ε_t is close to zero, $\log \varepsilon_t^2$ will be a negative number of large magnitude. The mean and variance of $\log \varepsilon_t^2$ are -1.27 and 4.93 respectively, according to the statistical property of a log χ^2 distribution (see Davidian and Carroll (1987, p. 1088)).

Volatility filtering and smoothing The specification in (2.5) implies that one cannot observe volatility but returns, and the log volatility h_t is not an exact function of past returns as in the case of GARCH models. In this sense, volatility is genuinely stochastic. In time series analysis, the exercise of extracting the latent variable from the observable can be conducted within the framework of state space models: see Chapter 4 for more on

[19] It is clear that it is convenient to remove the conditional mean of μ_t from the return decomposition. Because the logarithm of the absolute value $\log |y_t| = \log |\mu_t + \sigma \exp(h_t/2)\varepsilon_t|$ cannot be decomposed. If the return is measured over daily or weekly interval, μ_t is negligible, or one can demean the return prior to estimation.

[20] The autocorrelation of squared returns y_t^2 in GARCH(1,1) model is of the form of an ARMA(1,1).

[21] It is difficult for an AR(1) model to generate significant autocorrelations at high lags, as its autocorrelations decay to zero quickly after the first a few lags.

state-space models. Conditional on different information, the knowledge of volatility varies. In particular, estimation of h_t based on information up to say day t is termed *filtering*, and estimation of h_t based on all information up to the end of sample is termed *smoothing*. Both filtering and smoothing are important for volatility inference. The current level of volatility is critical in forecasting volatility next period given highly persistent volatility dynamics. If $\log \varepsilon_t^2$ is normal, (2.5) is linear and Gaussian, and filtering can be conducted with the celebrated Kalman filter (Kalman (1960)).[22] While Kalman filter can deliver the best linear estimator if the normality assumption does not hold, it is inefficient and far from being optimal, even under the unrealistic assumption that all the parameters are known. It is not easy to derive the optimal filter, and this is an active and evergrowing research area. The current focus is on simulation-based algorithms, in particular, *particle filtering*. The implementation of particle filtering on discrete SV models has been examined by Kim et al. (1998) and Pitt and Shephard (1999); see also Johannes and Polson (2008) and references therein. Alternative filtering schemes, such as importance sampling, have been discussed by Danielsson (1994) and Sandmann and Koopman (1998).

Extensions of the basic SV models The basic SV model discussed so far is restrictive, and some extensions have been proposed in the literature. For example, there is no reason to preclude y_t being a vector. In other words, the model is readily available for multivariate settings as discussed in Harvey et al. (1994). Another possibility is to allow h_t to follow a more complicated ARMA process; see the review in Shephard (1996, section 1.3.4) for other specifications. Also, h_t can be specified to follow a long memory process such as fractional integration, introduced by Breidt et al. (1998) and Harvey (1998), which is examined in Chapter 4 with a Bayesian treatment. The leverage effect, discussed in Section 2.3, is an important stylized fact of returns, especially for equity markets. In the family of GARCH models, the popular EGARCH model due to Nelson (1991) can capture the asymmetry of responses. The counterpart is the asymmetric SV model of Hull and White (1987), which allows for nonzero correlation between the innovations of ε_t and η_t. One important refinement of the SV model is closed form representation under temporal aggregation: this

[22] Harvey (1989) provides a helpful treatment of state space modeling on economic time series.

is to address the empirical observation that the same asset returns measured at different frequencies, say daily and weekly, results in different parameter estimates. The model proposed by Meddahi and Renault (2004) has the advantage of being closed under temporal aggregation, in either multivariate or cross sectional settings.[23]

We briefly compare GARCH and SV models before we close this section. In general, the two models complement each other. The GARCH model is less difficult to estimate, which is an advantage over the SV model; e.g. even if the distribution of ε_t, innovations in the mean equation is misspecified, quasi maximum likelihood estimation is still applicable as shown by Bollerslev and Wooldridge (1992). The GARCH model, as noted by Shephard (1996), is less able to handle extremely large changes in returns. Hence, a fat-tailed distribution of ε_t is often used to capture the high kurtosis. Another drawback of the GARCH model is that it is not convenient for option pricing, unlike the SV model which is naturally suited to a continuous-time setting. However, as derived in a series of papers in Nelson (1990, 1992) and Nelson and Foster (1994), under certain conditions the GARCH model converges weakly to a diffusion process. In other words, the diffusion limit of the GARCH model is a continuous SV model. This justifies the use of GARCH models as an approximate volatility filter even when we believe SV is the underlying model. In terms of forecasting performance, no significant difference between these two model types has been observed in practice; see the discussion in Andersen et al. (2006, p. 814).

2.4.2 Stochastic Volatility Models: Estimation

To evaluate the likelihood function of SV models is difficult due to the latent structure of volatility. Let the vector of returns be denoted as $Y_{\mathbf{T}} = (y_1, y_2, \ldots, y_T)$, and collect the parameters in $\theta = (\sigma, \phi, \sigma_\eta^2)$, then the exact likelihood of the SV model in (2.4) is

$$f(Y_{\mathbf{T}}; \theta) = \prod_{t=1}^{T} f(y_t | Y_{\mathbf{t-1}}; \theta) = \prod_{t=1}^{T} \int f(y_t | h_t; \theta) f(h_t | Y_{\mathbf{t-1}}; \theta) dh_t$$

[23] GARCH type models are not closed under temporal aggregation, which is regarded as a theoretical drawback by some authors, see Andersen, Bollerslev, Christoffersen and Diebold (2007, p. 534). In particular, if daily asset returns follow a GARCH (p,q) process, the same process will not be applicable to the corresponding weekly returns. The weak GARCH class of models proposed by Drost and Nijman (1993) is an attempt to fill this gap.

where the density of $f(y_t|h_t;\theta)$ is of closed form, but not the density of $f(h_t|Y_{\mathbf{t-1}};\theta)$.[24] Therefore, the evaluation of $f(y_t|h_t;\theta)$ requires an integration over the full vector of h_t for $t = 1, 2, \ldots, T$, which is infeasible. An early approach to this problem involves estimating without necessarily evaluating the likelihood function; for example, Taylor (1986) used method-of-moments. Estimation methods have been greatly improved since, especially simulation-based methods, which we will discuss below.

Generalized method-of-moments (GMM) An early example of the implementation of GMM can be found in Melino and Turnbull (1990) and Andersen (1994), with an extension by Hansen and Scheinkman (1995) to continuous-time SV models. In general, Monte Carlo studies (see Jacquier et al. (1994), Breidt and Carriquiry (1996) and Andersen and Sørensen (1996)), show that GMM performs worse as the degree of volatility persistence increases. This is not desirable as highly persistent volatility is a stylized fact. Also, GMM only provides parameter estimates, while it is often the latent volatility that is of interest. As such, it is necessary to conduct a filtering method conditional on GMM point estimates, which further increases the uncertainty in volatility estimates. Shephard (1996, section 1.3.2) further discusses the weaknesses of GMM when applied to SV models.

Quasi maximum likelihood estimation (QMLE) Harvey et al. (1994) suggest a quasi-maximum-likelihood estimator which treats $\log \varepsilon_t^2$ in (2.5) as normal by ignoring the nonnormality. This is convenient in the sense that the system then fits nicely into a linear Gaussian state-space form, with the first equation being the *measurement (observation) equation* and the second being the *state equation*. It is a routine exercise to filter, smooth and evaluate the likelihood within a linear Gaussian state-space form. The Kalman filter delivers an approximation to the likelihood function $f(Y_{\mathbf{T}};\theta)$, parameter estimates are obtained by optimizing the approximated likelihood function, and volatility estimates are readily available as the byproduct of likelihood evaluation. QMLE is easy to conduct, and provides a consistent and asymptotically normal parameter estimator, although $\log \varepsilon_t^2$ is far from being normally distributed with a long left tail. One way to alleviate this

[24] In the case of continuous-time SV model, even $f(y_t|h_t;\theta)$ is not of closed form.

nonnormality is suggested by Breidt and Carriquiry (1996). They use the transformation of Fuller (1996, example 9.3.2) to reduce the kurtosis in the transformed observations by trimming the long tail. Their simulation study indicates the transformation works well especially in small samples.

Bayesian method: Markov chain Monte Carlo (MCMC) The first Bayesian treatment of the SV model is due to Jacquier et al. (1994), who together with the work of Kim et al. (1998) popularizes Bayesian estimation on SV models. When a Bayesian view is taken, the quantity of interest is $f(h_t|Y_{\mathbf{T}})$ or $f(\theta|Y_{\mathbf{T}})$, which is called the posterior density. Unfortunately, the posterior density is of high dimension and analytically intractable. The response is to simulate from the target density, then treat the simulated sample as a realization of the posterior density, based on which statistical inference is made. There are many simulation methods available in the literature, and MCMC is a popular choice for the SV model. (Section 3.2.1 in Chapter 3 and Section 4.3 in Chapter 4 further discuss the Bayesian methodology and MCMC implementation.) One benefit of the MCMC method is that it facilitates data augmentation (Tanner and Wong (1987)), by treating h_t as an additional variable. According to the Clifford-Hammersley theorem, the distribution of $f(h_t, \theta|Y_{\mathbf{T}})$ is fully determined by the conditional distributions of $f(h_t|\theta, Y_{\mathbf{T}})$ and $f(\theta|h_t, Y_{\mathbf{T}})$. Thus the full posterior density can be obtained by working with the two conditionals which might have closed form. The advantage is that the posterior density of latent volatility is obtained as an inherent part of estimation. Importantly, this distribution or other densities of interest such as the prediction density, is the quantity obtained after the uncertainty in parameter estimation being integrating out. It is always necessary to conduct convergence diagnostics on the MCMC chain to ensure the chain converges to the target density, although there is no universally accepted rule of convergence available yet. Another advantage of the MCMC method is that it can be generalized to continuous-time SV models: see Jones (1998), Eraker (2001), Elerian, Chib and Shephard (2001) and Roberts and Stramer (2001). The MCMC method is straightforward in principle, however its implementation requires caution. For example, the choice of single-move or block updating h_t appears to be a minor issue, but can make large differences to computational

efficiency. See the discussion in Shephard (1996) and Kim et al. (1998).[25]

Method of simulated moments This is a simulated maximum likelihood estimation method, and is also as computationally intensive, like the MCMC method. It is not Bayesian so there is no need to specify a prior. Danielsson (1994) and Danielsson and Richard (1993) provide early applications to SV models. An important refinement of this method, efficient method of moments (EMM) (see Gallant and Tauchen (1996, 1998)), is promising, as it can in principle be readily generalized to continuous-time SV models. Andersen et al. (2006, section 4.2) provide a detailed and accessible introduction to the application of EMM to SV models. The importance sampling method used in Durbin and Koopman (1997) is also a simulated likelihood method which evaluates the exact likelihood of the basic SV model in (2.4).

2.5 Realized Volatility

GARCH and SV models offer a convenient and parsimonious frameworks for modeling key stylized facts of returns. Since tick-by-tick (continuous-time) data became commonly available,[26] one motivation of volatility modeling is to incorporate highly informative intra-day data. However, neither type of model is capable of dealing with high frequency data. The specification in (2.4) is too restrictive to capture various pronounced and systematic intraday patterns, for example, intraday volatility periodicity and longer run persistence, as examined in Andersen and Bollerslev (1998*b*) and Dacorogna et al. (2001). As noted in Shephard (1996, p. 49), direct modeling of the ultra high-frequency volatility process is complex and cumbersome, because "continuous-time models are grossly untrue over short periods of time due to institutional factors". The response in the literature is to seek

[25] In general, it is believed that block updating is more efficient, as shown in Kim et al. (1998). To facilitate block updating, Kim et al. (1998) propose to approximate the distribution of $\log \varepsilon^2$ by a mixture of normals. The convergence diagnostics conducted in their paper appear promising. While Ghysels et al. (1996, section 5.6) argue that this is at the cost that the draws navigate in a much higher dimensional space due to the discretization effected, and the convergence of chains based upon discrete mixtures is sensitive to the number of components and their weights.

[26] Many authors, see Andersen et al. (2006), regard the pioneering work by Olsen & Associates on high-frequency data (summarized in Dacorogna, Gençay, Müller, Olsen and Pictet (2001)), helps to pave the way for many of the more recent empirical developments in the realized volatility area.

nonparametric measures of volatility that avoid the otherwise heavy parameterizations. This section is about one of the most popular non-parametric measures of volatility, *realized volatility* (RV).

In general, RV is a sum of squared intraday returns, and provides a consistent nonparametric estimate of price variability that has transpired over a daily interval. It is model-free because it relies on no explicit specification. It is applicable to any log-price process that is subject to a no-arbitrage condition and weak auxiliary assumptions, and we will discuss more on these general assumptions later. There are several motivations for RV modeling. Firstly, volatility is inherently latent. SV models have to rely on simulation-based methods such as particle filtering or the reprojection approach of Gallant and Tauchen (1998) to estimate volatility. The complications are more profound in the continuous-time scenario. In this regard, it is convenient to treat volatility as observable, and the development in quadratic variation theory provides such a volatility measure. Also, when evaluating the volatility forecasting performance of various models, it is desirable to have a model-free measure to be treated as the "realization of volatility". Secondly, high kurtosis is a pronounced stylized fact of returns. It is well documented that GARCH and SV models cannot remove all the nonnormality in unconditional returns, as returns standardized by conditional volatility estimated from both models still exhibit high kurtosis. As shown in Andersen et al. (2000*a*) and Andersen, Bollerslev, Diebold and Ebens (2001), returns standardized by RV approximate normality much better. This conditional normality combined with the frequently found lognormality of RV itself are consistent with MDH, and this has important implications in market risk management where the conditional nonnormality of returns has been a key stylized fact. Finally, GARCH and SV models are unable to incorporate intraday information without increased complexity, and hence, the volatility implied by both models is assumed to be constant over the "day". This is unrealistic and can be misleading; see the excellent example in Andersen, Bollerslev, Christoffersen and Diebold (2007, section 11.4.1). If we are interested in the price variability over a daily interval, therefore, then it is worthwhile to use sufficiently high-quality intraday price data to uncover this feature.

Early work on high-frequency data to measure volatility includes Schwert (1990),

Hsieh (1991) and Zhou (1996); the last of these uses cumulative squared intraday returns to approximate the underlying return variance but in a highly stylized way.[27] The theoretical and empirical results in Andersen and Bollerslev (1998*a*), Comte and Renault (1998), Andersen, Bollerslev, Diebold and Labys (2001) and Barndorff-Nielsen and Shephard (2002*a*) have greatly popularized the applications of RV. This is a vast and rapidly growing area. Recent reviews available include Andersen et al. (2006, p. 838), McAleer and Medeiros (2008), Andersen and Benzoni (2009) and Andersen et al. (2010). An intuitive introduction to RV can be found in Taylor (2005, sections 12.8-12.9).

There has been rapid development in both the theoretical and empirical aspects of RV, and we will review some of the main theoretical results and empirical evidence on the unconditional characteristics of RV in the remainder of this section.

2.5.1 Theoretical Results on Realized Volatility

The concept of RV is cast in continuous-time. In the following discussion, we first address the question "what does realized volatility measure?", in the setting of a continuous-time SV model. Then we show an informal but intuitive proof of the consistency of RV as a volatility estimator, under a general model-free framework. Our notation and derivations in this section follow Zivot (2005) and Andersen et al. (2006) closely.

2.5.1.1 What Does Realized Volatility Measure?

Suppose the underlying log-price evolves continuously with the form of the following stochastic differential equation

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) \tag{2.6}$$

where $\mu(t)$ denotes the drift, $\sigma(t)$ refers to the *spot volatility*, and $W(t)$ is a Wiener process (standard Brownian motion). Both the drift and spot volatility can be time-

[27] The precedent of using cumulative daily squared returns as monthly volatility measures can be found in French, Schwert and Stambaugh (1987) and Schwert (1989).

varying. Return over an (infinitesimal) small time interval Δ is approximated as

$$r(t,\Delta) \equiv p(t) - p(t-\Delta) \simeq \mu(t-\Delta)\Delta + \sigma(t-\Delta)\Delta W(t) \tag{2.7}$$

where $\Delta W(t) \equiv W(t) - W(t-\Delta) \sim N(0,\Delta)$. For return over a unit time interval, $\Delta = 1$ (say daily), and

$$r(t) \simeq \mu(t-1) + \sigma(t-1)z_t, \qquad z_t \sim N(0,1)$$

where we follow the convention of denoting the daily return $r(t)$ rather than $r(t,1)$. If the drift and spot volatility are constant over the interval $[t-1,t]$, then the daily return can be written as

$$r(t) \simeq \mu_{t|t-1} + \sigma_{t|t-1}z_t, \qquad z_t \sim N(0,1) \tag{2.8}$$

This resembles the return decomposition in the discrete-time SV model in (2.3), which is an approximation to the continuous-time counterpart. If the drift and spot volatility are allowed to evolve over time, as normally assumed in continuous-time models, the exact representation of daily returns is

$$r(t) = p(t) - p(t-1) = \int_{t-1}^{t} \mu(s)ds + \int_{t-1}^{t} \sigma(s)dW(s) \tag{2.9}$$

We introduce the notion of *integrated volatility* as

$$IV_t \equiv \int_{t-1}^{t} \sigma^2(s)ds$$

which is analogous to the notion of conditional volatility $\sigma^2_{t|t-1}$ in a discrete-time setting. The quantity IV_t is central to certain models for option pricing as it represents the price variability transpiring over a given discrete interval; see Hull and White (1987). It is the dynamics of spot volatility that determines IV_t. Direct modeling of spot volatility is difficult, however, as continuously recorded prices are not available in practice. Even for extremely liquid markets, microstructure effects, such as discrete price grids and bid-ask bounce effects, prevent observation of the true continuous price realization. RV avoids

these difficulties by providing a consistent nonparametric estimator of IV_t and simultaneously preserving the intraday information.

2.5.1.2 An Informal Proof of the Consistency of Realized Volatility

The concept of RV is developed in a very general framework, and the underlying model is not necessarily included in the family of SV models. Let $p(t)$ denote a univariate continuous log-price process,[28] and if $p(t)$ is in the class of special semi-martingales then it has the unique decomposition

$$p(t) = p(0) + A(t) + M(t)$$

where $A(t)$ is a predictable drift component of finite variation which allows for stochastic evolution, and $M(t)$ is a local martingale. The purpose of the semi-martingale requirement is to rule out arbitrage opportunities. Let the unit interval be daily, and m denote a positive integer indicating the number of returns obtained by sampling $m = 1/\Delta$ times per day. The continuously compounded return over the period $[t - \Delta, t]$ is

$$r(t, \Delta) = p(t) - p(t - \Delta)$$

Formally, realized volatility is defined as

$$RV_t = \sum_{j=1}^{1/\Delta} r^2(t - 1 + j \cdot \Delta,\ \Delta)$$

The *quadratic variation* (QV) of the return process at time t is defined as

$$QV(t) = p \lim_{m \to \infty} \sum_{j=0}^{m-1} \{p(s_{j+1}) - p(s_j)\}^2$$

with $0 = s_0 < s_1 < \ldots < s_m = t$. QV measures the realized sample path variation of the squared return process. It is a unique and model-free *ex-post* volatility measure. From

[28] The theoretical results are readily applicable to a multivariate case. The case of extension to a multivariate framework, realized covariance, is another promising aspect of RV modeling, as standard multivariate GARCH models are too heavily parameterized to be useful in realistic large-scale problems.

the theory of quadratic variation for semi-martingales (see e.g., Back (1991) and Protter (1990)), the following convergence result holds:

$$RV_t \to QV(t) - QV(t-1) \equiv QV_t \qquad \text{as } m \to \infty$$

That is, daily RV converges in probability to the daily increment in quadratic variation. In the case of pure diffusion in (2.6), $QV_t = IV_t$.

Let us show why this convergence holds for (2.6). According to (2.7)

$$r^2(t,\Delta) \simeq \mu^2(t-\Delta)\Delta^2 + 2\Delta\mu(t-\Delta)\sigma(t-\Delta)\Delta W(t) + \sigma^2(t-\Delta)\big(\Delta W(t)\big)^2$$

Over (infinitesimal) small time intervals, Δ, the drift term is negligible. Then

$$Var[r(t,\Delta) \mid \mathcal{F}_{t-\Delta}] \simeq E[r^2(t,\Delta) \mid \mathcal{F}_{t-\Delta}] \simeq \sigma^2(t-\Delta)\Delta \tag{2.10}$$

where $\mathcal{F}_t$ denotes the information available at time t. The result in (2.10) can be generalized to the multi-period setting

$$\begin{aligned}\sum_{j=1}^{1/\Delta} E[r^2(t-1+j\cdot\Delta,\ \Delta) \mid \mathcal{F}_{t-1+(j-1)\cdot\Delta}] &\simeq \sigma^2\big(t-1+(j-1)\cdot\Delta\big)\cdot\Delta \\ &\simeq \int_{t-1}^{t}\sigma^2(s)ds.\end{aligned}$$

where the last approximation is due to the sum converging to the corresponding integral as the size of Δ shrinks toward zero, i.e. $m \to \infty$. Therefore, $RV_t \to IV_t$.

It is clear that RV is computed without reference to any specific model, with the assumption only that the underlying log price is a semi-martingale. The above derivation only indicates the consistency of RV, without evaluating its accuracy. Barndorff-Nielsen and Shephard (2002*b*) provide some distributional results for RV, which show that the measurement errors of RV are approximately uncorrelated, and this has important implications for time series modeling. More details of asymptotic results and multivariate generalizations of RV can be found in, e.g., Meddahi (2002), Barndorff-Nielsen and Shep-

hard (2001, 2002*a*, 2004*a*), and Andersen et al. (2003).

2.5.2 Empirical Results and Modeling of Realized Volatility

The statistical characteristics of RV appear to be consistent across a range of assets: log RV is nearly normal and daily returns standardized by RV are approximately normal. Andersen et al. (2000*a*, 2003) show that when RV is used, the distribution of the standardized exchange rate series is almost Gaussian. Andersen, Bollerslev, Diebold and Ebens (2001) observe that a similar result holds for stock returns. The log-normality of RV implies the volatility of volatility is increasing in the level of volatility because log RV is approximately homoscedastic, which is reminiscent of the popular EGARCH and log SV models discussed in Section 2.4.1.1.

The consistency of RV as a volatility measure, and its uncorrelated measurement error, together with the approximate log-normal distribution, suggests a simple alternative volatility modeling strategy: RV can be treated as an observable time series, which fits into the fully-developed standard autoregressive moving average (ARMA) modeling framework. Also, the state-space form of RV in Barndorff-Nielsen and Shephard (2002*a*) allows the measurement error to be taken into account. A pronounced feature is the degree of temporal dependence in the autocorrelation structure of RV (or log RV), which is consistent across different assets, time periods, and countries. Correlogram plots for log RV show a distinct hyperbolic decay, resembling the stylized fact discussed in Section 2.3.4. The conventional response is to specify a fractionally integrated ARMA (ARFIMA) process for log RV. In Chapter 5 we seek an alternative specification to capture the long memory property of log RV. The memory parameter estimated from log RV is usually around 0.4, indicating a stationary but highly persistent volatility process.

A practical issue in RV modeling is to determine the sampling frequency of returns used to construct RV. The convergence result of RV requires a frequency as high as possible, since this is the condition for RV to approximate QV arbitrarily well. However, this is infeasible in reality. Two factors prevent us from using the highest possible frequency. First, continuously recorded price does not exist in reality although it is convenient for

the derivation of theory. Actual prices are discrete due to rounding and ask and bid size. Barndorff-Nielsen and Shephard (2002*a*) address the inevitable discretization error of RV by structurally modeling the measurement errors based on the asymptotic theory derived in their paper. The other factor is market microstructure noise at the very highest return frequencies, the effect of which violates the no-arbitrage semi-martingale property. The list of microstructure effects is not short: at intraday level, markets experience regular and systematic fluctuations in the quotes and transactions intensity across the trading day; for markets traded 24-hour worldwide, such as foreign exchange, various opening hours of institutions due to time zone differences introduce pronounced intra-day "seasonality"; the release of economic and financial news according to specific timetables tends to cause price jumps and volatility clustering, see, e.g., Andersen and Bollerslev (1998*b*). Other effects include bid-ask bounce, trades taking places on different markets, non-synchronous trading and data recording mistakes, etc.[29] This wide array of noise types induces spurious autocorrelations in intraday return series, the outcome of which is that RV measures would be inflated/biased.[30] Empirical evidence suggests that for actively traded assets, fixing Δ between 5 and 15 minutes typically works well, with 5 minutes being the most common choice and regarded as an optimal trade-off.[31] There are other procedures in the literature for eliminating the systematic bias: applying an ARMA filter to the raw returns to remove the serial correlation, for example, the exponential moving average filtering used in Corsi, Zumbach, Müller and Dacorogna (2001); the volatility signature plot proposed by Andersen, Bollerslev, Diebold and Labys (2000*b*) for diagnostics; an efficient sampling scheme as in Aït-Sahalia, Mykland and Zhang (2005); alternative estimators that are robust to microstructure noise, such as the range-based volatility measure discussed in Alizadeh, Brandt and Diebold (2002) and Brandt and Diebold (2006), or power transformations of

[29] A good example is provided in Andersen et al. (2006, Figure 7), which demonstrates the dangers of using too small a value for Δ in the RV estimation without adjusting for the bid-ask spread effect. As noted in this paper, we only observe the underlying price plus a random bid-ask spread in practice. The bid-ask bounces effect introduces artificial volatility in the observed prices. Accordingly, RV based on very finely sampled high-frequency squared returns produce upward biased volatility measures.

[30] See Campbell et al. (1997, Ch. 3) for a discussion on market microstructure. Descriptions on the adverse effect of market microstructure noises on RV accuracy are provided in, e.g. Andersen et al. (2006) Andersen, Bollerslev, Christoffersen and Diebold (2007, section 11.4.1), Shephard and Andersen (2009, section 6) and Andersen and Benzoni (2009, section 6).

[31] Depending on the market, prices sampled at 15 to 30 minute intervals are also used. In markets that are not 24-hour open, adjustments are required due to the fact that days following weekends and holidays tend to have higher-than-average volatility, see for example Hansen and Lunde (2005).

absolute returns rather than squared returns.

RV has become commonplace in volatility modeling. There are some aspects which have already generated much interest in the rapidly growing literature. First, attempts are made to circumvent inherent market microstructure frictions, including Hansen and Lunde (2006), Bandi and Russell (2008) and Barndorff-Nielsen, Hansen, Lunde and Shephard (2008), among others. Second, an early attempt to estimate the QV of the continuous component of prices in the presence of jumps using the so-called realized bipower variation process is Barndorff-Nielsen and Shephard (2004*b*). The recent work of Andersen, Bollerslev and Diebold (2007) indicates that diffusive volatility is much more persistent than the jump component, so it is helpful to separate jumps and diffusive volatility components in the RV process. Other important issues include multivariate extensions with the use of realized covariance measures, and how to best accommodate the long memory property observed in log RV.[32]

2.6 Conclusion

Volatility modeling is an important topic in financial econometrics, the literature on which is voluminous and this chapter cannot be exhaustive. Our review starts with the Mixture-of-Distributions Hypothesis and descriptions of stylized facts of asset returns. It is a fundamental requirement for successful volatility models to be able to capture the key stylized facts and also to maintain the underlying economic assumptions. Each generation of volatility models, including GARCH, SV models and the RV approach, has been trying to meet these criteria. Of the stylized facts considered, the emphasis here is on the long memory property in volatility, which is discussed in the next three chapters. This review concentrates on SV models, because the discrete-time SV model is the building block of the long memory SV model examined in Chapter 4, and the continuous-time SV model is closely related to the concept of RV, which is the focus of Chapters 3 and 5.

[32]The last issue has precedent in the GARCH modeling, for example Engle and Lee (1999) propose to use different autoregressive volatility components, displaying strong but varying degrees of persistence, to produce a volatility dependence structure that is indistinguishable from long memory.

Chapter 3

Realized Volatility, Long Memory and Bayesian Model Averaging

3.1 Introduction

Volatility is a concept used to measure the spread of all possible outcomes of asset returns. It is well recognized that volatility is time varying, and that volatility exhibits a degree of predicability despite the fact that returns are unpredictable. The popular GARCH models assume that the conditional volatility is fully determined by the realization of past returns, while SV models assume volatility is genuinely latent. Both GARCH and SV models are commonly applied to low-frequency data, such as daily, weekly or monthly data.

In recent years, however, the focus has shifted to the use of high-frequency data, mainly because intra-day data can provide rich information absent in daily returns. The direct modeling of returns based on intra-day data poses great challenges in practice. This is not only because the number of observations might be overwhelming, but also because intra-day returns are contaminated by various market microstructure effects, such as bid/ask bounce, nonsynchronous trading, and calendar effects, etc. For example, the absolute five-minute returns often exhibit significant periodic (intra-day and intra-week)

activity patterns, and it is rather difficult to incorporate these stylized facts directly into a parametric model. One possible way out of this dilemma is by the use of realized volatility. Realized volatility is constructed from squared returns within the horizon of interest, and is essentially a model-free measure used to estimate *ex post* realized volatility over the horizon. For example, currency markets are 24-hour traded, and if prices are sampled every five minutes, we have 288 five-minute returns per trading day. The sum of these 288 squared returns is the realized volatility on the corresponding trading day. As shown in a series of papers by Andersen, Bollerslev, Diebold and Labys (2001); Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002*a*), the *ex post* value of realized volatility is an unbiased estimator of conditional variance. One attractive feature of realized volatility is that one can treat it as observed rather than latent, as a result of which, routine time series methods are readily applicable.

In this chapter, we conduct an empirical investigation of the long memory property of realized volatility. Our exercise differs from previous studies in the aspect of model selection. We choose to deal with an *overall averaged* model rather than a single model with predetermined orders of autoregressive and moving average terms. A Bayesian treatment is provided. Posterior model probabilities, i.e. evidence from data, are used to determine the weight allocated to each model. Statistical inference for the memory parameter and out-of-sample predictions are the weighted average over the corresponding quantities from single models considered. Models to be averaged across include both long memory and short memory specifications, namely, autoregressive fractionally integrated moving-average ARFIMA(p,d,q) models and ARMA(p,q) models with different orders of p and q. It is clear that the overall model nests single models, so a more general specification is dealt with, allowing robust statistical inference as uncertainty introduced by single model selection is integrated out in a consistent way. Whilst in principle, model averaging appears promising, a natural question to ask is whether it is worth the extra effort to deal with a number of models rather than a single model with computational cost taken into account. The answer to this question is of practical importance given that this is a typical scenario in time series modeling, and in applications with a number of explanatory variables to be considered. This is a question to be addressed in this chapter.

We try to address this question with an empirical exercise conducted on daily Yen/USD realized volatility, where we focus on posterior inference on the memory parameter and forecasting performance evaluated at different horizons. We also seek evidence from a simulation study, where a few data generating processes are considered and comparisons of forecasting performance are made between Bayesian model averaging and single models. Both aspects support the usefulness of Bayesian model averaging.

We focus on model averaging for the following reasons. It is well documented in the literature that volatility possesses strong persistence in returns constructed from both low-frequency and high-frequency data, see for example, Ding et al. (1993), Breidt et al. (1998), Andersen and Bollerslev (1997*a*), and Bollerslev and Wright (2000). Using measures of volatility such as powers or logarithms of squared returns or realized volatility, these authors have found that the sample autocorrelation function of volatility decays slowly to zero at high lags. In the case of realized volatility modeling, the common practice used to capture this stylized fact is to select a single model from the ARFIMA family. However, as argued in Beran (1994), to find a suitable parametric model that is not too complicated is not always an easy task. A poor choice of p and q can lead to biased estimates of the long memory parameter. As a result, it is risky to select only one model and ignore uncertainty in model choice. Bayesian model averaging (BMA) offers the potential benefit of overcoming this difficulty. Our choice of the Bayesian approach is due to the formality and coherency incorporated in BMA. The implementation of BMA is straightforward once one obtains Bayesian estimates of each model. This is because the marginal likelihood, the critical tool for Bayesian model choice, is a byproduct of posterior density simulation, computed, for example, by Markov chain Monte Carlo (MCMC) simulation. An accessible introduction to BMA can be found in Hoeting, Madigan, Raftery and Volinsky (1999).

The rest of this chapter is organized as follows. In Section 3.2, we discuss the BMA methodology, while in Section 3.3, we focus on Bayesian estimation of ARFIMA models. In Section 3.4, we present an example based on the realized volatility of the Yen/USD market, and document in detail the comparisons between single models and the weighted average. This is done in terms of the posterior inferences on the memory parameter from the full sample, and forecasting performances at different horizons based on recursive estimation.

A Monte Carlo experiment is conducted in Section 3.5. Conclusions are summarized in Section 3.6.

3.2 Methodology

Let us denote y_t as realized volatility observed on day t with $t = 1, \ldots, T$. A time series generated by ARFIMA(p,d,q) process can be represented as

$$\phi(B)(1-B)^d(y_t - \mu) = \theta(B)\eta_t \tag{3.1}$$

where μ is the process mean, B^i is lag operator defined as $B^i y_t = y_{t-i}$, p and q are orders of polynomial in B with $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$. The innovation η_t is assumed to be Gaussian white noise with mean zero and variance σ_η^2. To ensure stationarity and identifiability, d is assumed to be in the range of (0, 0.5), the roots of $\phi(z) = 0$ and $\theta(z) = 0$ lie outside the unit circle and $\theta(z)$ has no roots in common with $\phi(z)$. The positivity of d is to ensure positive autocorrelations of realized volatility as positive autocorrelations are a stylized fact of volatility. We also denote C_p and C_q as the stationary and invertible regions of $\phi(B)$ and $\theta(B)$, respectively.

The parameter d controls the way $\rho(k)$ converges to zero with increasing lag k. The process reduces to the standard ARMA(p,q) model when d=0, with autocorrelations decaying to zero at exponential rate, so ARMA(p,q) is often referred as a short memory process. When $d > 0$, ρ decays more slowly so that the autocorrelations are not geometrically bounded, and the process displays a long memory property. The parameters from the AR and MA parts capture short memory dynamics of a process. By introducing the memory parameter, the fractional ARMA model can allow a broad class of autocorrelation functions. However, this complicates the estimation since the fractional difference operator $(1-B)^d$ is a binomial expansion of the form as,

$$(1-B)^d = \sum_{j=0}^{\infty} \pi_j B^j, \tag{3.2}$$

with

$$\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \Pi_{0<k\leq j}\frac{k-1-d}{k}, \quad j = 0, 1, 2, \ldots, \tag{3.3}$$

and $\Gamma(\cdot)$ being the gamma function. The right side of (3.2) is a polynomial of infinite orders. In this regard, the fractional integrated process can be treated as an autoregressive (AR) or a moving-average (MA) process with infinite orders. Intuitively, this complicates the likelihood evaluation. More details on long memory time series can be found in Brockwell and Davis (1991, Ch. 13), Beran (1994) and Palma (2007).

3.2.1 Bayesian Inference

Under Bayesian methodology, the unknown quantity of interest ω, such as the parameter vector in our case $\omega = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, d, \mu, \sigma_\eta^2)$, is assumed to be a random variable. The ultimate objective in Bayesian inference is to obtain the conditional distribution of ω given observations $y = (y_1, \ldots, y_T)$, namely, the posterior distribution $p(\omega|y)$. According to Bayes' probability law,

$$\begin{aligned} p(\omega|y) &= \frac{p(\omega)p(y|\omega)}{p(y)} \\ &\propto p(\omega)p(y|\omega) \end{aligned} \tag{3.4}$$

where $p(\omega)$ is the prior distribution of the parameters, which incorporates information about parameters before seeing the data, $p(y|\omega)$ is the likelihood function defined by the model. A posterior distribution is fully determined by the data given the prior. The marginal likelihood p(y) is calculated as

$$p(y) = \int p(y|\omega)p(\omega)d\omega \tag{3.5}$$

which is a *normalizing constant* for an individual model, since it is a quantity with parameter uncertainty being integrated out. The marginal likelihood depends only on the prior and the likelihood, and is the key component in Bayesian model comparison and model averaging.

Bayesian prediction proceeds via the predictive density,

$$p(y_{T+1}|y) = \int p(y_{T+1}, \omega|y)d\omega = \int p(y_T|\omega)p(\omega|y)d\omega \tag{3.6}$$

from which random samples can be drawn to make out-of-sample predictions.

Textbook treatment of Bayesian inference can be found in Gelman, Carlin, Stern and Rubin (2004), while those particularly focusing on Bayesian econometrics can be found in Koop (2003) and Geweke (2005).

3.2.2 Bayesian Model Averaging

Section 3.2.1 deals with the case when only one model is involved. In practice, the data generating process might be more complicated so that each model could be treated as an approximation to certain aspects of the latent data generating process (DGP). Under this assumption, one might want to make inferences based on a "weighted average model" if it is available, or equivalently, integrate out the model choice uncertainty introduced by selecting only one model. The extension to BMA is straightforward under the setting in Section 3.2.1.

Suppose we have a set of models under consideration, $M = \{M_1, \ldots, M_J\}$, then the posterior distribution of ω averaged across J models is

$$p(\omega|y) = \sum_{j=1}^{J} p(\omega|y, M_j)p(M_j|y) \tag{3.7}$$

where $p(\omega|y, M_j)$ is obtained through (3.4), and $p(M_j|y)$ is the posterior probability of model M_j given the data. If we treat M_j as nothing more than a set of parameters, according to the same logic as in (3.4),

$$p(M_j|y) = \frac{p(y|M_j)p(M_j)}{p(y)} \tag{3.8}$$

$$\propto p(y|M_j)p(M_j) \tag{3.9}$$

where the division of $p(y)$ is to ensure probabilities in (3.8) sum to one. The prior proba-

bility of model M_j is $p(M_j)$, indicating one's prior knowledge of the model before seeing the data, and $p(y|M_j)$ is the marginal likelihood of model M_j, obtained through (3.5) as

$$p(y|M_j) = \int_{\Theta_{\omega_{M_j}}} p(y|\omega_{M_j}, M_j)p(\omega_{M_j}|M_j)d\omega_{M_j} \tag{3.10}$$

If one chooses a non-informative prior for $p(M_j)$, in particular, $p(M_j) = 1/J$ for $j = 1, \ldots J$, i.e. no prior preference for any single model under consideration, then the marginal likelihood of each model is sufficient to evaluate (3.9).

In terms of forecasting, the posterior mean and variance of y_{T+1} are

$$\begin{aligned} E(y_{T+1}|y) &= \sum_{j=1}^{J} E(y_{T+1}|y, M_j)p(M_j|y) \\ Var(y_{T+1}|y) &= \sum_{j=1}^{J}[Var(y_{T+1}|y, M_j) + E^2(y_{T+1}|y, M_j)]p(M_j|y) - E^2(y_{T+1}|y) \end{aligned}$$

which are derived according to probability laws (see, for example Hoeting et al. (1999)).

These generic derivations indicate the versatility of BMA, since there is no requirement for models to be nested, and parameters from each model might be of different forms, such as state variables, or parameters with constraints. The remaining issue is computational, i.e. posterior simulation in (3.4) and integration in (3.10), which mainly relies on numerical methods, such as Markov Chain Monte Carlo (MCMC) simulation. This is the approach taken in this chapter.

3.3 Bayesian Estimation of ARFIMA Models

The literature on long memory time series estimation is extensive. The classical approach to estimating ARFIMA models can be classified as either semi-parametric methods, such as the GPH estimator proposed in Geweke and Porter-Hudak (1983), or parametric methods, for example, approximate maximum likelihood estimation, such as Whittle's approximate MLE, or approximate MLE based on the AR representation, and exactly MLE

proposed in Sowell (1992). The GPH estimator is easy to implement, and does not require one to specify the orders of AR and MA terms, however, it does require one to determine the number of frequencies involved in estimation, which might be critical in some cases.

As shown in (3.4), the likelihood function evaluation given a set of parameters is an essential part of Bayesian inference. The likelihood evaluation for a long memory process is more involved compared with its short memory counterpart. In this chapter, we take the exact likelihood evaluation in Sowell (1992) despite the fact that it might be computationally costly. The Bayesian treatment of ARFIMA models can be found in Koop, Ley, Osiewalski and Steel (1997), Pai and Ravishanker (1998) and Palma (2007), some details of which will be described in the rest of this section.

3.3.1 Likelihood Evaluation

Let us partition the parameter vector ω of a ARFIMA(p,d,q) process into σ_η^2, and $\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, d)$. Throughout this chapter, we will use sample mean as an estimator of μ, which is justified by the argument in Yajima (1988) that for a long memory model there is only a small loss in efficiency if the mean is used. So we are working on the demeaned time series.

The exact Gaussian likelihood of ARFIMA(p,d,q) with $\mu = 0$ is

$$L(\beta, \sigma_\eta^2) = (2\pi)^{-T/2} |\Sigma|^{-1/2} \exp\{-\frac{1}{2} y' \Sigma^{-1} y\} \tag{3.11}$$

where Σ is the covariance matrix of y, with each element being expressed as $\sigma_\eta^2 \gamma(i-j)$ for $i, j = 1, \ldots, T$ and $\gamma(\cdot)$ is a function of β.

The evaluation of $\gamma(\cdot)$, which is necessary for both likelihood evaluation and out-of sample forecasts, is not an easy task when $d \neq 0$. One common approach to estimating $ARMA(p,q)$ models is via state-space representation, the likelihood function of which can be easily delivered by Kalman filter. If the orders of p and q are reasonable, the dimension of state variables will be feasible as well. However, given the binomial expansion in (3.2), a long memory process is essentially equivalent to a MA process with the order

of q being infinity, which makes the state-space representation cumbersome and being an approximation at its best.

In this chapter, we use the Ox package Arfima to directly evaluate (3.11) given a set of (β, σ_η^2), where the evaluation of autocovariance function is based on the algorithm proposed in Sowell (1992), see Doornik and Ooms (2004, 2006*a*) for more details. For a long time series, the storing and inversion of Σ might be costly and numerically unstable since it is a $T \times T$ matrix. Given the stationarity condition, Σ is a Toeplitz matrix, so the number of calculations required to invert Σ can be reduced by choosing a suitable algorithm. The Arfima package allows one to store the whole covariance matrix efficiently rather than in its raw form, and uses Durbin's algorithm to invert Toeplitz matrix. This gain in computation efficiency is critical especially when we conduct MCMC to simulate from the posterior distribution.

3.3.2 Prior

Throughout this chapter, we use the prior following the arguments in Koop et al. (1997). More specifically, assuming independent prior structure,

$$\begin{aligned} p(\beta, \sigma_\eta^2) &= p(\beta)p(\sigma_\eta^2) \\ &\propto p(\beta)\sigma_\eta^{-2} \end{aligned} \tag{3.12}$$

where we choose a uniform prior for β by constraining $(\phi_1, \ldots, \phi_p) \in C_p$, $(\theta_1, \ldots, \theta_p) \in C_q$, and $d \in (0, 0.5)$, and improper prior on σ_η^2. The problem that might be caused by using an improper prior of σ_η^2 when conducting Bayesian model comparison is overcome by the use of a proper non-informative prior on β.

In terms of model prior, we allocate equal prior model probability to each model by setting $p(M_j) = 1/J$. The posterior model probability then simplifies to $p(M_j|y) = \frac{p(y|M_j)}{\sum_{j=1}^{J} p(y|M_j)}$, so as a result, the marginal likelihood is sufficient to determine the weight attached to each model for model averaging.

3.3.3 Posterior Simulation

Since the analytical form of the posterior distribution is not available, we have to rely on numerical methods to evaluate $p(\beta, \sigma_\eta^2|y)$ and $p(y|M_j)$, i.e. to simulate from the posterior distribution, then make inference based on the draws from simulation rather than providing a full analytical treatment. In this chapter we apply a hybrid MCMC method, Metropolis-Hastings within Gibbs, the details of which we now turn to. The discussions are from the perspectives of posterior parameter simulation and posterior prediction simulation, respectively.

The basic tool in the Gibbs sampler is blocking. Instead of direct sampling from $p(\beta, \sigma_\eta^2|y)$, we sample from the conditional posterior density $p(\beta|\sigma_\eta^2, y)$ and $p(\sigma_\eta^2|\beta, y)$ separately. More specifically, given an initial value of $\sigma_\eta^{2^0}$, we sample β^1 from $p(\beta|\sigma_\eta^{2^0}, y)$, Given β^1, sample $\sigma_\eta^{2^1}$ from $p(\sigma_\eta^2|\beta^1, y)$. Iterating the exercise M times after a *burn-in* period to remove the effect of initialization, we obtain a Markov chain $\{\beta^{1,\dots,M}, \sigma_\eta^{2^{1,\dots,M}}\}$, the distribution of which approximates that of the posterior density $p(\beta, \sigma_\eta^2|y)$ under general conditions. See Geweke (2005) for more details.

In our case, drawing samples from $p(\sigma_\eta^2|\beta, y)$ is a standard Bayesian exercise, the conditional posterior of σ_η^2 is inverted gamma distribution IG(a,b), where the values of a and b are chosen following the argument in Palma (2007, Ch. 8) as

$$a = \frac{T - r + 4}{2}, \quad b = \frac{\hat{\sigma}_\eta^2 (T - r + 2)}{2}, \tag{3.13}$$

with r being the number of model parameters, and $\hat{\sigma}_\eta^2$ being the MLE of σ_η^2.

The conditional posterior of β is not of known form, so the Metropolis-Hastings algorithm is used to update β. The conditional posterior of β is proportional to the following equation up to a normalizing constant

$$p(\beta|\sigma_\eta^2, y) \propto |\Sigma|^{-1/2} \exp\{-\frac{1}{2} y' \Sigma^{-1} y\} p(\beta) \sigma_\eta^{-2}. \tag{3.14}$$

When conducting the Metropolis-Hastings algorithm, the proposal density of β is the

multivariate normal density $\beta \sim N(\hat{\beta}, c\Omega)$, where $\hat{\beta} = (\hat{d}, \hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q)$ is the MLE estimates, Ω is the inverse of the Hessian at $\hat{\beta}$ and c is a scaling constant. Given $\beta^{(m)}$ in the Gibbs sampler, we draw a candidate ξ from $N(\hat{\beta}, c\Omega)$, which is accepted with probability $\min\{p(\xi|\sigma_\eta^2, y)/p(\beta^{(m)}|\sigma_\eta^2, y), 1\}$, with $p(\cdot)$ evaluated by (3.14).

When we sample β from the proposal density $N(\hat{\beta}, c\Omega)$, we discard those samples out of the stationarity region of $C_p \times C_q \times C_d$, and initialize the Gibbs sampler with MLE estimates. The scaling parameter c is tuned to obtain an acceptance rate in the range of 30-60% for each model.

Posterior prediction simulation follows the same logic as posterior parameter simulation, if we treat out-of-sample prediction as an unknown parameter in the model. Given $(\beta^{(m)}, \sigma_\eta^{2(m)})$ in the Gibbs sampler, one can draw a sample from the predictive density, which is a normal density as

$$p(y_{T+1}|y, \beta^{(m)}, \sigma_\eta^{2(m)}) \sim N(0, \Sigma_{11} - \Sigma_{1T}\Sigma^{-1}\Sigma'_{1T}) \tag{3.15}$$

where $\Sigma_{11} = \sigma_\eta^2\gamma(0)$, and Σ_{1T} is a row vector with each element being $\sigma_\eta^2\gamma(T+1-j)$ for $j = 1, \ldots, T$. The extension to multi-step ahead prediction is straightforward.

The moments of posterior prediction from each individual model can be weighted by (3.8), to obtain estimates of moments averaged across all the models under consideration.

3.3.4 Marginal Likelihood Approximation

The marginal likelihood of each model is approximated by the method of density ratio marginal likelihood approximation. Geweke (2005, Ch. 8) discusses the method in detail. There are also other approximation methods available, such as approximation using importance sampling, for example, Koop et al. (1997). However, as argued in Geweke and Whiteman (2006), one problem with the importance sampling approximation is that the more the dimensions of integration involved, the greater the loss in approximation efficiency, which is a weakness inherent in any numerical approximation method.

In particular, in order to approximate (3.10), we save the output from MCMC simulation $(\beta^{(m)}, \sigma_\eta^{2(m)})$ and evaluate the corresponding value of

$$k(\beta^{(m)}, \sigma_\eta^{2(m)}|M_j) = p(y|\beta^{(m)}, \sigma_\eta^{2(m)}, M_j) \times p(\beta^{(m)}, \sigma_\eta^{2(m)}|M_j) \tag{3.16}$$

given the model j. According to Geweke (1999),

$$M^{-1} \sum_{m=1}^{M} f(\beta^{(m)}, \sigma_\eta^{2(m)})/k(\beta^{(m)}, \sigma_\eta^{2(m)}) \quad \overrightarrow{a.s.} \quad [p(y|M_j)]^{-1} \tag{3.17}$$

where $f(\cdot)$ can be constructed from posterior simulation to ensure the support of which is the same as that of posterior distribution of (β, σ_η^2). In this chapter, we choose f to follow a multivariate normal distribution with mean $\hat{\omega}$ and variance $\hat{H}$, truncated to its highest density region of size $100(1-\alpha)\%$,[1] where $\hat{\omega}$ is the MCMC approximation of $E[\omega|y, M_j]$, $\hat{H}$ is the $MCMC$ approximation of $Var[\omega|y, M_j]$, and ω denotes the parameter vector of a given model.

3.4 Empirical Analysis

We begin this section with some notation to show the construction of realized volatility. Let P_i denote the logarithm asset price at time i, where i is a regularly spaced time interval: for example, we take five-minute intervals in this chapter.[2] However, prices are actually observed irregularly, so we need an algorithm to construct regularly-spaced prices, details of which are given below. Let Δ denote the fraction of a trading day corresponding to the implied sampling frequency, i.e. five minutes, and $m = 1/\Delta$ denotes the number of sampled prices per trading day. The currency markets are 24-hour traded so that $\Delta = 5/(24 \times 60)$, and $m = 288$ indicating 288 five-minute intervals per trading day.

Let T denote the number of days in the sample, which gives $m \times T$ five-minute asset

[1] The value of α is chosen as 0.05 in this chapter, the results of which are almost identical to that when $\alpha = 0.1$.

[2] To construct realized volatility from five-minute returns is a common choice for currency markets. When we tried to construct realized volatility from thirty-minute returns, the time series obtained does not differ much.

prices in total. Five-minute returns are defined as

$$r_{i+\Delta} = P_{i+\Delta} - P_i$$

Realized volatility constructed from five-minute returns on day t is defined as

$$y_t = \sum_{j=1}^{288} r^2_{t-1+j\Delta}, \qquad \text{for} \quad t = 1, \ldots, T.$$

3.4.1 Raw Data

The raw data used in this chapter are the over-the-counter (OTC) quotes of the Yen/USD rate which appeared on Reuter's FXFX page during the sample period January 2003 through August 2008. Each quote consists of a *bid* and *ask* price and the times at which they are recorded to the nearest millisecond. Raw data are often contaminated by recording errors. So we first filter the data based on the algorithms proposed by Dacorogna et al. (1993). We also follow the argument in Müller et al. (1990) to define log price P_i as

$$P_i = \frac{\log P_{bid,i} + \log P_{ask,i}}{2} \tag{3.18}$$

where $P_{bid,i}$ and $P_{ask,i}$ are the *bid* and *ask* prices.

To construct the price for each five-minute interval, we follow the convention to use the last price from the interval, although this might make returns not exactly evenly spaced. We ignore this minor irregularity given that the data used in this chapter are of high liquidity with average time between available prices being much less than one minute.[3] Alternatively, the price for each five-minute interval can be obtained by linearly interpolating from the last price from this interval and the first price from next interval. We did not follow this method because this can cause spurious predictability. See Taylor (2005, Ch. 12) for more details.

The days on which trading is extremely thin have been removed. For example, the

[3]The number of quotes of one trading day can be overwhelming; for example, the average daily number of quotes in 2007 exceeds 20,000.

weekend period between Friday 21:05 GMT and Sunday 21:00 GMT,[4] fixed holidays such as Christmas (24-26 December) and New Year (31 December -2 January). We ignore these thin trading days following Andersen, Bollerslev, Diebold and Labys (2001); Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002*a*). We also remove those days contaminated by *data holes* in the Reuter's data feed, which are indicated by long sequences of zero five-minute returns, where the quotes are missing. Whenever we remove a trading day, we always cut from 21:05 GMT on one night to 21:00 GMT the next evening.

In total, we obtain 1436 complete days, with $288 \times 1436 = 413\,568$ five-minute returns. Based on these 413 568 observations, daily returns and realized volatility are constructed.

3.4.2 Preliminary Data Analysis

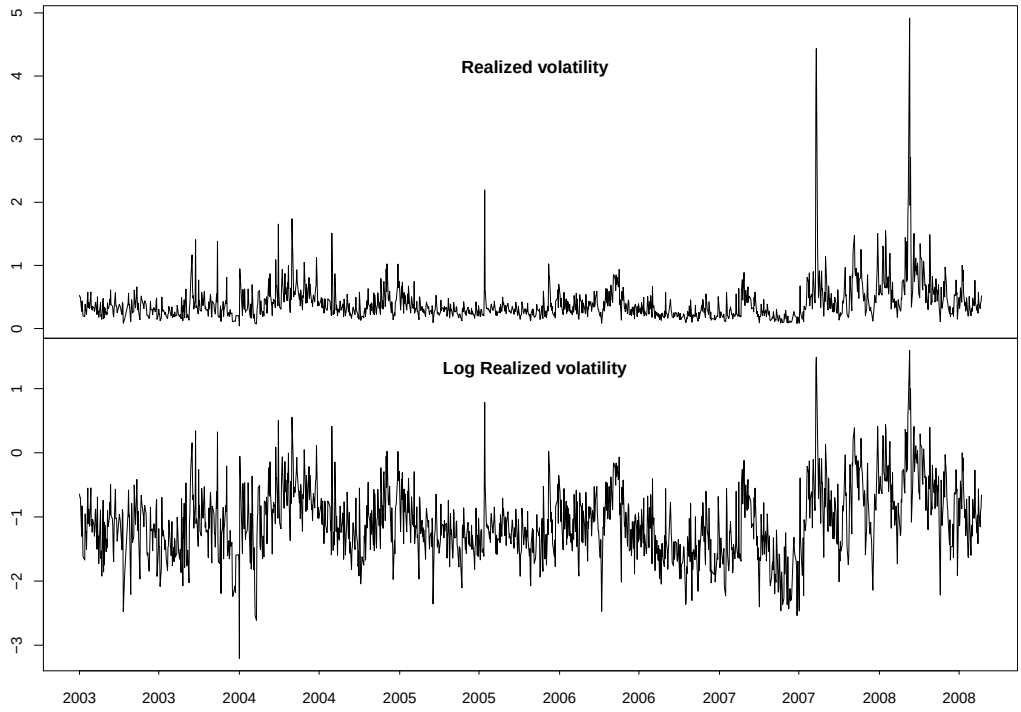
The time series and ACF plots of daily realized volatility and its logarithm are presented in Figures 3.1 and 3.2, respectively. The pattern in the ACF plots indicates substantial positive autocorrelations for both series, with slowly-decaying autocorrelation, even at lags of 100, which is a typical characteristic of a long memory process. The autocorrelations of logarithm realized volatility appear to be smoother and more persistent compared with that of realized volatility. This is as expected given quite a few spikes observed in the time series plot of realized volatility, and that the effect of additive noise is to reduce the autocorrelations of the volatility process. A close examination of the ACF plot also indicates minor seasonality at weekly frequency.

We conduct the modified rescaled range tests proposed in Lo (1991) and Giraitis, Kokoszka, Leipus and Teyssière (2003); the results of both tests reject the null hypothesis of short memory at 5% significance level. The GPH estimator gives a memory parameter of 0.4129 with standard deviation being 0.0259 when the number of periodogram values is chosen at 700.

In terms of the models fitted, we vary the order of p and q from 0 to 3 for both ARFIMA(p,d,q) and ARMA(p,q) models. Since the data are observed on daily frequency, we also extend the AR order to $p = 4, 5$ with $q = 0$ to capture possible weekly season-

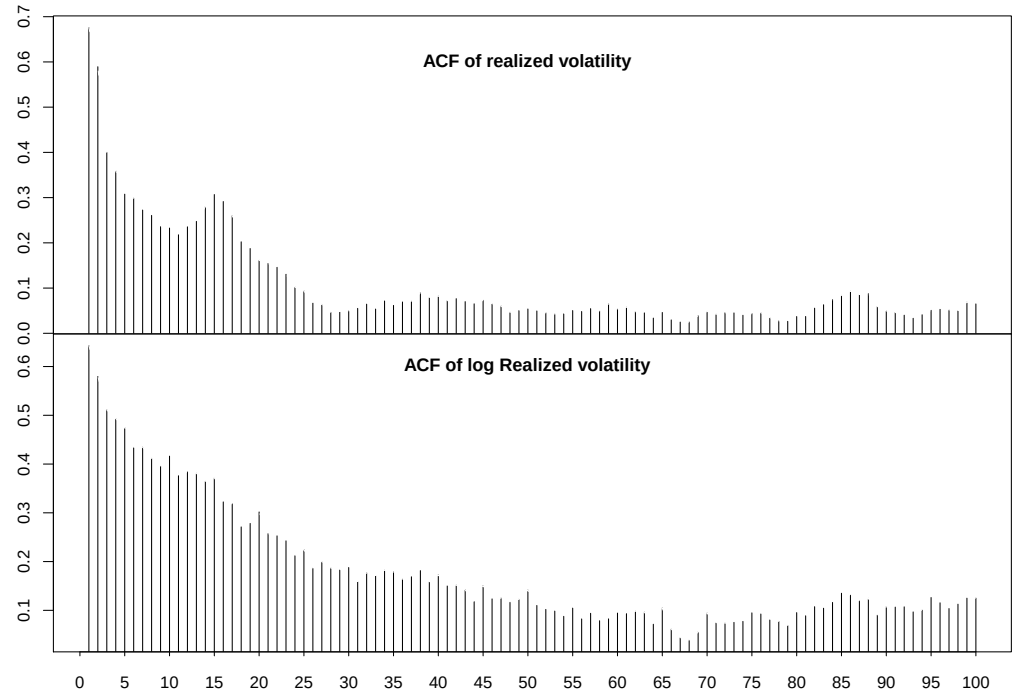
[4]Less than 0.1% of the quotes are made between 21:05 GMT on Friday and 21:00 GMT on Sunday.

Figure 3.1: Daily realized volatility for the Yen/USD (07/01/2003 - 23/08/2008)



[a]The sample period is from 07/01/2003 to 23/08/2008, with 1436 daily observations in total.

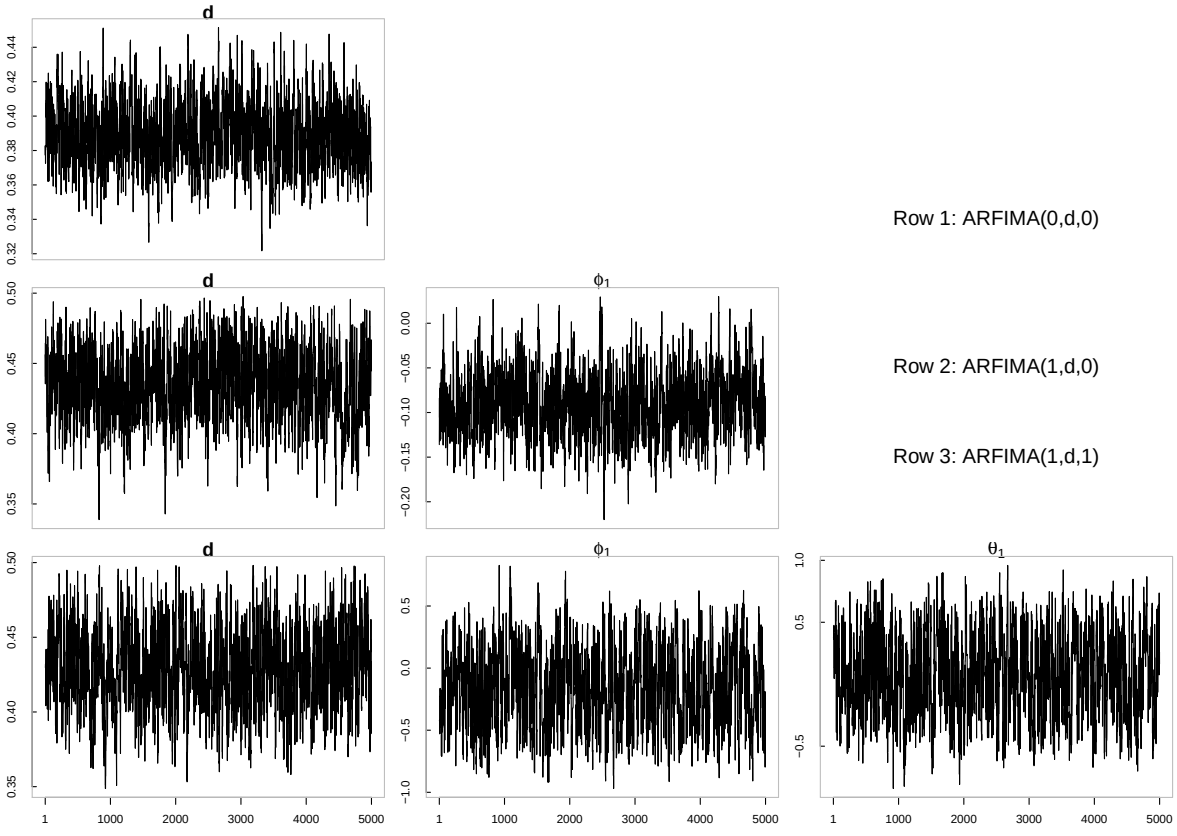
Figure 3.2: ACF of the daily Yen/USD realized volatility (07/01/2003 - 23/08/2008)



[a]The sample period is from 07/01/2003 to 23/08/2008, with 1436 daily observations in total.

ality. The logarithmic transformation is used throughout the estimation to ensure the positiveness of predicted volatility. When conducting MCMC simulation for each model, we always make inferences based on 5000 iterations after discarding the first 5000 runs. The convergence rate of most models is rapid, except for the case of ARFIMA(3,d,2) and ARFIMA(3,d,3); we therefore exclude these two cases from the BMA exercise, which leaves 33 models in total. The MLE estimates also have convergence problems for these two models, with the MA roots of MLE estimates for ARFIMA(3,d,2) model being outside the invertibility region. The MCMC iterations for some parameters are provided in Figures 3.3, 3.4 and 3.5, where the convergence of most parameters appears to be reasonably rapid. The estimates of innovation variance σ_η^2 from different specifications of the long memory process are similar to each other, with the posterior mean being around 0.17.

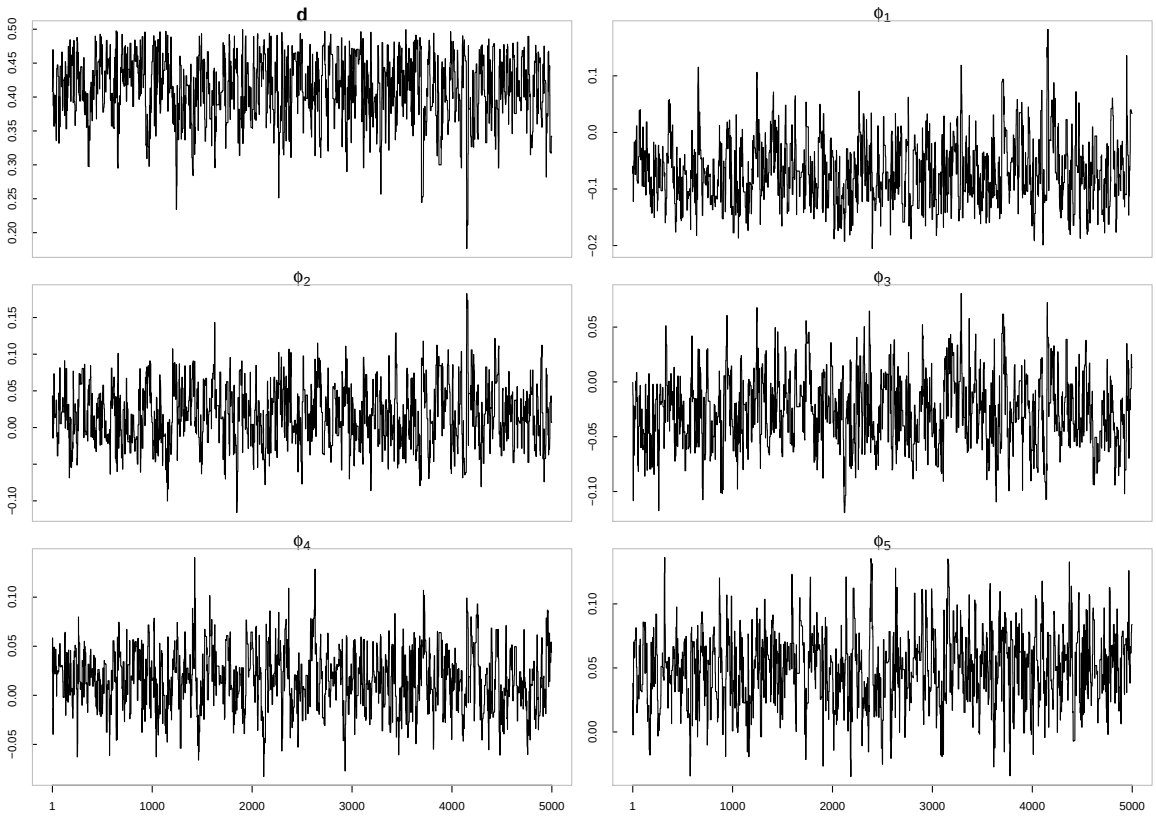
Figure 3.3: MCMC iterations of parameters of ARFIMA(0,d,0), ARFIMA(1,d,0) and ARFIMA(2,d,3) models of full sample



[a]The plotted 5000 MCMC iterations are obtained after discarding the first 5000 runs.

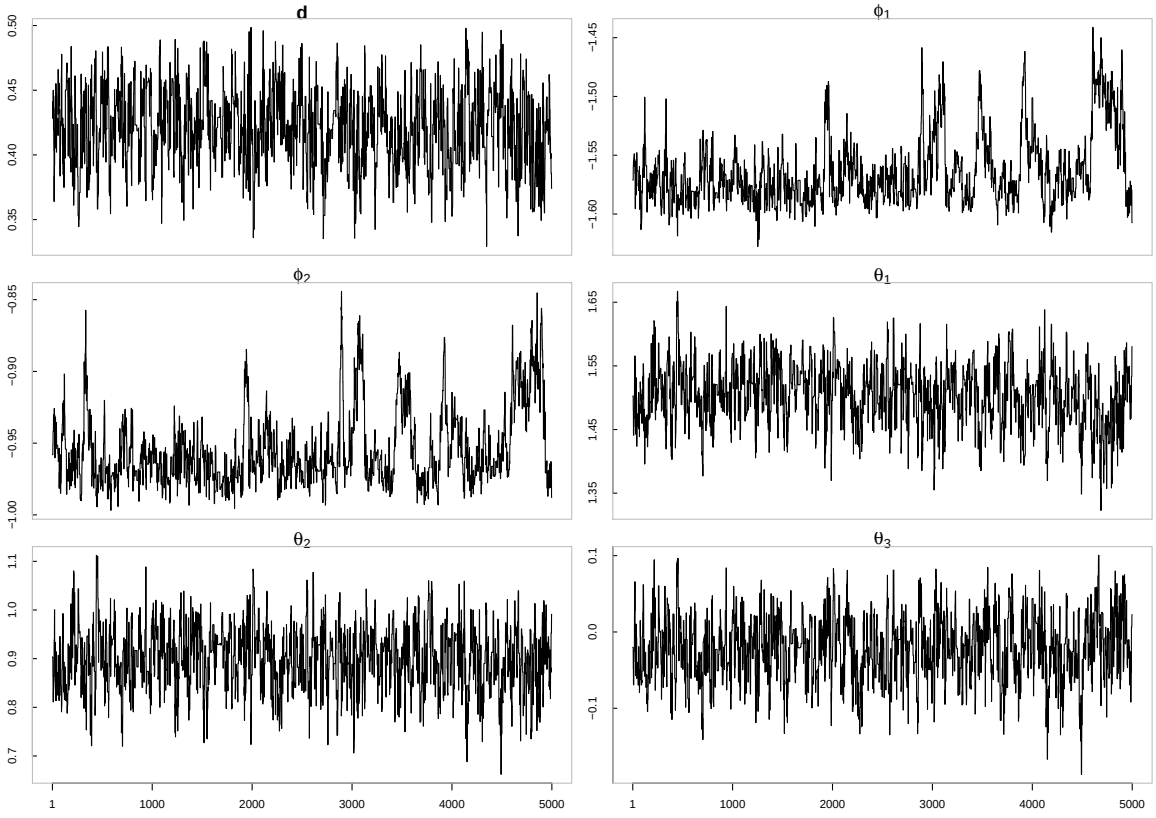
In the following sections, we first discuss the estimation results from the full sample,

Figure 3.4: MCMC iterations of parameters of ARFIMA(5,d,0) model of full sample



[a]The plotted 5000 MCMC iterations are obtained after discarding the first 5000 runs.

Figure 3.5: MCMC iterations of parameters of ARFIMA(2,d,3) model of full sample



[a]The plotted 5000 MCMC iterations are obtained after discarding the first 5000 runs.

with 1436 daily observations. In order to compare the empirical results of a single model and those of a weighted average model, we focus on the aspects of posterior inference on the long memory parameter d and out-of-sample predictive density, in Sections 3.4.3 and 3.4.4 respectively. A recursive estimation is also conducted to compare the selective forecasting performance of the single model and weighed average model; this is the topic of Section 3.4.5.

3.4.3 Memory Parameter Estimates and Posterior Model Probability

The posterior model probabilities and AIC are given in Table 3.1, and posterior means and standard deviations of the memory parameter are given in Table 3.2. In terms of posterior model probability, the evidence supporting long memory specifications is overwhelming, with 96% model probability for ARFIMA(p,d,q) models compared with 4% for ARMA(p,q) models, which is consistent with the ACF plots and the modified rescaled range statistics. Among the long-memory models, ARFIMA(2,d,3) receives the largest posterior model probability with 58%, which is also the best fitted model as indicated by AIC. The remaining posterior model probability is shared fairly evenly among ARFIMA(1,d,0), ARFIMA(0,d,1), ARFIMA(1,d,1), and ARFIMA(2,d,1) models, indicating comparable fits for these models. As to seasonality, the posterior mean of the AR coefficient at lag 5 is positive, but of fairly small magnitude as shown in Figure 3.4, which is also confirmed by the corresponding zero posterior model probability.

Table 3.1 highlights the usefulness of BMA in model choice. According to the classical approach, one might choose a model according to an information criterion, such as AIC. In this example, AIC does pick up the model with highest posterior model probability, however, it is not as informative as the posterior model probability in describing the relative merits of alternative models. For example, we found significant evidence supporting long memory specification according to posterior model probability, and yet the AIC of short memory models, with higher values of p and q, tend to give comparable and even smaller AIC values than those of their long memory counterparts. Besides this, while ARFIMA(2,d,3) appears to be the best fit under consideration, one still can not

completely rule out the other models given that the other long memory models account for 38% posterior model probability, and 4% posterior model probability is allocated to short memory specifications.

Table 3.1: Posterior model probabilities and AIC of full sample

p,q	ARFIMA(p,d,q)		ARMA(p,q)	
	Posterior Model Probability	AIC	Posterior Model Probability	AIC
0,0	0.0328	1.0965	—	—
1,0	0.0585	1.0939	0.0000	1.2263
2,0	0.0050	1.0952	0.0000	1.1490
3,0	0.0012	1.0948	0.0000	1.1346
4,0	0.0001	1.0961	0.0000	1.1176
5,0	0.0000	1.0961	0.0000	1.1067
0,1	0.0597	1.0940	0.0000	1.4046
1,1	0.0748	1.0950	0.0001	1.1025
2,1	0.0822	1.0963	0.0040	1.0935
3,1	0.0021	1.0962	0.0035	1.0910
0,2	0.0049	1.0953	0.0000	1.2820
1,2	0.0175	1.0963	0.0007	1.0960
2,2	0.0396	1.0971	0.0043	1.0920
3,2	—	—	0.0068	1.0919
0,3	0.0012	1.0948	0.0000	1.2457
1,3	0.0029	1.0960	0.0058	1.0903
2,3	0.5764	1.0757	0.0060	1.0917
3,3	—	—	0.0099	1.0931
Total	0.9590		0.0410	

[a]MCMC convergence of ARFIMA(3,d,2) and ARFIMA(3,d,3) is poor, so we exclude both models from our weighted average model.

The posterior means and standard deviations of the memory parameter d in Table 3.2 do not differ greatly across long memory models with different short-term dynamics, except for the case with no AR and MA terms, where the memory parameter is smaller, which might occur because the memory parameter is forced to capture the short term dynamics missing in the model specification. As to posterior inference on the parameters of interest, we believe that BMA provides a sensible way to find model weights. In terms of the burden of computing, the weight of each individual model determined by the marginal likelihood is simply the byproduct from MCMC output, so does not impose any additional computational burden.

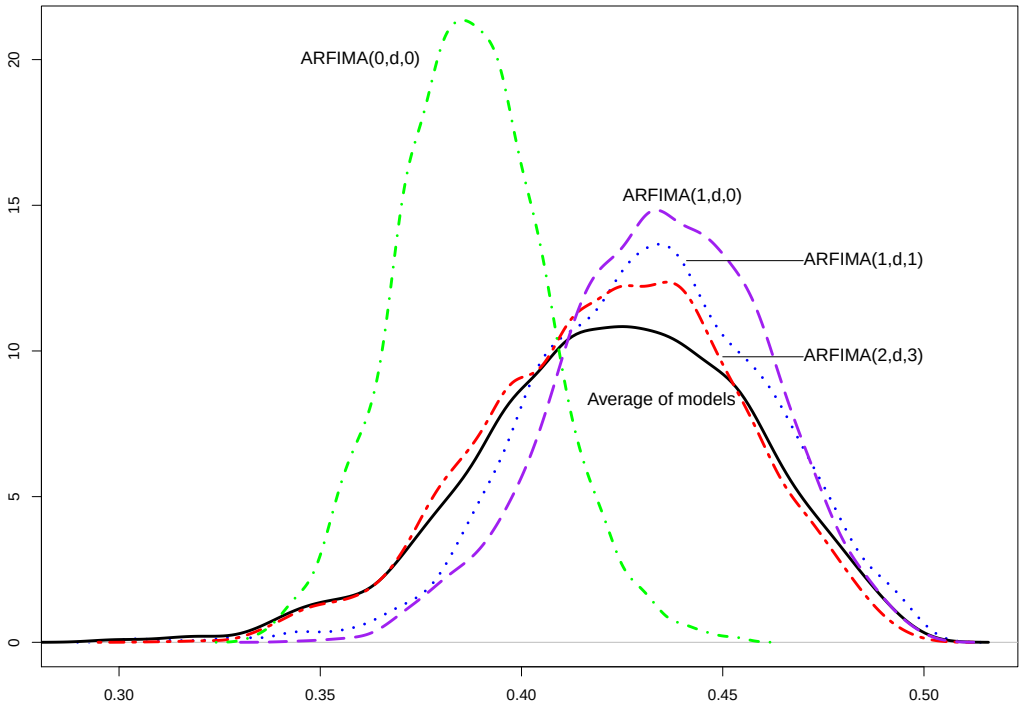
Table 3.2: Posterior mean and standard deviation of the memory parameter d for log realized volatility with full sample

p,q	Mean	Standard deviation
0,0	0.3873	0.0186
1,0	0.4349	0.0253
2,0	0.4278	0.0332
3,0	0.4441	0.0311
4,0	0.4412	0.0362
5,0	0.4169	0.0454
0,1	0.4382	0.0270
1,1	0.4316	0.0294
2,1	0.4196	0.0426
3,1	0.4445	0.0307
0,2	0.4294	0.0325
1,2	0.4355	0.0311
2,2	0.4212	0.0368
0,3	0.4491	0.0300
1,3	0.4447	0.0323
2,3	0.4223	0.0309
Weighted average	0.4064	0.0899

[a]MCMC convergence of ARFIMA(3,d,2) and ARFIMA(3,d,3) is poor, so we exclude both models from our weighted average model.

The posterior standard deviation of d weighted across 33 models is 0.0899, which is roughly 3 times that of each individual model. This is as expected, since a single model does not take account of model uncertainty and the precision of parameter estimates tends to be overstated. Figure 3.6 plots the posterior density of d from the weighted average and a selection of single models, thus highlighting the difference. It is clear from the density plots that the posterior inference of the memory parameter d differs substantially across models. The posterior mean of the weighted average model is 0.4064, shrinking towards zero relative to the dominant long memory specification, ARFIMA(2,d,3), due to the 4% posterior model probability of $d = 0$. As we know, to specify both short term and long term dynamics accurately is not always an easy task, so it might be risky to make an inference about d based on a single model, thereby ruling out the possibility of other short term dynamics.

Figure 3.6: Posterior density of memory parameter d for log realized volatility with full sample



3.4.4 Predictive Density

When evaluating forecasting performance, we follow the common approach of treating daily realized volatility as observed/actual of volatility despite the fact that volatility is essentially unobserved. The log volatility forecast from each MCMC iteration is back transformed into the original units, demonstrating the flexibility of MCMC in dealing with bias caused by transformation. In a classical setting, the back-transformed statistics, such as forecasting interval, might not provide unbiased estimates. Besides this, prediction intervals are readily available since the predictive density can be constructed from MCMC outputs.

Different forecasting horizons are examined. The one, five, ten and twenty-step ahead predictions of single and averaged models are generated, roughly representing short-term, medium-term and long-term forecasting. Since the data frequency is daily, five and twenty-step correspond to weekly and monthly horizons respectively. The predictive densities of different horizons constructed at the end of the sample are plotted in Figure 3.7, with the

posterior model probabilities taken from Table 3.1. For the ease of notation, let WAM, WAM_{ARFIMA} and WAM_{ARMA} denote models averaged across overall 33 models, long memory models and short memory models respectively.

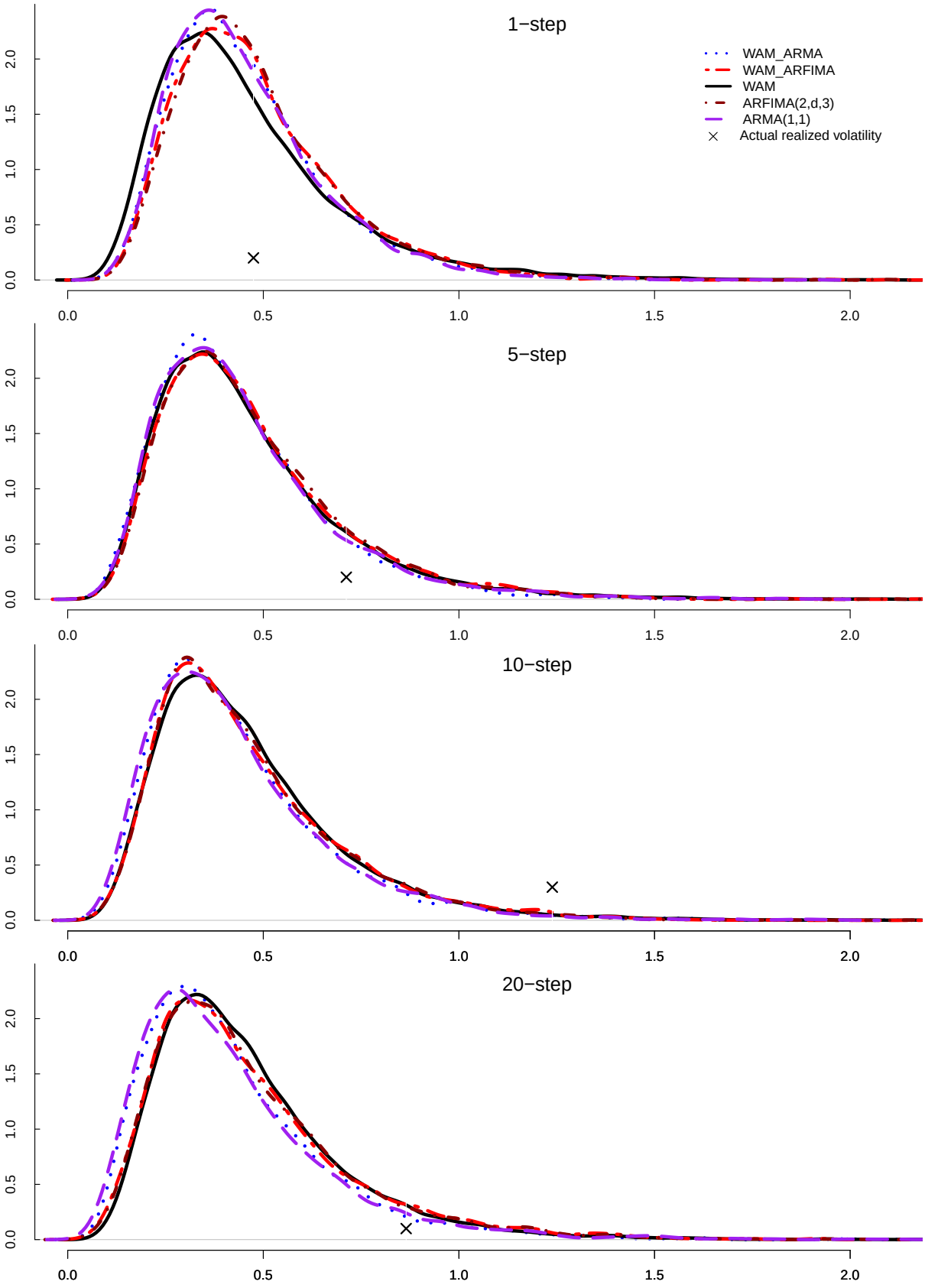
The main pattern in Figure 3.7 can be summarized as follows: as forecasting horizon increases, predictions from WAM, WAM_{ARFIMA} and ARFIMA(2,d,3) become very similar, but disparity between long memory and short memory models increases. Similarity exists between WAM_{ARFIMA} and ARFIMA(2,d,3) on all horizons examined. The predictive density plots of WAM_{ARMA} and ARMA(1,1) are also similar. For daily prediction, the WAM provides the largest standard deviation, indicating a wider prediction interval than a single model. At other horizons, the standard deviation of predictive density of WAM is also the largest, but it is not as obvious as in the case of daily horizon. To examine prediction accuracy, we also mark the observed realized volatility on the day being predicted in the plot. It is clear that volatility prediction becomes more difficult as horizon increases. If the mean of the predictive density is used as a point forecast, the distance between the predicted and the observed generally increases as the forecast horizon increases. For the four observed volatilities examined, ARFIMA(2,d,3) provides the most accurate one-step prediction, WAM is the best at ten and twenty-step horizons, and performance of each model at five-step is close. Short memory specifications of ARMA(1,1) and WAM_{ARMA} are always out-performed by others.

The comparisons in this section are based on estimation with the full sample, which might be risky for reliable statistical inference since the magnitude of prediction error is also determined by the actual realization, not just model specification. So it would be desirable to evaluate the average forecasting performance using recursive estimation, which is the focus of next section.

3.4.5 Recursive Estimation Results

To evaluate forecasting performances of averaged and single models, we split the whole data set into estimation and validation sets. Rolling sample estimation and prediction are conducted from the 1336th to the 1436th observations, corresponding to the period from

Figure 3.7: Predictive densities of realized volatility of full sample size



[a]The estimates obtained here are based on the posterior model probabilities in Table 3.1, with sample size being 1436.

10/04/2008 to 27/08/2008. For each rolling sample, the sample size is updated with one more observation, the 33 models are estimated, and the corresponding posterior model probabilities and multiple-step predictions are generated. The convergence of some models is not always good, especially the specifications of ARFIMA(2,d,1) and ARFIMA(2,d,2). As a result, both models are excluded from this exercise.

Figure 3.8: Mean squared forecast error of log realized volatility of recursive estimations

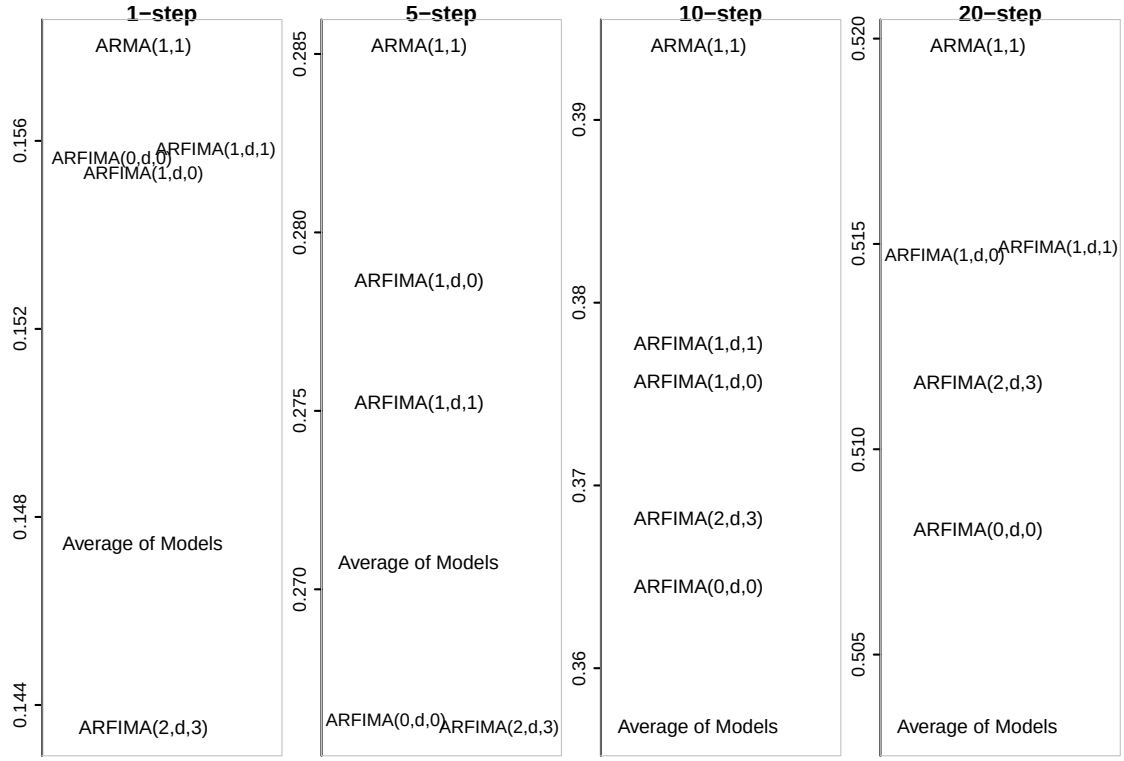


Table 3.3: Mean squared forecast error of log realized volatility of recursive estimations

	One-step	Five-step	Ten-step	Twenty-step
$ARMA(1,1)$	0.1558	0.2852	0.3940	0.5198
$ARFIMA(0,d,0)$	0.1556	0.2663	0.3644	0.5080
$ARFIMA(1,d,0)$	0.1553	0.2786	0.3756	0.5147
$ARFIMA(1,d,1)$	0.1558	0.2752	0.3777	0.5149
$ARFIMA(2,d,3)$	0.1435	0.2661	0.3681	0.5116
Average of models	0.1474	0.2707	0.3567	0.5032

[a]The $MSFE$ is averaged over the recursive estimation period from 10/04/2008 to 27/08/2008.

The posterior model probabilities of each model are averaged over the recursive estimation period, and the average weights for some models are as follows. The specifications

of ARFIMA(0,d,0), ARFIMA(1,d,0), ARFIMA(1,1), ARFIMA(2,d,3) and WAM_{ARMA} are allocated 4%, 10%, 15%, 37% and 13% model weights respectively. The ranking is largely in line with that in Table 3.1, although more weight is allocated to ARFIMA(1,d,1) and the weighted average of short memory processes. The combination of ARFIMA(2,d,3) and ARFIMA(1,d,1) accounts for more than half the model probabilities.

Mean squared forecast errors (MSFE) of some models are provided in Figure 3.8, with the mean of the predictive density used as the point forecast. The plot is the visual representation of the tabulation in Table 3.3 with MSFE on the vertical axis. In terms of magnitude, MSFE does not differ significantly across the single ARFIMA models and the weighted average model, indicating a relatively large proportion of variation caused by noise in realized volatility. The performance of ARMA(1,1) is clearly always the worst. In terms of ranking, the weighted average is the best at ten and twenty-step horizons, and the second best at one-step horizon. At five-step horizon, ARFIMA(2,d,3) and ARFIMA(0,d,0) perform almost the same, followed by the weighted average. The ranking is largely consistent with the case when predictions are made from the full sample as discussed in the previous section. There is no single model always performing the best, with ARFIMA(2,d,3) being the best at short horizons, and the performance of the parsimonious model ARFIMA(0,d,0) improves as the horizon increases. Overall, the weighted average appears to do a better job than single models. Since the ranking of a single model changes when forecasting horizon varies, inference relying on one particular model appears to be risky.

According to the empirical exercise conducted in this section, weighted average forecasting shows potential to improve accuracy at a long-term horizon, and is outperformed only by the ARFIMA model with the largest posterior model probability at daily (short-term) horizon. Also, the magnitude of MSFE does not appear to differ considerably across a single ARFIMA model and the weighted average of models. It would be useful to determine whether this observation is robust to different data sets, hence some simulation results are to be discussed in the following section.

3.5 A Simulation Study

The purpose of the simulations conducted in this section is to gauge the forecasting performance of a weighted average model compared to a single ARFIMA model. The simulation is designed as follows: the single ARFIMA model with relatively large model probability according to the recursive estimations reported in Section 3.4 is chosen as the data generating process. This includes ARFIMA(2,d,3), ARFIMA(1,d,1) and ARFIMA(1,d,0). Besides these, ARFIMA(0,d,0) is also considered for its property of parsimony. For each sample simulated, these four single models and the weighted average model are estimated, with corresponding single and multi-step forecasts generated. The mean of the prediction density is used as the point forecast, as described in the last section.

To be comparable with the empirical exercise in the previous section, recursive estimation is conducted in a similar fashion. In particular, given a data generation process, a sample of 1500 observations is simulated. Then recursive estimations from the 1336th to the 1436th observations are conducted for each model fitted, including the true model, the mis-specified models and the weight average model. The mean squared forecast error is recorded for each model fitted, meaning five measures of MSFE are obtained for one set of simulated data. Of each data generation process, 50 replications are carried, which results in $50 \times 4 = 200$ replications in total. The choice of number of replications is largely based on computational cost: about 22 000 hours computation is involved with 200 replications.[5] The parameters used to generate the ARFIMA models are taken from the average of the posterior mean over the recursive estimations in the previous section as[6]

$$
\begin{aligned}
ARFIMA(0,d,0) &: \quad d = 0.39 \\
ARFIMA(1,d,0) &: \quad d = 0.44, \phi_1 = -0.09 \\
ARFIMA(2,d,3) &: \quad d = 0.43, \phi_1 = -1.05, \phi_2 = -0.72, \theta_1 = 0.97, \theta_2 = 0.68, \theta_3 = -0.06.
\end{aligned}
$$

[5] All the simulations and computations in this section are run on the *Monash Sun Grid (MSG)*, which facilitates parallel computing. It takes about 360 hours to obtain the results discussed in this section.

[6] We also experiment with the DGP of ARFIMA(1,d,1), with d=0.40, ϕ_1= 0.07, θ_1=-0.12. For the corresponding samples generated by this process, the convergence of MCMC is poor when the specifications of ARFIMA(2,d,3) is fitted into the data. Accordingly, this case is excluded from our discussion. The specification of ARFIMA(2,d,3) is $(1+1.05B^{-1}+0.72B^{-2})(1-B)^{0.43}(y_t-\mu) = (1+0.97B^{-1}+0.68B^{-2}-0.06B^{-3})\eta_t$. This is a (covariance) stationary process because the inverted AR roots are $-0.52+0.67i$ and $-0.52-0.67i$, which lie inside the unit circle.

The unconditional mean of each simulated process is assumed to be zero, and the standard deviation of the innovation is the same for each process, with $\sigma_\eta = 0.42$. Short memory processes are not considered here, largely due to the computational cost. Also, the relatively small posterior model probabilities of ARMA specifications in Table 3.1 and obvious poorer forecasting performance as shown in Figure 3.8 indicate that it could suffice to focus on long memory specifications in this section.

Table 3.4: Average of MSFE using simulated data

		One-step	Five-step	Ten-step	Twenty-step
DGP: ARFIMA(0,d,0)					
Fitted model:	ARFIMA(0,d,0)	0.1760	0.2361	0.2530	0.2670
	ARFIMA(1,d,0)	0.1762	0.2369	0.2536	0.2678
	ARFIMA(1,d,1)	0.1761	0.2375	0.2547	0.2681
	ARFIMA(2,d,3)	0.1767	0.2384	0.2562	0.2703
	Average of models	0.1761	0.2364	0.2522	0.2656
DGP: ARFIMA(1,d,0)					
Fitted model:	ARFIMA(0,d,0)	0.1829	0.2424	0.2649	0.2869
	ARFIMA(1,d,0)	0.1826	0.2414	0.2639	0.2849
	ARFIMA(1,d,1)	0.1832	0.2426	0.2665	0.2885
	ARFIMA(2,d,3)	0.1836	0.2435	0.2675	0.2911
	Average of models	0.1829	0.2417	0.2644	0.2857
DGP: ARFIMA(2,d,3)					
Fitted model:	ARFIMA(0,d,0)	0.8950	1.1172	1.0962	1.0785
	ARFIMA(1,d,0)	0.7089	1.1121	1.0921	1.0754
	ARFIMA(1,d,1)	0.5328	1.0908	1.0854	1.0707
	ARFIMA(2,d,3)	0.3360	0.8910	1.0270	1.0671
	Average of models	0.3360	0.8910	1.0270	1.0671

[a]The parameters used to generate the ARFIMA(0,d,0) process is d = 0.39.

[b]The parameters used to generate the ARFIMA(1,d,0) process is d = 0.44 and $\phi_1 = -0.09$.

[c]The parameters used to generate the ARFIMA(2,d,3) process is d = 0.43, $\phi_1 = -1.05$, $\phi_2 = -0.72$, $\theta_1 = 0.97$, $\theta_2 = 0.68$ and $\theta_3 = -0.06$.

[d]The unconditional mean of each process is zero, and the standard deviation of the innovation σ_η is equal to 0.42.

[e]Each entry in the table is the average of MSFE over 50 replications.

The main results are summarized in Table 3.4, where each entry is the average of MSFE over 50 replications for one particular data generating process. The corresponding graphical representations are in Figures 3.9, 3.10 and 3.11. Some observations from the table can be summarized as follows.

The true model is usually the best of the candidate models. However, the weighted

Figure 3.9: Average of MSFE with ARFIMA(0,d,0) being the data generation process

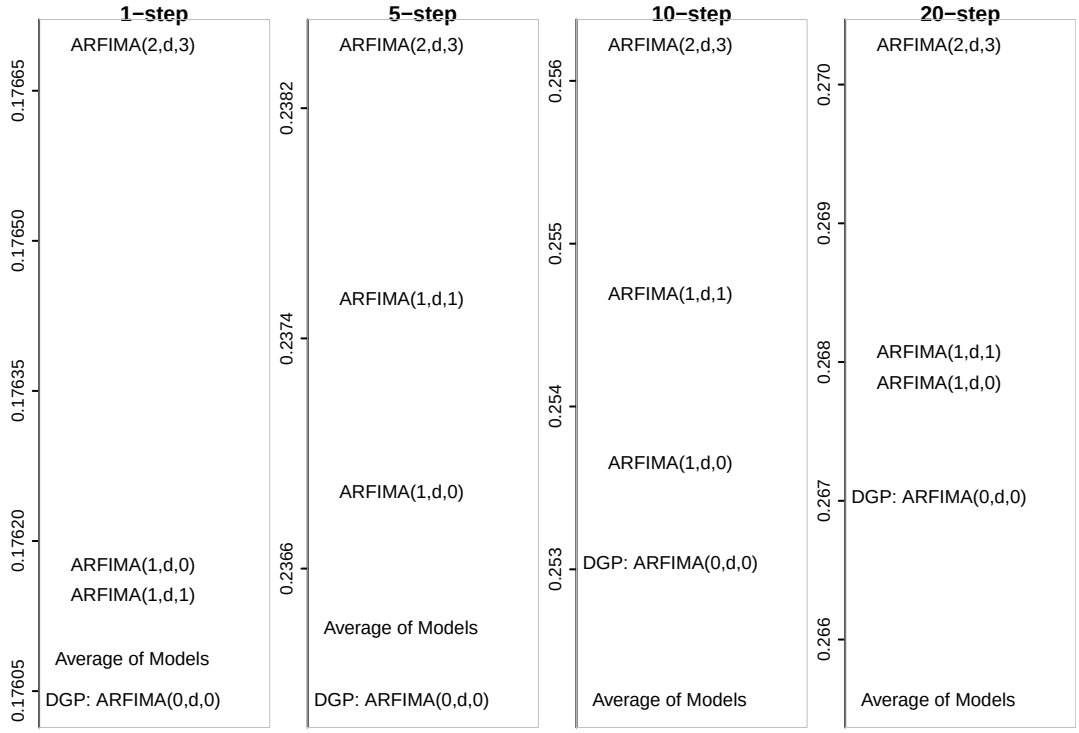


Figure 3.10: Average of MSFE with ARFIMA(1,d,0) being the data generation process

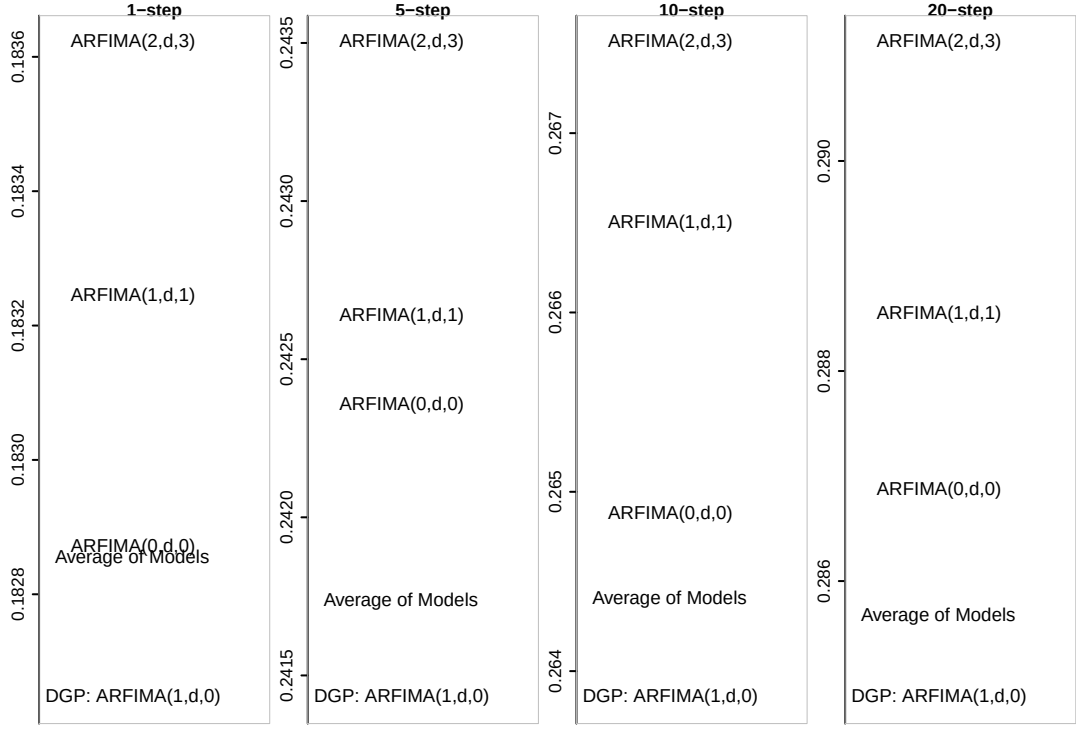
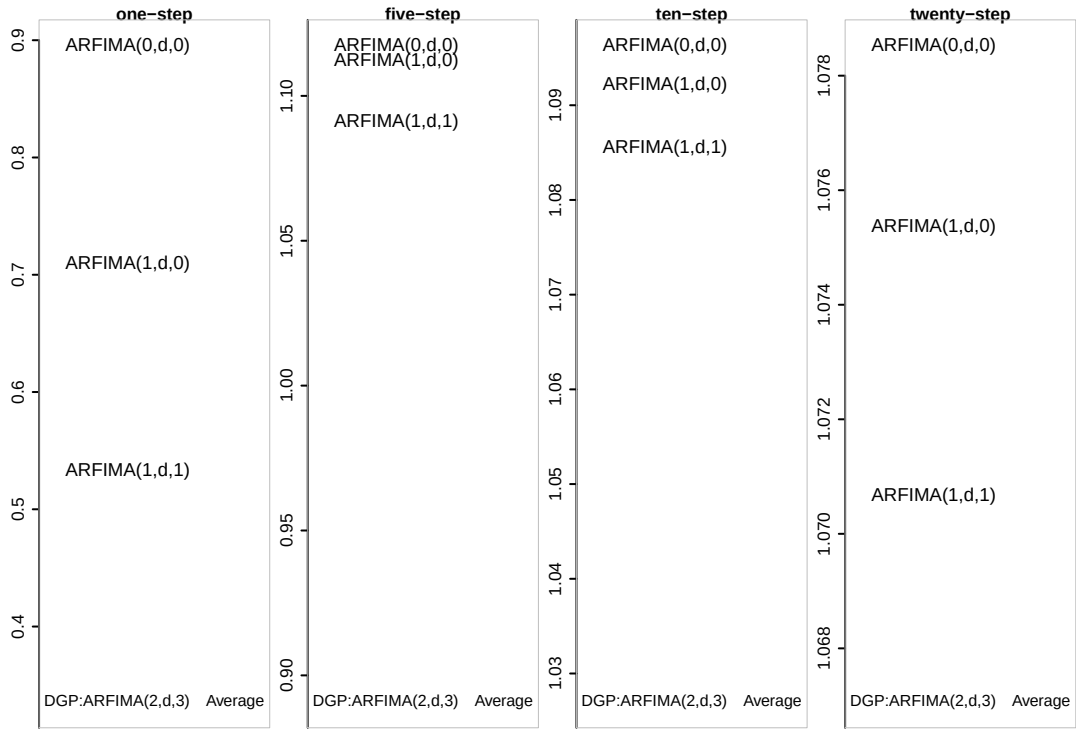


Figure 3.11: Average of MSFE with ARFIMA(2,d,3) being the data generation process



average model has the ability to pick up the best model, by attaching most weight to the true model in the weighting process. As a result, the MSFE of weighted average is always close to that of the true model. For example, in the case when ARFIMA(2,d,3) is the true model, the posterior model probability allocated to ARFIMA(2,d,3) is so close to one that the weighted average forecast coincides with that of the ARFIMA(2,d,3) model. The weighted average forecasting even slightly out-performs the true model at the horizons of ten and twenty-step when ARFIMA(0,d,0) is the underlying true process. The performance of an over-simplified model in dealing with the data generated by a more complicated model is rather poor for short-term forecasts, as indicated by the fairly large gap in MSFE between them. The largest gap is observed when ARFIMA(0,d,0) model is fitted to the data generated by ARFIMA(2,d,3) model for one-step ahead forecast.

The simulation conducted in this section assumes the DGP is one single model. For the parameters considered, the weighted average model does a reasonably good job in forecasting at different horizons when mean squared forecasting error is treated as the loss function. To sum up, both our empirical example and simulation study indicate

the usefulness of Bayesian model averaging in improving forecasting accuracy at different horizons.

3.6 Conclusions and Extensions

In this chapter, we conduct a Bayesian model averaging exercise on daily realized volatility of the Yen/USD exchange rate in the period from January 2003 to August 2008. We found significant evidence of long memory in volatility when an overall model averaged across a number of ARFIMA(p,d,q) and ARMA(p,q) specifications are considered. The increased uncertainty of the posterior density for the averaged memory parameter, according to the weights determined by posterior model probabilities, indicates the robustness to model mis-specification of the averaged model compared with a single model. Statistical inference obtained from competitive single models tends to provide different pictures of the posterior density, indicating it is risky to rely on a single model specification. In terms of forecasting, the average of models provides the most accurate forecasting at long-term horizons, and it is out-performed only by the single best long memory model at the daily horizon. No single model remains the best across different horizons considered. A simulation study is conducted and the results also indicate the risk reductions that can be achieved by Bayesian model averaging. These observations imply that it is worth the extra effort to deal with an average of models than a single model.

The models considered in this chapter only deal with ARMA and ARFIMA time series specifications; an extension of this approach could involve incorporating the feature of jumps commonly observed in high frequency data (see for example, Barndorff-Nielsen and Shephard (2004*b*, 2006)). With this different set of models to be averaged, the performance of model averaging may be quite different; this is one direction for future study.

Chapter 4

Bayesian Estimation of a Long Memory Stochastic Volatility Model

4.1 Introduction

A large contribution to the topic of volatility modeling has been produced by stochastic volatility models, where volatility is treated as a latent (state variable), and evolves according to an assumed data generating process. Under this approach, volatility is unobservable, but can be extracted from a model of asset returns. The basic stochastic volatility (SV) model of Taylor (1986) assumes volatility follows an autoregressive (AR) process of order one, and the autoregressive coefficient is often constrained to be less than one, to ensure stationarity of volatility. If the coefficient is on the unit boundary, the model is a random walk SV process. Extensive reviews of SV models can be found in Taylor (1994), Ghysels et al. (1996) and Shephard (1996). Two extreme values of the autoregressive coefficient are zero and one, with the former suggesting a volatility process with no memory and the latter implies that a shock exerts its effect permanently and volatility is not mean-reverting. Except for these two cases, the strength of persistence in volatility is determined by the magnitude of the AR coefficient. Most empirical work on stock and exchange rate returns has found the coefficient to be close to one. In such cases the SV model appears to resemble a GARCH(1,1) process with the sum of coefficients

close to one. According to the auto-covariance function of a stationary AR process, the effect of shocks or innovations will die out at an exponential rate, although much empirical evidence indicates that volatility might possess a longer memory than that described by an exponentially decaying auto-covariance function. For example, see Ding et al. (1993), Breidt et al. (1998), Bollerslev and Wright (2000) and Andersen, Bollerslev, Diebold and Labys (2001); Andersen et al. (2003). Some of these papers use volatility approximations, such as squared, log squared or absolute daily returns, and others use realized volatility constructed from intra-day returns. Regardless of the way volatility is measured, results in these papers suggest that volatility is mean-reverting and highly persistent with autocorrelations decaying to zero slowly. In other words, volatility appears to have long memory rather than the short memory characteristic suggested by an autoregressive SV model.

In time series modeling, fractional integration processes have long been employed to represent long memory, beginning with Hurst (1951) and Granger (1966), with the latter focusing on economic variables. The extension of short memory SV to long memory SV models is first proposed by Harvey (1998) and Breidt et al. (1998). Despite the different rates of decay of autocorrelations captured by short and long memory specifications, transitions from stationarity to non-stationarity described by the two processes proceed in different ways (see Harvey (1998)). Alternative specifications of long memory include the fractionally integrated GARCH model in Baillie et al. (1996) and the long memory moving average model in Robinson (2001).

In this chapter, we focus on the models proposed in Harvey (1998) and Breidt et al. (1998) for their natural conjunction to SV models. Our aim is to provide a Bayesian treatment and compare the performance of the Bayesian estimators with that of their classical counterparts. The classical treatment of long memory SV models is extensive, and includes Bollerslev and Wright (2000), Sun and Phillips (2004), Arteche (2004) and Andrews and Sun (2004) for semiparametric approaches. For parametric estimation, the frequency domain method is employed by Harvey (1998) and Breidt et al. (1998), and the time domain method by Ferraz and Hotta (2007). Semi-parametric methods are robust to mis-specifications in short-term dynamics of volatility, although they rely heavily on volatility approximations, such as log squared returns or power transformations of absolute

returns (see Ding et al. (1993)), which are a noisy approximation in some cases. Less work has been done on Bayesian inference for long memory SV models: to our knowledge, the only references are Jensen (2000) and Xu, Liu and Nie (2006) which are based on the frequency domain methods, and Chan and Giovanni (2000) which is based on time domain approximation, take a Bayesian approach.

Our decision to take a Bayesian approach is determined by the objectives of SV modeling, particularly where latent volatility is of primary interest. The Bayesian methodology facilities data augmentation by treating volatility as a parameter to be estimated, and posterior simulation methods, such as Markov chain Monte Carlo (MCMC) simulation, provide the potential to obtain posterior estimates of volatility by integrating out other parameters. For classical inference, volatility estimation is usually obtained conditional on parameter point estimates. To integrate out the inherent uncertainty in point estimates is not an easy task for the classical method. Our Bayesian estimation is based on a state space representation of the SV model, with fractional Gaussian noise being approximated by a moving average process of high order. This treatment is close to that of Chan and Giovanni (2000) but differs in the choice of prior and implementation of simulation algorithms. We propose to update the parameters of interest in one block, rather than individually, to improve efficiency in estimation, as it is well recognized that a block updating scheme is helpful in variance reduction. In particular, a hessian matrix, obtained with quasi maximum likelihood estimation, is used to construct the proposal density for the random walk Metropolis-Hastings algorithm, to facilitate MCMC simulation. A Monte Carlo simulation study indicates the performance of the proposed estimator is comparable to that of classical methods. With small to medium sample sizes, our estimator outperforms the classical estimates in terms of root mean square error for most of the parameter settings considered.

In SV models, part of the difficulty in estimation arises because volatility is latent (see the review in Section 2.4.2). For the long memory SV model considered in this chapter, log squared returns are treated as a long memory process with additive noise.[1] There are

[1]Hsu and Breidt (2003) conducts Bayesian estimation of fractionally integrated autoregressive and moving average models with additive noise. Their Bayesian estimation is based on importance sampling.

two series of random noise to be dealt with under this setup, one is the innovation to latent volatility and the other is the noise in volatility measurement. The ratio of the variances of the two random terms, *signal-to-noise ratio*, is expected to have an effect on estimation accuracy. So it is desirable to address this question in a Monte Carlo study. As part of our Monte Carlo exercise, a sensitivity analysis is conducted to examine the impact of varying parameter values on estimation, the result of which is useful for empirical applications of the long memory SV model. We argue that the parameter settings in our Monte Carlo study are of more practical relevance than previous research. For example, the magnitude of the innovation to volatility is chosen as 1.0 in Breidt et al. (1998), which is much higher than in most empirical findings. A magnitude of less than 0.5 is not uncommon in practice, and it is important to select realistic values for parameters when a Monte Carlo study is conduced.

An important issue in volatility modeling is the choice of sampling frequency, since prices of financial assets are available at different frequencies, say from tick-by-tick to yearly. This chapter only considers daily frequency. Most discrete-time SV models are not closed in the sense that given the same asset, the model estimates of say daily returns are different from those of the corresponding weekly returns.[2] We follow the discussion in Meddahi and Renault (2004): the long memory SV model is specified for daily frequency by implicitly assuming that it is the appropriate model for this frequency. This assumption is consistent with stylized facts of returns observed at daily frequency, and also is the common approach in practice. Also, to examine a series spanning over almost three decades, as in our empirical exercise in this chapter, would be infeasible for intra-day returns.

The plan for the remainder of this chapter is as follows. In Sections 4.2 and 4.3 we discuss the estimation of long memory SV model, from the perspectives of both classical and Bayesian methods. In Section 4.4, a Monte Carlo simulation experiment is conducted

[2]As discussed by Bollerslev and Wright (2000), Meddahi and Renault (2004) and references therein, SV model is not closed under temporal aggregation in the sense that volatility observed at different frequencies might not be characterized by the same parametric class. According to the definition given in Bollerslev and Wright (2000), a parametric class of time series model is said to be closed under temporal aggregation if a parametric model from the same class, but with different parameter values, characterizes the data generating process across all observation frequencies. The ARMA class of model processes this property, while common GARCH and SV models are not closed under temporal aggregation. In general, we say that a model is closed under temporal aggregation if the model keeps the same structure, with possibly different parameter values for any data frequency (Meddahi and Renault (2004)).

to examine the performance of different estimators. Section 4.5 presents an example based on the daily Yen/USD exchange rate series. Our conclusions are found in Section 4.6.

4.2 Long Memory Stochastic Volatility Model

Section 4.2.1 discusses the specification of the long memory SV model. Sections 4.2.2 and 4.2.3 focus on the estimation of long memory SV models, from the perspectives of time domain and frequency domain respectively.

4.2.1 Specification of the Long Memory Stochastic Volatility Model

Let $\log(p)$ denote the log-price of asset, and the continuously compounded return at time t for $t = 1, \ldots, T$ is defined as,

$$y_t = 100 \times [\log(p_t) - \log(p_{t-1})] \tag{4.1}$$

where t is assumed to be of daily frequency. Volatility of asset returns is the concept used to measure the spread of all possible return outcomes. Both daily return and conditional volatility change over time. It is well recognized in the literature that volatility is predictable though daily return is unpredictable. Section 2.4.1.1 in Chapter 2 has discussed the basic SV model of Taylor (1982, 1986), reproduced here for convenience,

$$y_t = \sigma \exp(h_t/2)\varepsilon_t \qquad \varepsilon_t \sim NID(0,1) \tag{4.2}$$

$$h_t = \delta + \phi h_{t-1} + \eta_t, \qquad \eta_t \sim NID(0, \sigma_\eta^2) \tag{4.3}$$

The log volatility $h_t = \log(\sigma_t)$ is assumed to follow a first order autoregressive process. The log transformation is to ensure positivity of the conditional variance σ_t^2. Positive autocorrelation is one of the stylized facts of volatility, so the sample estimate of ϕ is usually found to be positive. Also, it is convenient to assume volatility is stationary by constraining $\phi < 1$ (see the discussion in Shephard (1996)). One might assume ε_t and η_t to be correlated in order to capture the so-call "leverage effect"; in this paper we assume

$Cov(\varepsilon_t, \eta_t) = 0$. This is consistent with the observation that leverage effect in exchange rate markets is not as pronounced as that in the stock market. The quantity of σ_η^2 is termed *volatility of volatility*. The literature on SV models is extensive; see the review in Section 2.4.2 in Chapter 2.

In order to incorporate the slowly-decaying autocorrelation in various volatility approximations, a natural extension to (4.3) is to assume log volatility follows a long memory process rather than a short memory process, which is found to be consistent with the empirical findings of slow-decaying autocorrelation of various volatility approximations. The basic long memory SV model proposed by Harvey (1998) assumes log volatility to be generated by fractional ARFIMA(0,d,0) process[3]

$$(1 - B)^d h_t = \eta_t \tag{4.4}$$

where B is lag operator, and d is the memory parameter. The condition for a fractional process to be stationary is $|d| < 0.5$. It is reasonable to further assume autocorrelations of volatility to be positive by the condition $0 < d < 0.5$, which is analogous to the stationarity assumption usually made in (4.3) where ϕ is restricted to be positive and smaller than one. When d is equal to zero, h_t is a white noise, representing a process with no memory. The degree of volatility persistence is suggested by the magnitude of the memory parameter d. Throughout, we impose the stationarity assumption on volatility by following Harvey (1998). Accordingly, the evolution of volatility is mean-reverting, which is consistent with the observations in empirical applications.

One of the appealing features of the long memory process in (4.4) is its parsimony. The autocorrelation function $\rho(k)$ indicated by the process decays slowly at a hyperbolic rate, asymptotically proportional to k^{2d-1} with k being the index of lags. The autocorrelation function decays so slowly that the autocorrelations are not summable in the sense that $\Sigma_{k=-\infty}^{\infty}\rho(k) = \infty$. To describe such a slowly decaying autocorrelation is a difficult task for ARMA(p,q) models because the process needs to be approximated by an ARMA(p,q) pro-

[3]Long memory has been observed on many economic time series, early evidence can be found in Granger (1966) who observed that for economic time series, the typical shape of the spectral density is (at least in good approximation) a function with a pole at the origin, which is a stylized feature of long memory process.

cess with high orders of autoregressive and moving-average terms, while the specification in (4.4) allows the slowly decaying autocorrelations to be fully determined by the memory parameter. See Beran (1994) for a text-book treatment of long memory processes.

Despite all these desirable properties, exact maximum likelihood estimation of the long memory SV model is not easy. The reasons are twofold. One is due to the nonlinearity of (4.2): this is a problem common to the short memory SV model as well. The other is due to the computational cost in evaluating the exact likelihood function of a long memory process.[4] The natural solution is via approximation.

Let us address the two difficulties in order. The nonlinearity in (4.2) can be dealt with by linearization as

$$\log(y_t^2) = \mu + h_t + \xi_t \tag{4.5}$$

where $\mu = \log(\sigma^2) + E[\log(\varepsilon_t^2)]$ with $E[\log(\varepsilon_t^2)] = -1.27$. The term of ξ_t is the square of Gaussian noise, which is non-Gaussian with zero mean and variance equal to $Var[\log(\varepsilon_t^2)] = \pi^2/2$.[5] To reduce the number of parameters involved in the numerical optimization, μ can be estimated from the sample mean of $\log(y_t^2)$. As argued by Yajima (1988), the loss in efficiency is small if the sample mean is used to estimate the unconditional mean.[6] In particular, let ω_t denote the demeaned series of $\log(y_t^2)$. The long memory SV model is then transformed into

$$\omega_t = h_t + \xi_t \tag{4.6}$$

$$h_t = (1-B)^{-d}\eta_t \tag{4.7}$$

The expected value of the noise term ξ_t is zero and the variance is $\pi^2/2$. However, the distribution of ξ_t is not normal. In the literature on SV model estimation, the ways to

[4] Long memory is a process with infinite memory, as a result of which the corresponding covariance matrix used to evaluate likelihood function is of order $T \times T$, with large sample size typically used in financial econometric applications, the burden on estimation is not trivial.

[5] Given normality assumption of ε_t, $\log(\varepsilon_t^2)$ follows log chi-square $\log(\chi_1^2)$ distribution, which is negatively skewed, with $E[\log(\chi_1^2)] = -1.27$ and $var[\log(\chi_1^2)] = \pi^2/2$.

[6] According to the simulation results of short memory SV model in Breidt and Carriquiry (1996), the reduction in number of parameters to be estimated also helps to improve the performance of numerical optimization.

handle this non-normality can be summarized as:

1. Approximate the distribution of ξ_t by a normal distribution with the same mean and variance. This is the method proposed in Harvey et al. (1994).

2. Use a mixture of normal distributions to approximate the distribution of ξ_t as in Kim et al. (1998).

3. Evaluate the exact likelihood of (4.2) via Monte Carlo simulation as in Durbin and Koopman (1997) and Jungbacker and Koopman (2007).

We take the first approach for two reasons: compared with the other two methods, which are relatively more complicated and computationally intensive, especially the method of likelihood simulation, the normal approximation can be easily implemented. Also, the normal approximation provides estimators which are consistent as shown by Harvey et al. (1994). We leave the relaxation of this normality assumption as the subject of future research.[7]

If the volatility is observable, or can be treated as observed as in the methodology taken by realized volatility, the method of Sowell (1992) can be applied to evaluate the exact likelihood function of the long memory process. However, this method is infeasible in the case of SV estimation since the long memory process of h_t is unobservable and it is only observed through the time series of ω_t. The solution to this problem is to approximate the likelihood function. The likelihood function of the long memory SV model can be approximated according to its frequency domain representation or its time domain representation. The estimation method proposed in Harvey (1998) and Breidt et al. (1998) are based on the frequency domain approximation, which is to be discussed in Section 4.2.2. Our Bayesian method is based on the time domain approximation with relevant details documented in Section 4.2.3.

[7]There is a large literature in the semi-parametric estimation of long memory SV models. One of the common approaches taken by semi-parametric estimation is to estimate the memory parameter of ω_t or $\log(y_t^2)$ by the method proposed by Geweke and Porter-Hudak (1983). One of the appealing features of the semi-parametric method is that it is robust to the deviation from the normality assumption of ξ_t. However, this method suffers from a drawback of underestimating the memory parameter as is well documented in the literature. We conduct a Monte Carlo exercise to gauge the performance of the GPH estimator when it is used to estimate the memory parameter of long memory SV model, the downward bias is obvious even when the sample size is reasonably large. The simulation results are available upon request.

In the following sections, the parameters of the long memory SV model to be estimated are denoted as $\beta = [d, \sigma_\eta^2]'$.

4.2.2 Frequency Domain QML Estimation of Long Memory SV Model

Spectral estimation is based on frequency domain quasi-maximum likelihood (QML) estimation proposed by Whittle (1953) and introduced in a long memory context by Fox and Taqqu (1986). The implementation of this method is computationally simple.

Given the stationarity assumption of the process, its autocorrelation function carries the same information as its spectral density. Accordingly, the likelihood function to be optimized can be constructed in the form of a spectral likelihood function in the frequency domain. As argued in Harvey (1998), the simplicity of this approach is a great advantage, in that QML estimation in the frequency domain is no more difficult than it is in the short memory SV case.

Based on discrete Whittle approximation, the frequency domain (quasi) log-likelihood function of ω_t is,

$$\log L_W(\beta) = -2\pi T^{-1} \sum_{j=1}^{T^\star} \{\log f(\lambda_{j,T}; \beta) - \frac{I(\lambda_{j,T})}{f(\lambda_{j,T}; \beta)}\} \tag{4.8}$$

where $T^\star$ denotes the integer part of $(T-1)/2$, and $I(\lambda_{j,T})$ is the normalized periodogram ordinate which can be calculated by the fast Fourier transformation. The term of $\lambda_{j,T}$ denotes the jth Fourier frequency as

$$\lambda_{j,T} = 2\pi T^{-1}, \qquad j = 1, \ldots, T^\star \tag{4.9}$$

and $f(\lambda_{j,T}; \beta)$ denotes the spectral generating function for ω_t with the form of

$$f(\lambda_{j,T}; \beta) = \sigma_\eta^2 [2(1 - \cos \lambda_{j,T})]^{-d} + \sigma_\xi^2 \tag{4.10}$$

where $\sigma_\xi^2 = \pi^2/2$. See Beran (1994, Ch. 6) for general justification of the discrete Whittle approximation.

The frequency domain QML estimators of β are given by the values maximizing (4.8) and numerical optimization is required. Breidt et al. (1998, Appendix A) provides the proof of strong consistency for spectral likelihood estimators, though the asymptotic distribution of the estimator is unknown.

4.2.3 Time Domain QML Estimation of the Long Memory SV Model

QML estimation in the time domain relies on the representation of (4.6) and (4.7) in the state space form (SSF). Under the setting of state space modeling, the series of h_t are treated as latent variables. Equation (4.6) is termed as the *observation equation*, or *measurement equation*, indicating the fact the latent variables cannot be observed directly, but need to be extracted from the measurable ω_t. A large body of the SV modeling literature is under the setting of the state space form. See, for example, the early work of Harvey et al. (1994) on QML estimation of the short memory SV model, and more recent work on Bayesian inference presented in Kim et al. (1998) with the applications of MCMC simulation.

For a long memory SV model, QML estimation in the time domain becomes relatively more complicated compared with that discussed in Section 4.2.2. This is because the SSF can only be used by expressing h_t as a moving average or autoregressive process truncated at a suitably high lag. As addressed by Harvey (1998), the computational cost of the time domain QML is increased, though the initial state covariance matrix can be easily constructed, and the truncation does not affect the asymptotic properties of the estimators. Given the rapid improvements in modern computing facilities, the increased computational burden introduced by the truncations at high lag can be easily dealt with. Therefore we discuss in details the estimation of QML in the time domain in this section. Our simulation results to be presented in Section 4.4 indicate that the time domain estimator's performance is comparable to that of the frequency method. In terms of computing time on numerical optimizations, no great differences are observed between the two methods. Another reason we discuss the time domain method is that it is the basis on which we evaluate the likelihood function in our Bayesian method.

4.2.3.1 The State Space Representation of Long Memory SV Model

The basic idea for the time domain method is to approximate the fractional Gaussian process in (4.7) by an ARMA representation. In particular, h_t can be represented as an infinite moving average process with coefficients explicitly determined by its memory parameter[8]

$$h_t = (1-B)^{-d}\eta_t = \sum_{i=0}^{\infty} \varphi_i B^i \eta_t \tag{4.11}$$

where the fractional differencing operator is defined through the formal binomial series expansion

$$(1-B)^{-d} = \sum_{i=0}^{\infty} \frac{\Gamma(i+d)}{\Gamma(i+1)\Gamma(d)} B^i$$

Accordingly,

$$\varphi_i = \frac{\Gamma(i+d)}{\Gamma(i+1)\Gamma(d)} = \prod_{0<l\leq i} \frac{l-1+d}{l}, \qquad i = 1, 2, \ldots, \tag{4.12}$$

with $\varphi_0 = 1$. If the value of d is fractional, the summation in (4.11) is genuinely over an infinite number of indices. It is infeasible to use $MA(\infty)$, so a natural solution is truncation at a suitably high lag. For a fractional process, according to the asymptotic properties proved by Chan and Palma (1998), the truncated lag does not have to be very high. Their simulation results indicate even an order less than 10 can achieve a reasonably accurate estimate for a long memory process. Hence it appears realistic to assume their theoretical results are applicable to the case when the long memory process is latent. Suppose (4.11) is truncated at lag m as

$$h_t = \eta_t + \varphi_1\eta_{t-1} + \varphi_2\eta_{t-2} + \ldots + \varphi_m\eta_{t-m} \tag{4.13}$$

[8]Fractional Gaussian noise can be equivalently expressed as an infinite autoregressive (AR) process. We conducted a small-scale Monte Carlo experiment and the results indicate that the moving average representation provides better estimates than the autoregressive case. This is consistent with the observations in Brockwell and Davis (1991, Ch. 12).

In this way, h_t is treated as a linear combination of past innovations. The difference between an ordinary moving average time series and h_t is that the latter is not observed directly, but observed with additive noise ξ_t. This provides a standard application of state space modeling framework. Join the measurement equation (4.6) with the MA approximation of h_t,

$$\begin{aligned} \omega_t &= h_t + \xi_t \\ h_t &= \eta_t + \varphi_1\eta_{t-1} + \varphi_2\eta_{t-2} + \ldots + \varphi_m\eta_{t-m} \end{aligned}$$

and the second equation is called the *state equation* as it specifies the evolution of the latent variable (or *state variable*) h_t. The two equations can be rewritten as[9]

$$\begin{aligned} \omega_t &= \begin{bmatrix} \varphi_m & \ldots & \varphi_2 & \varphi_1 & 1 \end{bmatrix} x_t + \xi_t \\ x_t &= \begin{bmatrix} 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \\ 0 & 0 & \ldots & 0 \end{bmatrix} x_{t-1} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \eta_t \end{aligned} \tag{4.14}$$

for $t = 1, 2, \ldots, T$. The state vector

$$x_t = \begin{bmatrix} \nu_{t-m} & \ldots & \nu_{t-2} & \nu_{t-1} & \nu_t \end{bmatrix}'$$

has m+1 dimensions. The state is initialized with x_0 following a normal distribution with zero mean and variance matrix of $\sigma_\eta^2 I_{m+1}$, with I_{m+1} denoting $(m+1) \times (m+1)$ identity matrix.[10]

Under the normality assumption of the error terms ξ_t and η_t, the state space model in (4.14) is linear and Gaussian, the likelihood evaluation of which is readily available by the

[9]The state space representation for a moving average process is not unique, see, for example, Brockwell and Davis (1991, Ch. 12) and Hamilton (1994, Ch. 13).

[10]Alternatively, one may truncate the differenced process $(1 - B)h_t$ rather than h_t as discussed in Chan and Palma (1998), since the moving average coefficients of which converges to zero faster than the coefficients of φ_i. We did not take this approach for the easy of interpreting parameters.

celebrated Kalman filter. The Kalman filter algorithms are provided in the next section for convenience.

4.2.3.2 Kalman Recursions

For most applications of state space modeling, the main focus is on statistical inference for the latent state. For example, the ultimate goal of most volatility modeling might be the volatility extraction and forecasting, because volatility is such an important input to financial applications. This provides an attraction of the time domain method over the frequency domain method. Volatility extraction and prediction can be conducted by routine exercises in state space models, which are the by-products of likelihood evaluation.[11]

To discuss the Kalman filter in a generic form, the state space form of the long memory SV model in (4.14) can be expressed as

$$\begin{aligned} \omega_t &= \Phi x_t + \xi_t, & \xi_t \sim NID(0, \sigma_\xi^2) & \quad \textit{observation equation} \\ x_t &= \Omega x_{t-1} + \Psi \eta_t, & \eta_t \sim NID(0, \sigma_\eta^2) & \quad \textit{state equation} \end{aligned} \tag{4.15}$$

where Φ is a row vector which is conformable to the state vector, and ξ_t is serially uncorrelated with $E(\xi_t) = 0$ and $Var(\xi_t) = \sigma_\xi^2$. The state variable x_t is unobservable, the evolution of which is characterized by the state equation where Ω is a square matrix conformable to the state vector, and Ψ is row vector with the same dimension as x_t. The state disturbance is serially uncorrelated with $E(\eta_t) = 0$, $Var(\eta_t) = \sigma_\eta^2$ and $E(\xi_t \eta_t) = 0$. To complete state space specification, one need to assume the initial state x_0 with $E(x_0) = m_0$ and $Var(x_0) = P_0$.

Suppose the aim of the analysis is to infer the properties of h_t from the data $\omega_{1,\ldots,T} = \{\omega_1, \omega_2, \ldots, \omega_T\}$ and the model. Let $\omega_{1,\ldots,t} = \{\omega_1, \omega_2, \ldots, \omega_t\}$ be the set of observations available at time t with $t < T$. Three types of state inferences are considered in state

[11] Several statistics and econometric software have add-in packages to conduct the state filtering and smoothing algorithms for linear and Gaussian state space models. For example, *EViews*, *Ox* and the free GNU project *R*.

space modeling:

- State filtering

  To filter means to uncover x_t conditional on $\omega_{1,\ldots,t}$. It is concerned with updating the knowledge of the state each time a new observation ω_t is brought in. The conditional distribution of the state vector is normal given linear and Gaussian assumptions. Thus it is fully characterized by its first two moments. At time $t-1$, the mean and covariance of the one-step ahead prediction for the state vector are

$$\begin{aligned} E(x_t|\omega_{1,\ldots,t-1}) &= x_{t|t-1} = \Omega x_{t-1}, \\ Var(x_t|\omega_{1,\ldots,t-1}) &= P_{t|t-1} = \Omega P_{t-1}\Omega' + \Psi\sigma_\eta^2\Psi' \end{aligned}$$

  assuming that the model is known, including all parameters. As a new observation ω_t becomes available, the mean and covariance matrix are updated as

$$\begin{aligned} E(x_t|\omega_{1,\ldots,t}) &= x_t = x_{t|t-1} + P_{t|t-1}\Phi' F_t^{-1}(\omega_t - \Phi x_{t|t-1}) \\ Var(x_t|\omega_{1,\ldots,t}) &= P_t = P_{t|t-1} - P_{t|t-1}\Phi' F_t^{-1}\Phi P_{t|t-1}, \text{ with } F_t = \Phi P_{t|t-1}\Phi' + \sigma_\xi^2. \end{aligned}$$

  Derivations of the above equations follow directly from standard results on the multivariate normal distribution. Kalman filtering is recursive in the sense that it runs for all observations through $t = 1, \ldots, T$. See Harvey (1989, pp. 105-106).

- State prediction

  Prediction is concerned with forecasting ω_{t+h} conditional on $\omega_{1,\ldots,t}$, for $h > 0$. At time $t-1$, the one-step ahead predictive distribution of ω_t is normal with mean $\hat{\omega}_{t|t-1} = \Phi x_{t|t-1}$ and covariance matrix F_t. The corresponding prediction error is $\upsilon_t = \omega_t - \hat{\omega}_{t|t-1}$, which is serially independent distributed as $\upsilon_t \sim NID(0, F_t)$.

  The likelihood function for the state space model is constructed based on the one-step ahead prediction errors as

$$\log L(\beta) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log F_t - \frac{1}{2}\sum_{t=1}^{T}\upsilon_t^2/F_t \tag{4.16}$$

and this is the objective function to be optimized for the QML estimates of β.

- State smoothing

  Smoothing means to estimate x_t given $\omega_{1,\ldots,T}$. A smoothing algorithm is a backward recursion. Given the filtered state vector, the smoothed state vector is obtained via a recursion run for $t = T, \ldots, 1$. See Harvey (1989, p. 154).

The algorithms discussed here appear cumbersome. However, the implementation of them is straightforward with add-on packages, such as the free *Ox* package *SsfPack* developed by Koopman, Shephard and Doornik (1998). In this paper, the likelihood evaluation is obtained via *SsfPack*.[12] Durbin and Koopman (2001*a*) provide a text book treatment of time series analysis by state space methods.

The limitation of Kalman filter is that it is only applicable to linear and Gaussian state space model. There is no easy generalization when the linearity and normality assumptions do not hold. A growing body of literature in state space modeling has been addressing non-linear and/or non-Gaussian applications, for example, the sequential Monte Carlo method of particle filtering. See the recent book-length treatment in Doucet, de Freitas and Gordon (2001). In our example, a possible extension can be made by dealing with

$$\begin{aligned} y_t &= \sigma \exp(h_t/2)\varepsilon_t \\ h_t &= \eta_t + \varphi_1\eta_{t-1} + \varphi_2\eta_{t-2} + \ldots + \varphi_m\eta_{t-m} \end{aligned}$$

where the state equation is linear while the observation equation is non-linear. In principle, the likelihood of this state space form can be exactly evaluated by a simulation method, such as the importance sampling method proposed by Sandmann and Koopman (1998) in the context of the short memory SV model.[13] However, the results we have obtained so far are not encouraging; this we leave for future research.

[12]The *SsfPack* package applies the smoothing algorithms proposed by de Jong and Shephard (1995) to obtain smoothed estimates of the state vector.

[13]Thanks to Koopman for providing us with the Ox code for their paper.

4.3 Bayesian Estimation

Bayesian inference is concerned with the posterior distributions of the parameters given the observed data. The analytical forms of posterior distributions are rarely available in most applications. Our Bayesian method is based on the state-space form of (4.15). In this case, the state space form appears linear and Gaussian, while the memory parameter enters the state equation in a non-linear fashion. Each of the moving average coefficients is a non-linear transformation of the memory parameter. These preclude the possibility of obtaining analytic forms of the joint posterior distribution. The solution is to draw a sample from the posterior distribution with a reasonably large sample size, rather than derive the closed distribution. Statistical inference is then made upon these large number of draws. For example, the sample mean can be treated as the posterior mean of the parameter of interest. The density estimates for the sample can be treated as an approximation to the posterior density of interest. One popular device for obtaining such large numbers of draws which can be treated as the sample from the posterior distribution is MCMC simulation.[14]

The basic strategy to conduct Bayesian state space modeling is through data augmentation (Tanner and Wong (1987)) by treating state variables as unknown parameters. Let $x = \{x_1, x_2, \ldots, x_T\}$ denote the matrix of state variables, each element of which denotes the state vector at time t for $t = 1, 2, \ldots, T$. Then the parameters to be estimated are augmented to x and β. The reason for the argumentation is that it facilitates the implementation of MCMC simulation, in particular, the Gibbs sampler (Gelfand and Smith (1990)).

According to Bayes' rule, the joint posterior distribution of x and β conditional on $\omega_{1,\ldots,T} = \{\omega_1, \ldots, \omega_T\}$ is

$$p(x, \beta|\omega_{1,\ldots,T}) \propto p(\omega_{1,\ldots,T}|x, \beta)p(x|\beta)p(\beta)$$

where $p(\beta)$ is the prior, defining the prior belief on the parameters before $\omega_{1,\ldots,T}$ are

[14] Gamerman and Lopes (2006) provides an excellent introductory to MCMC simulation.

available. The posterior simulation can be achieved by applying a Gibbs-based algorithm, sampling iteratively from the two conditionals, $p(x|\beta, \omega_{1,\ldots,T})$ and $p(\beta|x, \omega_{1,\ldots,T})$.[15] In particular, suppose the MCMC chain is initialized with $(\beta^{(0)}, x^{(0)})$. Given the MCMC output obtained at the *mth* iteration $(\beta^{(m)}, x^{(m)})$, we repeatedly cycle through the two simulation steps:

Step 1. generate $x^{(m+1)}$ from $p(x|\beta^{(m)}, \omega_{1,\ldots,T})$

Step 2. generate $\beta^{(m+1)}$ from $p(\beta|x^{(m+1)}, \omega_{1,\ldots,T})$

for $m = 1, 2, \ldots$. After discarding a large number of *burn-in* iterations the samples of x and β can be treated as being generated from the marginal posteriors of $p(x|\omega_{1,\ldots,T})$ and $p(\beta|\omega_{1,\ldots,T})$. The Gibbs sampler helps to reduce the parameter dimensions of the distributions to be dealt with by iterating between the two full conditionals. This also indicates the advantage of MCMC method: volatility estimates are the by-product of parameter estimation. The focus of some applications might be volatility rather than the parameters. For the classical method, however, the smoothed estimates of volatility are conditional on the QML point estimates. Inference based on $p(h_t|\omega_{1,\ldots,T})$ is less risky compared with the classical counterpart because the latter does not integrate out the uncertainty of parameter estimates.

4.3.1 Simulations from Conditionals

We now discuss how to simulate from the two conditionals. To simulate from the conditional $p(x|\beta, \omega_{1,\ldots,T})$ is a routine exercise in Bayesian state space modeling, because the state space form in (4.15) is linear and Gaussian conditional on the value of β. This can be done via the efficient *forward filtering, backward sampling* algorithm proposed by Carter and Kohn (1994) and Frühwirth-Schnatter (1994). The advantage of this algorithm is that the latent variables $(x_1, x_2, \ldots, x_T)$ can be updated in one block rather than a single-state updating scheme. For SV models, volatility is typically highly persistent. Thus the single-move sampler is of low efficiency given the highly correlated state variables as discussed

[15] The distributions of $p(x|\beta, \omega_{1,\ldots,T})$ and $p(\beta|x, \omega_{1,\ldots,T})$ are full conditional distributions, i.e., conditional distributions given the data and the other parameters.

in Carter and Kohn (1994) and Kim et al. (1998).[16]

The parameters d and σ_η^2 can be drawn from their full conditionals separately. By choosing the prior of σ_η^2 properly, say, inverse Gamma distribution, the full conditional distribution of σ_η is of closed form, the simulation from which can be easily implemented. The full conditional of d is not closed since the the moving average coefficients are non-linear transformations of d. Metropolis Hastings algorithm (Hastings (1970)) is useful for dealing with this difficulty. However, the two parameter estimates appear to be correlated based on our simulation results. Therefore, we propose updating the two parameters in one block with Random Walk Chain Metropolis-Hastings algorithm, in an attempt to improve convergence of the MCMC chain.

We choose a noninformative prior as

$$p(d, \sigma_\eta^2) \propto \frac{1}{\sigma_\eta^2} \times C_d$$

with C_d denoting the stationary region of $d \in (0, 0.5)$. To ensure positivity and stationarity, parameters d and σ_η^2 are transformed into unconstrained parameters $\tilde{\beta} = [\tilde{d}, \tilde{\sigma_\eta^2}]'$ with

$$\tilde{d} = \log \frac{2d}{1-2d} \qquad \tilde{\sigma_\eta^2} = \log(\sigma_\eta^2)$$

Since simulation from the density $p(\tilde{\beta}|x, \omega_{1,\ldots,T})$ is impractical, we simulate from a proposal density $q(\cdot)$ which permits simulation while at the same time being as close as possible to $p(\tilde{\beta}|x, \omega_{1,\ldots,T})$. This proposal density is generally referred to as *candidate generating density*. We correct for the fact that the candidate density is different from the posterior by not accepting all the candidate's draws. To illustrate, suppose, at the *mth* iteration, to move the chain to a new value of $\tilde{\beta}$,

a. Generate $\tilde{\beta}^\star$ from the candidate density $q(\tilde{\beta}^{(m-1)}, \cdot)$.

b. Accept $\tilde{\beta}^\star$ with the probability $\alpha = \min\left\{\frac{p(\tilde{\beta}^\star|x,\ \omega_{1,\ldots,T})}{p(\tilde{\beta}^{(m-1)}|x,\ \omega_{1,\ldots,T})}, 1\right\}$.

[16] Kim et al. (1998) applies this algorithm conditional on additional latent indicator variables. Jacquier et al. (1994), one of the earliest papers on Bayesian estimation of SV models, is based on a single-state updating schemed.

The second step is performed after independently generating u from Uniform$(0,1)$. If $u \leq \alpha$, $\tilde{\beta}^{\star}$ is accepted and the MCMC chain is updated as $\tilde{\beta}^{(m)} = \tilde{\beta}^{\star}$. If $u > \alpha$, $\tilde{\beta}^{\star}$ is rejected and $\tilde{\beta}^{(m)} = \tilde{\beta}^{(m-1)}$. Neglecting constant terms, the logarithm of $p(\tilde{\beta}|x,\ \omega_{1,\ldots,T})$ is

$$\log p(\tilde{\beta}|x,\ \omega_{1,\ldots,T}) = \log L\left(d, \sigma_{\eta}^{2}\right) + \tilde{d} - 2\log[1 + \exp(\tilde{d})]$$

where $\log L(\cdot)$ is evaluated via (4.16).

Under mild regularity conditions, the MCMC chain constructed this way will converge to the equilibrium distribution of $p(\tilde{\beta}|x, \omega_{1,\ldots,T})$. (See Gamerman and Lopes (2006, Ch. 6) for detailed discussions on Metropolis-Hastings and Random Walk Chain Metropolis-Hastings algorithms.) A good choice of the candidate density is helpful in speeding up MCMC convergence. We choose the following candidate density for its simplicity

$$\tilde{\beta}^{\star} = \tilde{\beta}^{(m-1)} + c\tau, \qquad \tau \sim N(0, V)$$

where c is a scale constant and the random increment τ follows a bivariate normal distribution. This option is implemented in many practical applications of Metropolis Hastings algorithms, see, for example, Müller (1991). We choose V as the negative inverse Hessian matrix obtained from maximizing the likelihood function of (4.16). In terms of initializations, the MCMC chain is initialized by the values at which V is obtained, and the initial state x_0 is drawn from the unconditional distribution of the state variable.

The value of c is fine-tuned so that a reasonable acceptance rate can be achieved. There are no universal rules to determine the optimal acceptance rate for the random walk chain. Roberts, Gelman and Gilks (1997) suggest that if the target and candidate densities are normal, the optimal acceptance rate be around 45% for one-dimension problems, around 25% in as many as 6 dimensions and approximately 23% as the number of dimensions approaches infinity. The optimal rate recommended by Müller (1991) is around 50%.[17] These figures serve as rough guides in our analysis.

This completes the details of the implementation of the Metropolis-within-Gibbs al-

[17] See the discussion in Chib and Greenberg (1995).

gorithm for our Bayesian estimation of the long memory SV model.

4.3.2 Diagnostics in MCMC Convergence

Convergence is an important aspect of MCMC estimation. According to the theory of Markov chains, our MCMC chain is expected to eventually converge to the equilibrium whose stationary distribution is our target distribution, say, the posterior distribution. A natural question to ask is: at what point do we know that the chain has converged to the stationary distribution? In other words, how do we choose the length of the *burn-in* period. The objective of MCMC convergence diagnostics is to see whether the estimated results are reliable. Cowles and Carlin (1996) provides a detailed review and comparisons of different diagnostics. Gilks, Richardson and Spiegelhalter (1996, Ch. 8) and Zellner and Min (1995) are also important references, with the latter focusing on convergence criteria for the Gibbs sampler. Koop (2003, section 4.2.4) also provides an accessible introduction.[18] We discuss some commonly used diagnostics below.

Visual inspection We can see how well the chain behaves through visual inspection of the *trace plot* for every parameter. A trace plot is the time series plot of the draw of the parameters at each iteration. We can see whether our chain gets stuck in certain areas of the parameter space because the MCMC chain is expected to explore the parameter space thoroughly so that a good picture of the posterior density is obtained.

Autocorrelation We can examine the autocorrelations between the draws of a MCMC chain. This is similar to examine the ACF plot of a time series to see whether the series appears to be stationary or non-stationary. The draws between consecutive iterations of the chains may be highly correlated; for example in the case of Metropolis-Hastings sampler, whether to accept a candidate draw is largely dependent on the draw obtained from the previous iteration. However, if autocorrelations are high even at large lags, this indicates a high degree of correlation between draws. With slow mixing of the chain, a

[18] Most MCMC diagnostics are helpful in detecting obvious nonstationarity but non-obvious nonstationarity. See the intuitive discussion on http://www.stat.umn.edu/∼charlie/mcmc/diag.html.

greater number of iterations are required to obtain reliable result on the posterior density.

Tests There are some formal tests we can perform to evaluate whether the chain appears to have converged. Popular tests include the Gelman and Rubin diagnostic (Gelman and Rubin (1992)) and the Geweke diagnostic (Geweke (1992)). The Gelman and Rubin diagnostic requires one to run more than two MCMC chains parallel, and then compare the within-chain and between-chain variation. The basic logic underlying the Geweke diagnostic test is that if the chain converges to the stationary distribution (the null hypothesis), the means of two non-overlapping parts taken from the chain should be equal. A significant difference between means indicates the two parts of the chain are not from the same distribution, accordingly, the equilibrium of the Markov chain has not been reached. The test is similar to the test of difference in means of two populations. However, the test has to take into account sample autocorrelation since the draws of the MCMC chains are not *iid*. The Geweke test estimates the standard error from the spectral density at zero frequency to adjust for any autocorrelation. Geweke's statistic has an asymptotically standard normal distribution, and the test statistic is a standard Z-score. We use this test in the Monte Carlo exercise and empirical application in this chapter due to its simplicity.[19]

4.4 Monte Carlo Experiment

This section reports on a Monte Carlo study to evaluate the performances of the estimators discussed so far: frequency domain QML, time domain QML and Bayesian method. Frequency QML serves as the benchmark when comparisons are made since it is one of the most commonly used parametric methods in the estimation of the long memory SV model. The posterior mean calculated via the Bayesian method is compared with QMLs.

Besides these three methods, we also consider the Geweke-Porter-Hudak (GPH) es-

[19]Some MCMC diagnostics, for example, the test proposed in Kim et al. (1998), lack asymptotic critical values of the test statistics. They are useful but less feasible to be implemented when a large number of Monte Carlo replications are conducted and the convergence of each replication needs to be checked. The Geweke test is implemented by the add-in package *coda* in *R* in this chapter.

timator (Geweke and Porter-Hudak (1983)), since this is one of leading semiparametric estimators of long memory in volatility and has been studied in detail for linear long-memory time series. The GPH estimator is robust to model misspecification of short-term dynamics, and also invariant to the mean of the time series to be estimated since it is based on the log periodogram regression at nonzero Fourier frequencies. But it cannot provide an estimator of σ_η. As discussed by a number of papers, see for example, Andersen and Bollerslev (1997*a*) and Deo and Hurvich (2001, 2003), for the long memory SV model specified in (4.4), the correlations of the absolute returns raised to any power r have hyperbolic decay and always decay at the same rate which is governed by d. In other words, any positive power of $|y_t|$ possesses long memory with the same value of d in (4.4). In this regard, the memory parameter estimated from these power transformations can provide a consistent estimator of d. According to the linearization in (4.5), $\log(y_t^2)$ has the same memory parameter as h_t. Therefore, some studies also use $\log(y_t^2)$ as the volatility approximation to obtain the GPH estimator of d.

One practical issue with the GPH estimator is the choice of bandwidth m and there is no universally accepted rule for determining the optimal value of m in the literature. In our simulation studies, we try the values of (0.3, 0.4, ..., 0.8). The choice of $m = 0.5$ appears to provide a reasonable tradeoff between bias and variance. In terms of the choice of volatility approximation, we experiment with $\log(y_t^2)$ and $|y_t|^r$ with a range of the values of r as (0.5, 0.75, 1, 1.25, 2). Our simulation results indicate the performances of the various transformations are largely similar, although the case with $r = 1$ appears to be better in general. In some cases, for a given value of m, the squared returns show the least persistence. This is consistent with other observations in the literature (see Deo and Hurvich (2003)). Further discussion around the use of the GPH estimator in a volatility context can be found in Bollerslev and Wright (2000) and Taylor (2005, Ch. 12). In the following discussions of the simulation results, we focus on the GPH estimator on $|y_t|$ with $m = 0.3$ and 0.5. The results for other scenarios are available on request.

4.4.1 Simulation Design

The assumptions about parameters that underly classical and Bayesian methods differ greatly: classical methods treat a parameter, say, β as a fixed but unknown quantity, while Bayesians assume β is a random quantity. In this regard, posterior means are not expected to have identical repeated frequency behavior to those of QML estimators.[20] As noted in Shephard (1996), the comparison of estimators based on simulations implicitly treats the Bayesian method as an empirical Bayes procedure by treating the mean of the posterior distribution as an estimator of β. The intention of our simulation study is to compare the performance of the proposed Bayesian point estimator to that of the classical counterparts. The following parameter values are considered:

$$\begin{aligned} \mathrm{d} = 0.3: &\quad \sigma_\eta = 0.2, \quad \sigma_\eta = 0.5, \quad \sigma_\eta = 1 \\ \mathrm{d} = 0.4: &\quad \sigma_\eta = 0.2, \quad \sigma_\eta = 0.5, \quad \sigma_\eta = 1 \end{aligned}$$

with the scale parameter $\sigma = 1$ in all cases. The values of the parameters are chosen primarily based on empirical evidence, where typical estimates of d for volatility are around 0.4. It is desirable to choose the values of σ_η so they are empirically realistic, as our simulation study indicates that its magnitude has an effect on the performance of each estimator. The three values considered are according to the following reasoning: in the simulation study of Breidt et al. (1998), only the case of $\sigma_\eta = 1$ is considered, while this value is somewhat larger than the commonly observed values in practice, as noted in Pérez and Ruiz (2001). The value of $\sigma_\eta = 0.5$ is chosen because this is the benchmark value used in Bollerslev and Wright (2000) when the long memory SV model is considered. Finally, it is not uncommon for estimates of σ_η to be around 0.2 when the short memory SV model is estimated in practice.

Sample sizes of 2000, 4000 and 8000 are considered. The simulation of the fractional Gaussian process and the GPH estimates are obtained via the *fracdiff* package in *R*. For each estimation method, a total of 1000 replications is conducted given a set of parameter values and sample size. The order of moving average approximation to fractional noise is

[20] Posterior means with proper diffuse priors tend to behave similar to MLEs when sample size is large.

chosen as 20 for the state-space representation of long memory SV model.[21] The burn-in period and iterations for MCMC chains are 1000 and 4000 respectively.[22]

4.4.2 Simulation Results

Simulation results are summarized in Tables 4.1 and 4.2, where the bias and root mean squared error (RMSE) of each method are reported, with $Bias(\hat{\beta}) = \beta - E[\hat{\beta}]$ and $MSE(\hat{\beta}) = E[(\beta - \hat{\beta})^2]$. The estimated densities of the point estimators and posterior means are plotted in Figures 4.1-4.6. Note that the estimates of σ_η are not available for the GPH method.

Let us first look at the convergence in estimation before we discuss the tables and figures. MCMC convergence is considered by the Geweke test and visual inspection on a Markov chain. Whenever a Geweke diagnostic test is conducted, we follow the convention of testing for equality of the means of the first 10% and the last 50% of a Markov chain. We conduct diagnostics for d and σ_η for all 1000 replications. The results from visual inspection and the Geweke test are largely consistent. The magnitudes of d and σ_η play a role in convergence for the parameters considered. With $\sigma_\eta = 0.5$ and 1, there is little difficulty in achieving convergence regardless of the value of d and sample size. The corresponding results of MCMC and other methods, which are tabulated in the middle and bottom panels in Tables 4.1 and 4.2, are based on 1000 replications.

[21] We also try the value of 30 with a small scale simulation study, there is no great improvement observed. However, the choice of 30 significantly increases computing time.

[22] The lengths of burn-in period and MCMC iterations are largely based on computation cost. With total replications of 1000×3×6 for each method, the computing time, especially for MCMC method, is not a trivial factor. For most replications, the MCMC chain appears well-behaved after the chosen burn-in period, indicating the choice is reasonable. Time domain QML and Bayesian estimations are conducted with *Ox* and make use of the state space modeling package *SsfPack*. The *Ox* maximization package is used for time domain QML. Frequency domain QML is conducted in *R*. The BFGS algorithm is chosen for numerical optimization. A summary of computing time (in hour) for each method with 1000 replications is provided here for referencee:

	T=2000	T=4000	T=8000
QML (frequency domain)	0.18hr	0.37	1.75
QML (time domain)	8	10	13.5
Bayesian (MCMC)	500	667	1050

These numbers are provided as rough guides, and there is certainly room for improvement especially for Bayesian methods, for example, choosing software more computationally efficient such as C^{++}, or using dynamic link library, which we will not pursue here. All the simulations and computations are run on the *Monash Sun Grid (MSG)*, which facilitates parallel computing.

Poor convergence is observed when $\sigma_\eta = 0.2$ regardless of the magnitude of d, although the convergence is better with higher value of d. To illustrate, in the scenario of $\sigma_\eta = 0.2$ and $d = 0.3$, the number of converged MCMC chains of 1000 replications are 95, 123 and 157 when $T = 2000$, 4000, 8000 respectively; when $\sigma_\eta = 0.2$ and $d = 0.4$, the number of converged MCMC chains of 1000 replications are 153, 232 and 362 when $T = 2000$, 4000, 8000 respectively. The results reported for MCMC method (the top panels in Tables 4.1 and 4.2) are based on these converged replications, and the results of other methods are based on 1000 replications.[23] For both the frequency and time domain QMLEs, numerical optimizations can be obtained for each of the 1000 replications. However, difficulty in estimation is implied by the observation that numerical optimizations frequently converge to the parameter boundaries, as shown in the density plots in Figures 4.1 and 4.4. For example, the estimated density of d for each sample size with $\sigma_\eta = 0.2$ has a global mode which is on the lower or upper boundary, i.e. 0 or 0.5.

The difficulty in estimation when $\sigma_\eta = 0.2$ is also seen in the QMLEs being sensitive to the initial values of optimization. The results reported in Tables 4.1 and 4.2 are obtained with initial values of $d_0 = 0.3$ when the true value of d is 0.4, and $d_0 = 0.4$ with the true value being 0.3, and $\sigma_\eta = 0.6$ in all the cases. A Monte Carlo exercise is also conducted in exactly the same way but with the true parameters used as initial values.[24] No obvious difference is observed when $\sigma_\eta = 0.5$ and 1. For $\sigma_\eta = 0.2$, both bias and RMSE for QMLEs decrease when optimizations are initialized with the true values, while the magnitude of the decrease becomes negligible as sample size increases. Compared with QMLEs, Bayesian methods are not sensitive to initial values in terms of bias and RMSE of posterior means, although the proportions of converged chains increases with good initial values.

In general, the performance of each method improves as the magnitude of σ_η increases given the same memory parameter. Both the bias and RMSE decrease as σ_η increases in Tables 4.1 and 4.2. Given the same value of σ_η except when $\sigma_\eta = 0.2$, all the estimators except the GPH estimator improve as the value of d increases. These observations are

[23] The results of all the methods based on the replications when MCMC chain converges are tabulated in Table 4.3. We will discuss these results later.

[24] The tables and figures obtained when true parameter values are used for initializations are not reported here in order to save space, and they are available on request.

Table 4.1: Bias and RMSE for d and σ_η with true value of $d = 0.3$

Parameters			d		σ_η	
d=0.3			Bias	RMSE	Bias	RMSE
$\sigma_\eta = 0.2$ [a]	T=2000	QML (frequency domain)	0.1966	0.2454	-0.0579	0.2636
		QML (time domain)	0.0523	0.1711	-0.0455	0.2563
		Bayesian (MCMC)	0.0900	0.1113	-0.2933	0.3559
		GPH (m=0.5)	0.2292	0.2412	—	—
	T=4000	QML (frequency domain)	0.1915	0.2415	-0.0278	0.2189
		QML (time domain)	0.0351	0.1728	-0.0208	0.2117
		Bayesian (MCMC)	0.0781	0.1075	-0.2016	0.2491
		GPH (m=0.5)	0.2298	0.2383	—	—
	T=8000	QML (frequency domain)	0.1675	0.2259	-0.0119	0.1838
		QML (time domain)	0.0195	0.1699	-0.0158	0.1756
		Bayesian (MCMC)	0.0752	0.1109	-0.1591	0.1961
		GPH (m=0.5)	0.2273	0.2343	—	—
$\sigma_\eta = 0.5$	T=2000	QML (frequency domain)	0.0224	0.1438	0.0058	0.2271
		QML (time domain)	-0.0048	0.1568	0.0028	0.2180
		Bayesian (MCMC)	0.0290	0.0817	-0.0554	0.1700
		GPH (m=0.5)	0.1501	0.1798	—	—
	T=4000	QML (frequency domain)	0.0014	0.1000	0.0020	0.1688
		QML (time domain)	-0.0272	0.1272	0.0061	0.1696
		Bayesian (MCMC)	0.0010	0.0831	-0.0075	0.1377
		GPH (m=0.5)	0.1410	0.1646	—	—
	T=8000	QML(frequency domain)	-0.0028	0.0689	0.0039	0.1266
		QML(time domain)	-0.0398	0.0991	0.0119	0.1342
		Bayesian (MCMC)	-0.0266	0.0770	0.0160	0.1152
		GPH (m=0.5)	0.1271	0.1460	—	—
$\sigma_\eta = 1.0$	T=2000	QML (frequency domain)	-0.0015	0.0578	0.0084	0.1495
		QML (time domain)	-0.0121	0.0681	0.0131	0.1565
		Bayesian (MCMC)	-0.0155	0.0647	0.0302	0.1530
		GPH (m=0.5)	0.0861	0.1394	—	—
	T=4000	QML (frequency domain)	-0.0012	0.0400	0.0066	0.1066
		QML (time domain)	-0.0185	0.0490	0.0116	0.1126
		Bayesian (MCMC)	-0.0221	0.0492	0.0224	0.1133
		GPH (m=0.5)	0.0761	0.1181	—	—
	T=8000	QML (frequency domain)	-0.0026	0.0269	0.0087	0.0737
		QML (time domain)	-0.0248	0.0343	0.0135	0.0790
		Bayesian (MCMC)	-0.0268	0.0350	0.0192	0.0801
		GPH (m=0.5)	0.0635	0.0996	—	—

[a]When $\sigma_\eta = 0.2$, the number of converged MCMC chains of 1000 replications are 95, 123 and 157 when T=2000, 4000, 8000 respectively. The results reported for MCMC method are based on these converged replications. The results of other methods are based on 1000 replications.

[b]The initial values for optimizations are d_0=0.4 and $\sigma_{\eta 0}$=0.6 for frequency and time domain methods. Bayesian posterior mean is used as a point estimator. The order of moving average approximation to fractional noise is chosen as 20.

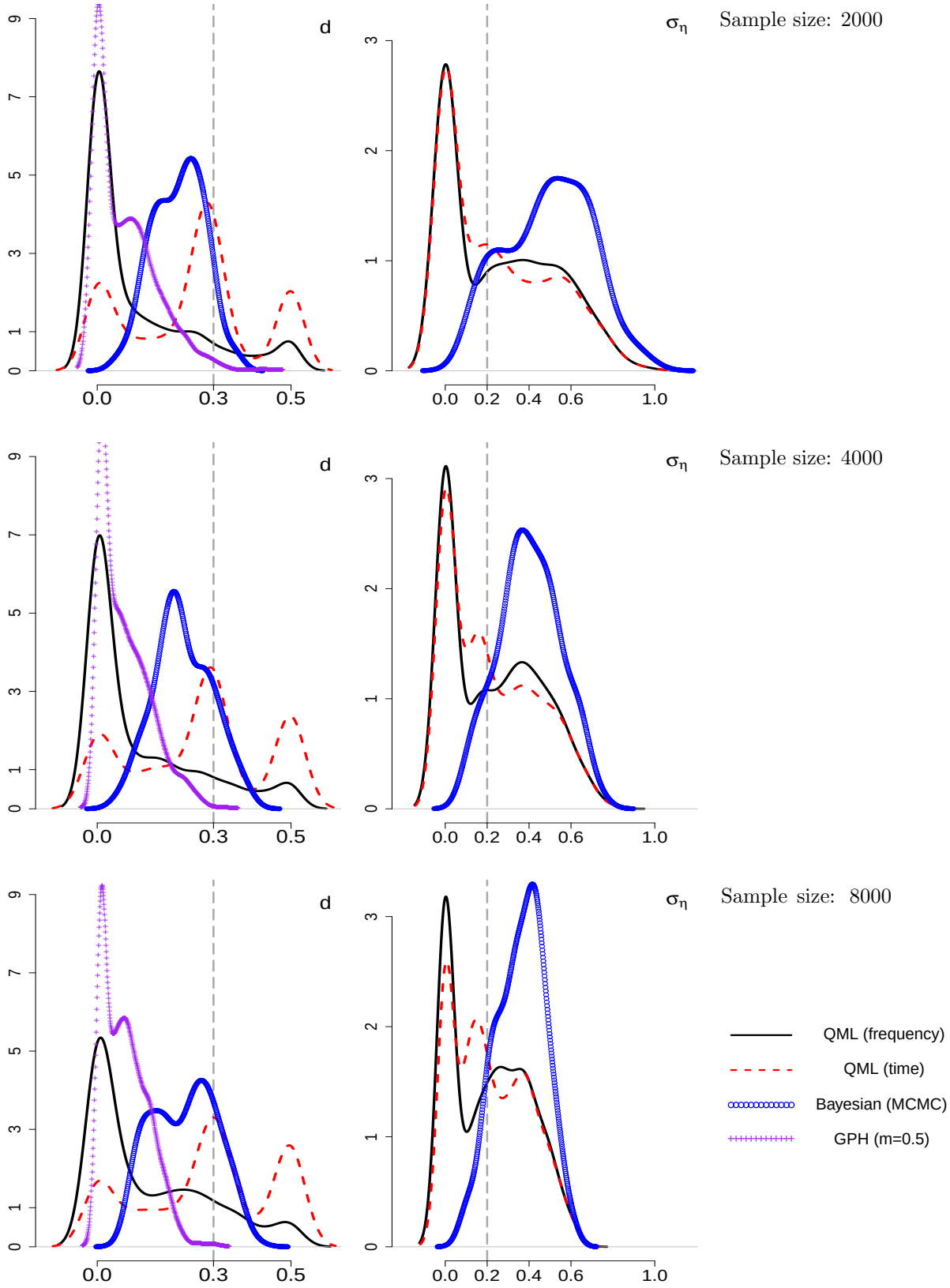
Table 4.2: Bias and RMSE for d and σ_η with true value of $d = 0.4$

Parameters			d		σ_η	
d=0.4			Bias	RMSE	Bias	RMSE
$\sigma_\eta = 0.2$ [a]	T=2000	QML (frequency domain)	0.2305	0.2932	-0.0806	0.2659
		QML (time domain)	0.1741	0.2454	-0.0916	0.2667
		Bayesian (MCMC)	0.1695	0.1849	-0.2857	0.3456
		GPH (m=0.5)	0.2830	0.2976	—	—
	T=4000	QML (frequency domain)	0.1832	0.2596	-0.0543	0.2142
		QML (time domain)	0.1227	0.2156	-0.0713	0.2124
		Bayesian (MCMC)	0.1462	0.1659	-0.1970	0.2499
		GPH (m=0.5)	0.2650	0.2780	—	—
	T=8000	QML (frequency domain)	0.1164	0.2085	-0.0225	0.1649
		QML (time domain)	0.0599	0.1712	-0.0538	0.1620
		Bayesian (MCMC)	0.1104	0.1383	-0.1303	0.1759
		GPH (m=0.5)	0.2481	0.2584	—	—
$\sigma_\eta = 0.5$	T=2000	QML (frequency domain)	0.0082	0.0927	-0.0104	0.1781
		QML (time domain)	-0.0101	0.1017	-0.0234	0.1725
		Bayesian (MCMC)	0.0338	0.0672	-0.0522	0.1433
		GPH (m=0.5)	0.1541	0.1909	—	—
	T=4000	QML (frequency domain)	-0.0014	0.0719	0.0002	0.1363
		QML (time domain)	-0.0329	0.0775	-0.0133	0.1252
		Bayesian (MCMC)	-0.0041	0.0548	-0.0334	0.1027
		GPH (m=0.5)	0.1280	0.1573	—	—
	T=8000	QML (frequency domain)	-0.0030	0.0531	0.0010	0.1023
		QML (time domain)	-0.0509	0.0589	-0.0055	0.0933
		Bayesian (MCMC)	-0.0308	0.0423	-0.0246	0.0758
		GPH (m=0.5)	0.1053	0.1297	—	—
$\sigma_\eta = 1.0$	T=2000	QML (frequency domain)	-0.0024	0.0515	0.0001	0.1383
		QML (time domain)	-0.0205	0.0594	-0.0001	0.1387
		Bayesian (MCMC)	-0.0136	0.0455	-0.0077	0.1216
		GPH (m=0.5)	0.0879	0.1436	—	—
	T=4000	QML (frequency domain)	-0.0028	0.0352	0.0021	0.0979
		QML (time domain)	-0.0341	0.0540	0.0074	0.1018
		Bayesian (MCMC)	-0.0305	0.0459	0.0029	0.0913
		GPH (m=0.5)	0.0730	0.1213	—	—
	T=8000	QML (frequency domain)	-0.0020	0.0255	-0.0011	0.0732
		QML (time domain)	-0.0438	0.0547	0.0074	0.0773
		Bayesian (MCMC)	-0.0399	0.0488	0.0050	0.0707
		GPH (m=0.5)	0.0560	0.0967	—	—

[a]When $\sigma_\eta = 0.2$, the number of converged MCMC chains of 1000 replications are 153, 232 and 362 when T=2000, 4000, 8000 respectively. The results reported for MCMC method are based on these converged replications. The results of other methods are based on 1000 replications.

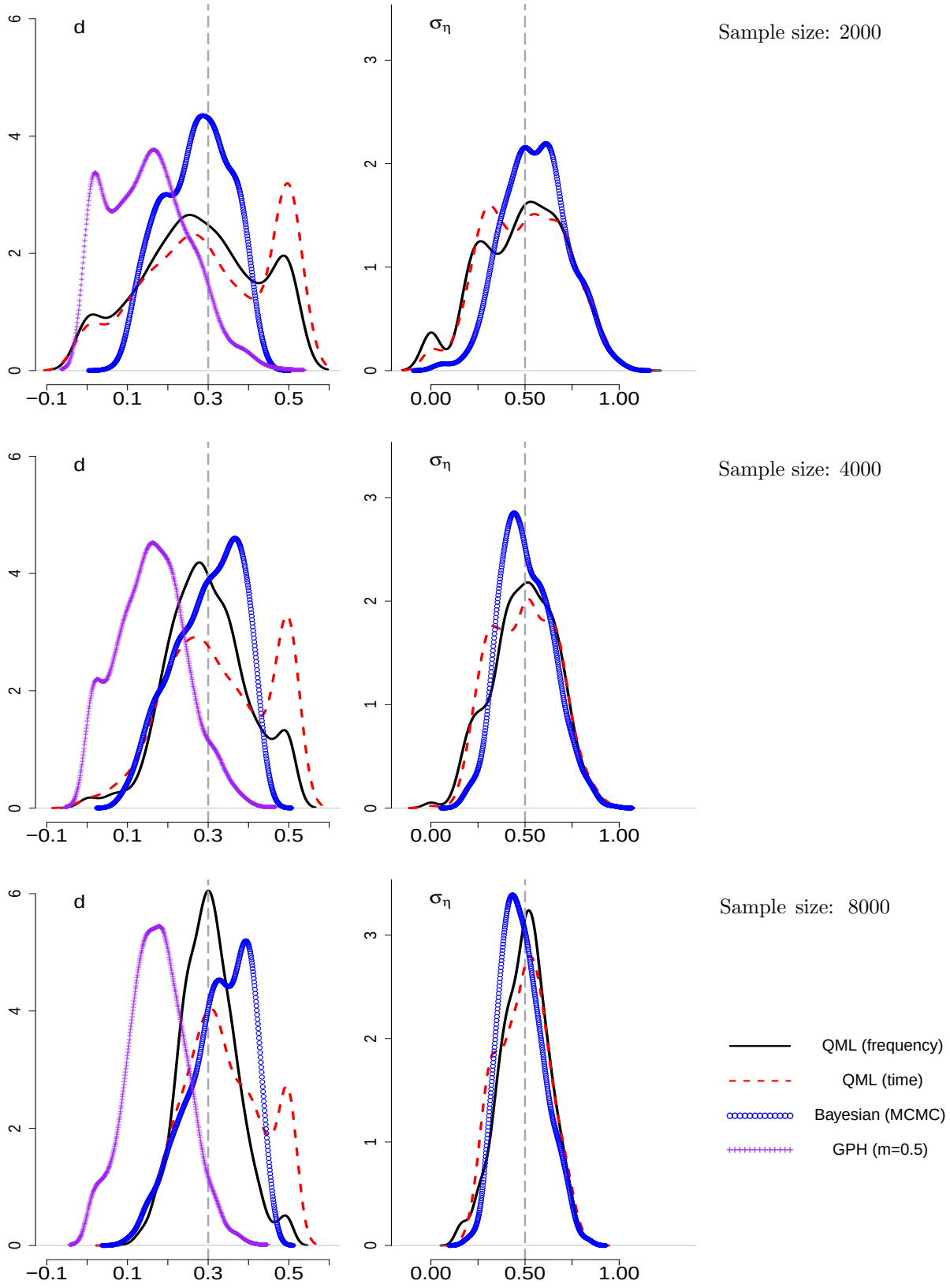
[b]The initial values for optimizations are d_0=0.3 and $\sigma_{\eta 0}$=0.6 for both frequency and time domain methods. Bayesian posterior mean is used as a point estimator. The order of moving average approximation to fractional noise is chosen as 20.

Figure 4.1: Estimated densities of d and σ_η with true values of d=0.3, σ_η=0.2



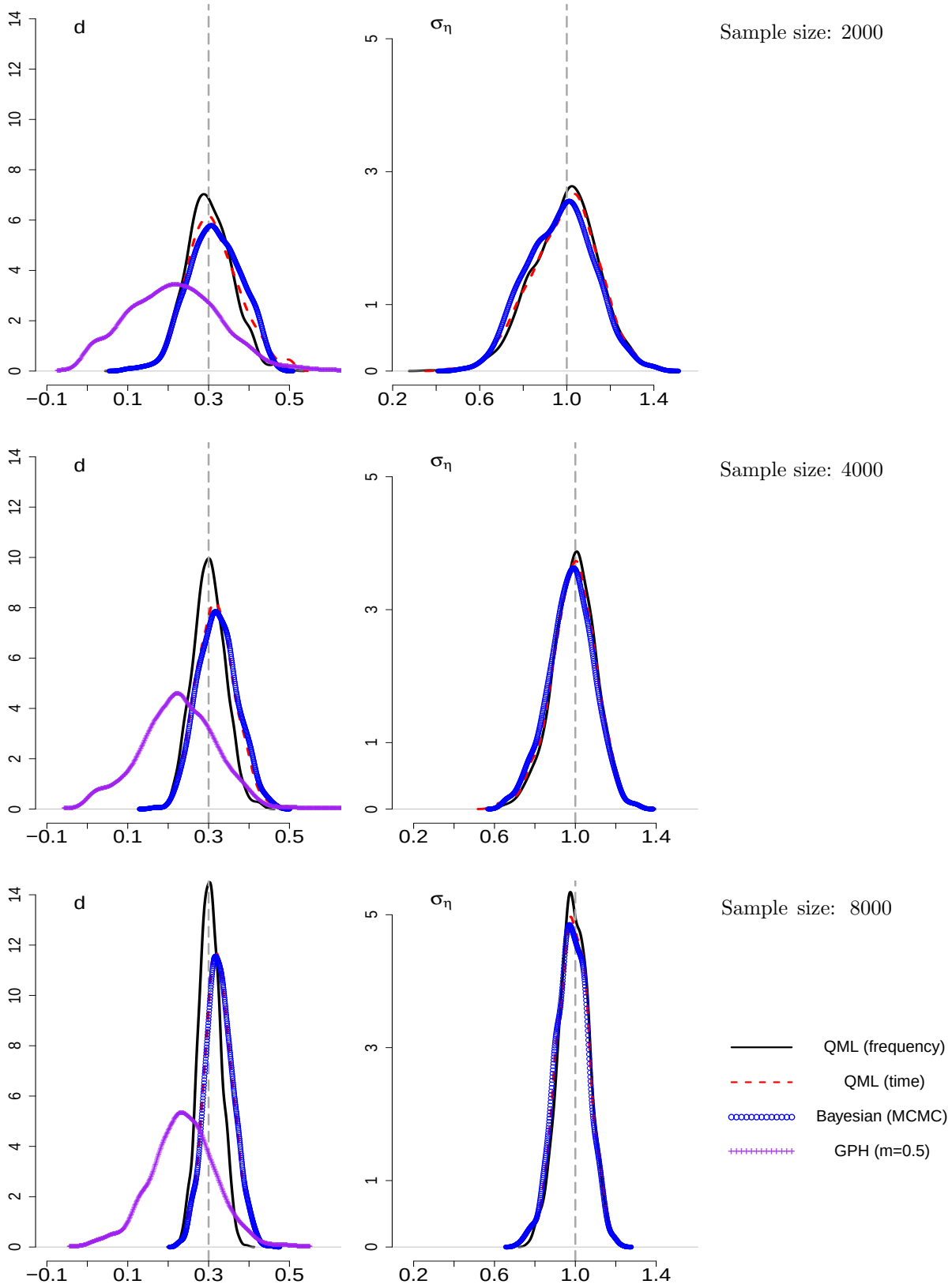
The density of posterior mean of MCMC method is based only on the replications where convergence was achieved: this occurred for 95, 123 and 157 samples when T = 2000, 4000, 8000 respectively. The results of other methods are based on 1000 replications. Vertical line drawn on each plot indicates the true parameter value. Bayesian posterior mean is used as parameter point estimator.

Figure 4.2: Estimated densities of d and σ_η with true values of d=0.3, σ_η=0.5



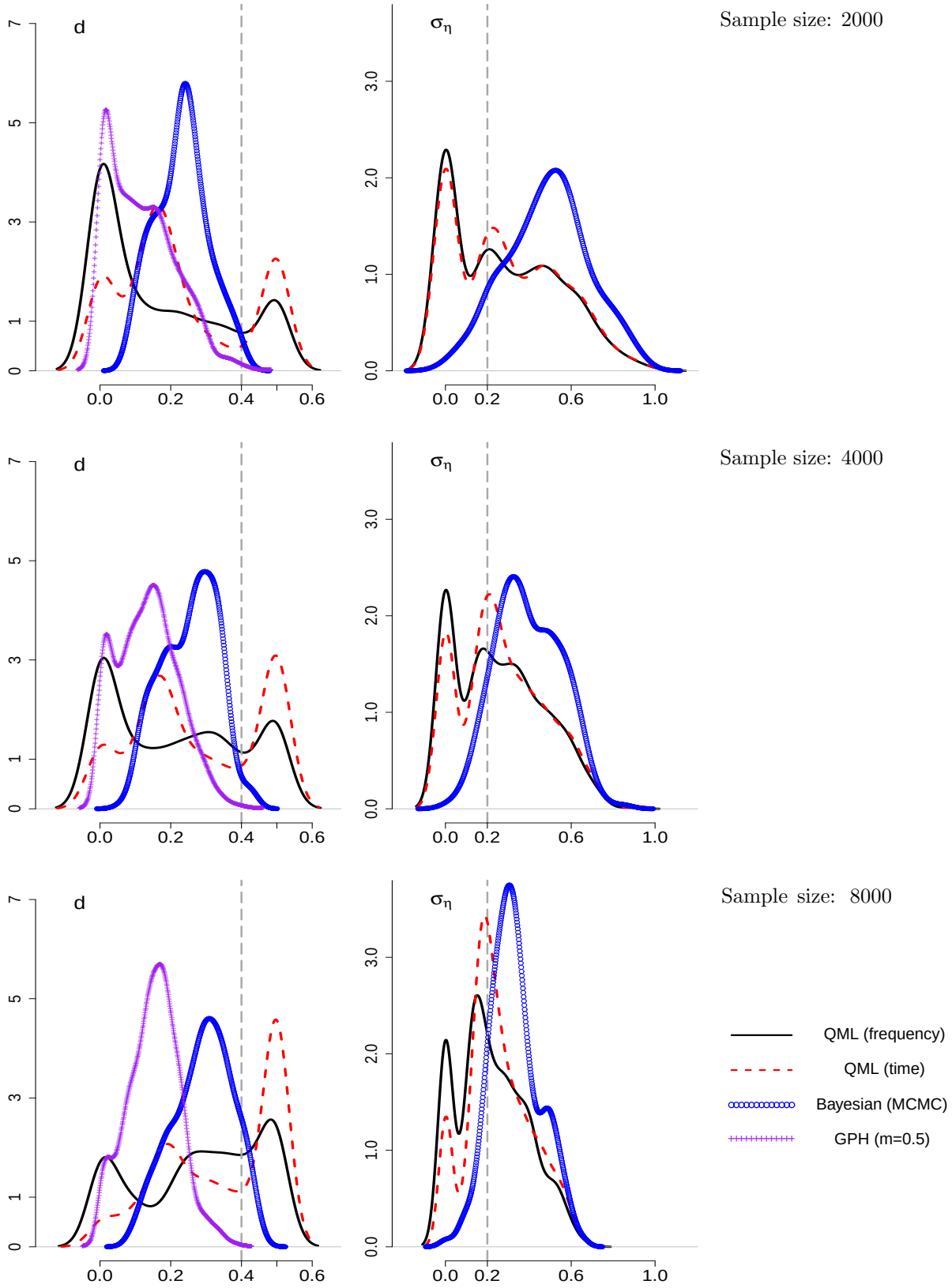
The results are based on 1000 Monte Carlo replications. Vertical line drawn on each graph indicates the true parameter value. Posterior mean is used as point parameter estimator for Bayesian method.

Figure 4.3: Estimated densities of d and σ_η with true values of d=0.3, σ_η=1



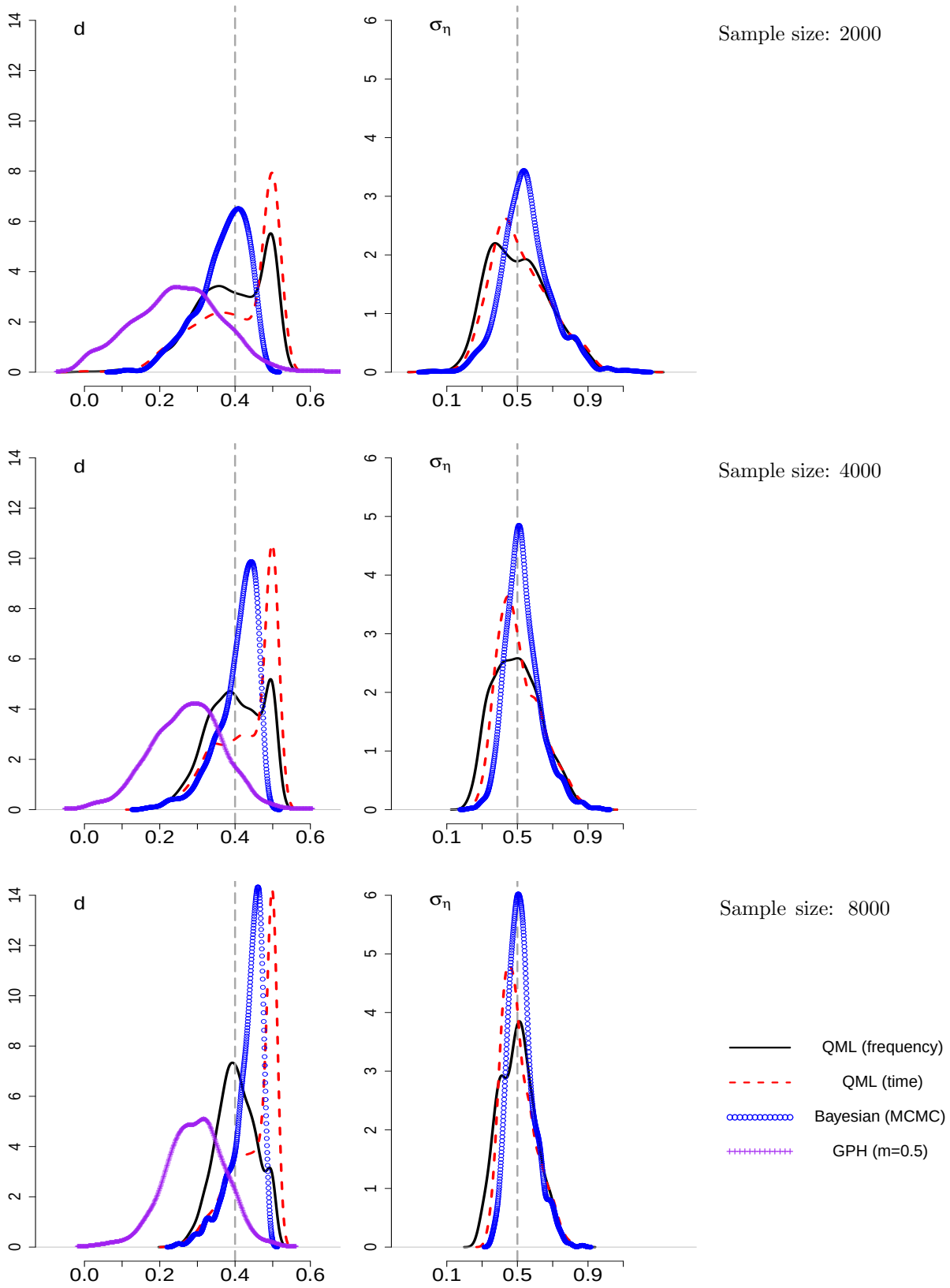
The results are based on 1000 Monte Carlo replications. Vertical line drawn on each graph indicates the true parameter value. Posterior mean is used as point parameter estimator for Bayesian method.

Figure 4.4: Estimated densities of d and σ_η with true values of d=0.4, σ_η=0.2



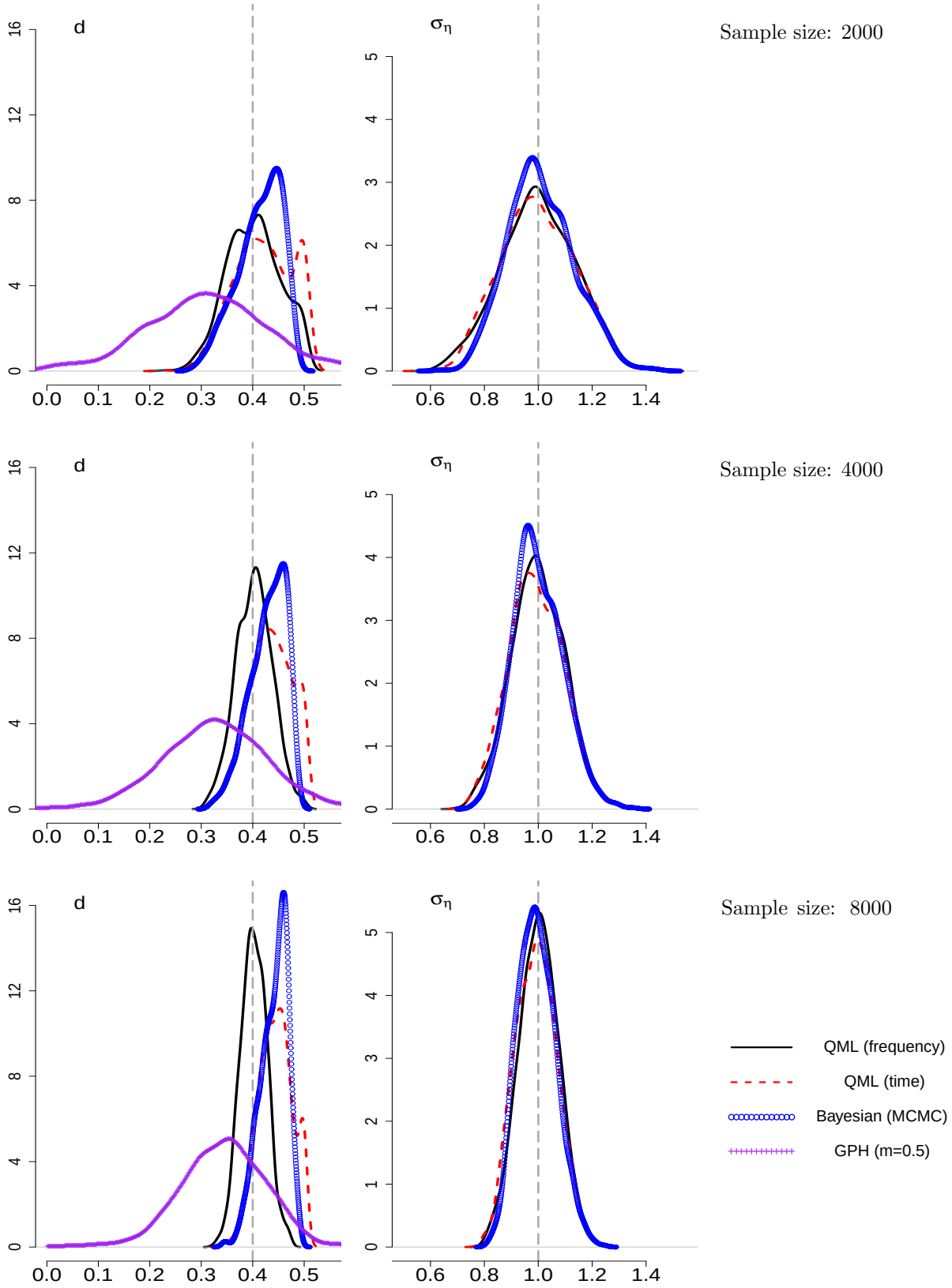
The density of posterior mean of MCMC method is based only on the replications where convergence was achieved: this occurred for 153, 232 and 362 samples when $T = 2000, 4000, 8000$ respectively. The results for other methods are based on 1000 replications. Vertical line drawn on each plot indicates the true parameter value. Bayesian posterior mean is used as parameter point estimator.

Figure 4.5: Estimated densities of d and σ_η with true values of d=0.4, σ_η=0.5



The results are based on 1000 Monte Carlo replications. Vertical line drawn on each graph indicates the true parameter value. Posterior mean is used as point parameter estimator for Bayesian method.

Figure 4.6: Estimated densities of d and σ_η with true values of d=0.4, σ_η=1



The results are based on 1000 Monte Carlo replications. Vertical line drawn on each graph indicates the true parameter value. Posterior mean is used as point parameter estimator for Bayesian method.

largely consistent with Pérez and Ruiz (2001) and Hsu and Breidt (2003). The GPH estimator is the worst in most of the scenarios, implied by relatively large bias and RMSE. The memory parameter of GPH is always underestimated, and this phenomenon is consistent with the previous studies; see for example, the simulation results in Breidt et al. (1998) and the asymptotic results predicted by the theorem in Deo and Hurvich (2003).

We provide an informal explanation of the effect of the magnitude of σ_η, volatility of volatility, on estimator performance. The parameterization of the long memory SV model in (4.5) indicates the assumption that volatility cannot be observed directly but through returns. The observed series $\log(y_t^2)$ is treated as a fractionally integrated process (i.e. the log volatility) with additive noise. The variance of the additive noise σ_ξ^2 is equal to $\pi^2/2$ according to (4.2). In the terminology of state-space models, the ratio of $\sigma_\eta^2/\sigma_\xi^2$ is called the *signal-to-noise ratio*. It increases as the value of σ_η increases as σ_ξ^2 is fixed. The extreme scenario of signal-to-noise ratio approaching infinity can be conceptually interpreted as a fractional integrated process with no additive noise. In such a case, the dynamics of the latent process are expected to estimated with more accuracy. On the other hand, a small signal-to-noise ratio indicates a relatively large magnitude for the additive noise. As such, it could be difficult to uncover the dynamics of the latent process with relatively large additive noise as they could mask the strength of autocorrelations. It is clear in the tables that all the estimators of d are downward biased when $\sigma_\eta = 0.2$. The simulation result in Hsu and Breidt (2003, pp. 503-4) indicates that the performance of the posterior mean of d improves as the signal-to-noise ratio increases when dealing with a fractionally integrated ARMA process with additive noise. Their result is consistent with the pattern observed here.

Comparisons between the classical and Bayesian methods are discussed here.

- With small σ_η, no method performs consistently well. In terms of the memory parameter, the posterior means of the Bayesian method appear to provide smaller RMSE, while the QMLE estimators often reach the lower and upper boundaries of 0 and 0.5 with non-negligible proportions even with the large sample size of 8000, as indicated in the density plots. For example, a global mode around 0 is observed

on the estimated density plots of d for the frequency method when $d = 0.3$ (Figure 4.1), and the corresponding density plots of d appear to be fairly flat between the two boundaries when $d = 0.4$ (Figure 4.4).

For the replications when MCMC chain converges, Table 4.3 is reported in order to compare between the classical and Bayesian methods. The Bayesian method outperforms other methods in terms of bias and RMSE. For both parameters, the RMSE of the Bayesian method is always the smallest, with the bias being the smallest in most of the cases. This tells a similar story in the previous Tables 4.1 and 4.2.

Table 4.3: Bias and RMSE for d and σ_η with true value of $\sigma_\eta = 0.2$

Parameters			d		σ_η	
			Bias	RMSE	Bias	RMSE
$d = 0.3$ [a]	T=2000	QML (frequency domain)	0.1211	0.1676	-0.4055	0.4454
		QML (time domain)	0.1073	0.1819	-0.4024	0.4430
		Bayesian (MCMC)	0.0900	0.1113	-0.2933	0.3559
		GPH (m=0.5)	0.2197	0.2361	—	—
	T=4000	QML (frequency domain)	0.1215	0.1712	-0.3038	0.3326
		QML (time domain)	0.0914	0.1774	-0.2956	0.3269
		Bayesian (MCMC)	0.0781	0.1075	-0.2016	0.2491
		GPH (m=0.5)	0.2389	0.2470	—	—
	T=8000	QML (frequency domain)	0.1140	0.1583	-0.2325	0.2567
		QML (time domain)	0.0836	0.1712	-0.2249	0.2521
		Bayesian (MCMC)	0.0752	0.1109	-0.1591	0.1961
		GPH (m=0.5)	0.2255	0.2328	—	—
$d = 0.4$ [b]	T=2000	QML (frequency domain)	0.1758	0.2213	-0.3730	0.4219
		QML (time domain)	0.1646	0.2248	-0.3802	0.4245
		Bayesian (MCMC)	0.1695	0.1849	-0.2857	0.3456
		GPH (m=0.5)	0.2688	0.2865	—	—
	T=4000	QML (frequency domain)	0.1345	0.1796	-0.2535	0.3012
		QML (time domain)	0.1187	0.1934	-0.2614	0.3063
		Bayesian (MCMC)	0.1462	0.1659	-0.1970	0.2499
		GPH (m=0.5)	0.2530	0.2671	—	—
	T=8000	QML (frequency domain)	0.0818	0.1403	-0.1441	0.1982
		QML (time domain)	0.0731	0.1645	-0.1660	0.2102
		Bayesian (MCMC)	0.1104	0.1383	-0.1303	0.1759
		GPH (m=0.5)	0.2380	0.2491	—	—

[a] When $\sigma_\eta = 0.2$ and $d = 0.3$, the number of converged MCMC chains of 1000 replications are 95, 123 and 157 when T=2000, 4000, 8000 respectively. The results reported for MCMC and other methods are based on these converged replications.

[b] When $\sigma_\eta = 0.2$ and $d = 0.4$, the number of converged MCMC chains of 1000 replications are 153, 232 and 362 when T=2000, 4000, 8000 respectively. The results reported for MCMC and other methods are based on these converged replications.

- With $\sigma_\eta = 0.5$, the Bayesian method outperforms QMLEs with smaller RMSE with sample sizes of 2000 and 4000. In the top and middle panels for the d parameter in Figures 4.2 and 4.5, local or global modes are evident around the parameter boundaries for both QMLEs. In other words, the numerical optimizations of both methods still have a good chance of converging to the boundaries. As sample size increases to 8000, the performance of the frequency QMLE improves, providing the smallest bias and comparable RMSE to that of the Bayesian method. This is as expected, as the simulation studies in Cheung and Diebold (1993) suggest that estimation based on the time domain likelihood approximation is more efficient in small samples compared to that of the frequency counterpart.

- With $\sigma_\eta = 1$, in terms of the estimators of d, the frequency QMLE provides the smallest bias and RMSE of all the sample sizes considered when $d = 0.3$. When $d = 0.4$, the RMSE of the Bayesian method is the smallest when sample size is 2000. In terms of the estimators of σ_η, the estimated densities of the QMLEs and Bayesian methods appear fairly similar regardless of sample size (Figures 4.3 and 4.6).

To sum up, for the values of d considered, when the magnitude of volatility of volatility is small, all the estimators examined perform poorly. In the other scenarios, the Bayesian method is capable of offering an estimator with smaller RMSE when sample size is 2000 and 4000, indicating a better efficiency over QMLEs. This superior performance decreases as sample size increases to 8000. The GPH estimator is inferior to the other three methods in virtually all the cases. The overall performance of frequency method is better when $\sigma_\eta = 1$, however this magnitude is larger than the commonly observed volatility of volatility in practice. Therefore the simulation results when σ_η is around 0.5 might be of more practical relevance.

Before closing this section, some comments are in order. The prior used is not very informative. As argued in Breidt and Carriquiry (1996),[25] the prior can influence the

[25] We also experiment with the robust transformation proposed by Breidt and Carriquiry (1996) to long memory SV model to alleviate the effect of *inliers* caused by log transformation of daily returns of small magnitude. Our Monte Carlo simulation indicates the transformation is helpful in reducing bias when $\sigma_\eta = 0.2$ but at the cost of increased RMSE. The results are not reported here and are available upon request.

performance of the Bayesian estimator. As such, a more informative prior could be useful. It is not unrealistic to expect that practitioners have some knowledge of the parameters to be estimated. A different choice of prior will not complicate the MCMC simulation. So there is room for the Bayesian method to improve. The form of the proposal density used in our MCMC simulation is also suitable for another popular posterior simulation method, importance sampling. If the focus of a Bayesian exercise is a posterior moment of the parameter, say the posterior mean of d, then importance sampling is suitable. It is less involved in terms of computational time and the implementation of the algorithm into computer code is straightforward. A benefit of importance sampling compared with MCMC simulation is that it does not require convergence checks.[26] We suggest when MCMC simulation has difficulty in converging, one can consider the importance sampling method. We examine the implementation of importance sampling when $\sigma_\eta = 0.2$, the result indicates it provides the smallest RMSE given the three sample sizes considered. Table 4.4 reports the importance sampling result when $\sigma_\eta = 0.2$. Compared with the

Table 4.4: Bias and RMSE for d and σ_η of the importance sampling method

	$d = 0.3, \sigma_\eta = 0.2$				$d = 0.4, \sigma_\eta = 0.2$			
	d		σ_η		d		σ_η	
Sample size	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
2000	0.0787	0.1092	-0.0975	0.2296	0.1637	0.1807	-0.1256	0.2300
4000	0.0752	0.1035	-0.0570	0.1782	0.1484	0.1667	-0.0858	0.1808
8000	0.0627	0.0963	-0.0126	0.1434	0.1164	0.1403	-0.0481	0.1401

corresponding figures in Tables 4.1 and 4.2, the RMSE of the importance sampling method is smaller. Having said that, one weakness of importance sampling is that it cannot provide a full picture of the parameter posterior density.

4.4.3 Sensitivity Analysis of Monte Carlo Results

As observed in our simulation study, the performance of each method examined varies with the values of the underlying parameters. The objective of this section is to conduct a sensitivity analysis to the Monte Carlo results by considering a larger set of values of

[26] See Ripley (1987) on the algorithm of importance sampling, Geweke (1989) on theoretical justifications and applications to econometric models, and Durbin and Koopman (2001*b*, Ch. 8 and 11) for the implementations of importance sampling for state space models.

the key parameters. This would be useful for empirical applications of the long memory SV models. The values of the parameters being examined are as follows,

Grid for d: 0.22 to 0.46 in steps of 0.04 with 7 grids

Grid for σ_η: 0.25 to 1.15 in steps of 0.10 with 10 grids

Given each pair of d and σ_η, 500 replications are conducted. The frequency QMLE is examined mainly because it is infeasible to conduct such an exercise by other methods as their computing time is overwhelming. Also, the overall pattern observed here is expected to be representative of all estimators. Bias and RMSE for the parameters of d and σ_η are reported in Tables 4.5 to 4.12, for the sample sizes of 2000, 4000, 8000 and 16000.

Let us first focus on the estimation of d. For a given magnitude of σ_η, the stronger the persistence in volatility (i.e. a more positive value of the true d), the better the estimation performance, indicated by smaller RMSE. The evidence becomes clearer as sample size increases. The exception is when $\sigma_\eta = 0.25$, where the results are mixed when the true value of d is smaller than 0.38, especially with sample sizes of 2000 and 4000. One possible explanation of this exception is that the numerical optimizations frequently converge to the parameter boundaries when $\sigma_\eta = 0.25$ and sample size is small, which masks the story told by RMSE. For a given value of d, the larger the magnitude of σ_η (i.e. increased strength of the volatility innovation relative to that of the additive noise), the smaller the RMSE and bias in estimation of d. The evidence is fairly consistent across different sample sizes and the true values of d. The only exception is the case of $d < -0.3$ and $\sigma_\eta < 0.35$ with sample size of 2000, where the bias decreases while the RMSE increases slightly as σ_η increases. In terms of the direction of bias, d tends to be overestimated when $\sigma_\eta < 0.55$, in particular when sample size is 2000. This upward bias tends to decrease as sample size increases, however the upward bias is still clear when $d < 0.22$ and $\sigma_\eta < 0.55$, even with the sample size as large as 16000.

Table 4.5: Sensitivity analysis for the memory parameter d with sample size of 2000

	d=0.22		0.26		0.30		0.34		0.38		0.42		0.46	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
σ_η=0.25	0.0737	0.1625	0.0783	0.1767	0.0956	0.1902	0.1018	0.1905	0.1011	0.1952	0.0841	0.1784	0.0894	0.1663
0.35	0.0560	0.1675	0.0606	0.1769	0.0705	0.1780	0.0588	0.1779	0.0620	0.1634	0.0378	0.1285	0.0410	0.1020
0.45	0.0437	0.1586	0.0296	0.1601	0.0322	0.1546	0.0266	0.1449	0.0119	0.1126	0.0157	0.1030	0.0217	0.0770
0.55	0.0287	0.1418	0.0133	0.1396	0.0111	0.1283	0.0086	0.1138	0.0029	0.0988	0.0034	0.0800	0.0139	0.0655
0.65	0.0028	0.1256	-0.0001	0.1149	0.0024	0.1006	-0.0063	0.0887	-0.0039	0.0829	0.0025	0.0703	0.0098	0.0560
0.75	-0.0053	0.1013	0.0028	0.0892	-0.0073	0.0828	-0.0096	0.0802	-0.0019	0.0724	0.0031	0.0628	0.0082	0.0503
0.85	0.0003	0.0882	-0.0029	0.0757	-0.0030	0.0706	-0.0032	0.0666	-0.0034	0.0644	0.0001	0.0564	0.0036	0.0441
0.95	0.0049	0.0662	-0.0030	0.0660	-0.0023	0.0584	-0.0041	0.0579	-0.0043	0.0561	-0.0008	0.0523	0.0032	0.0425
1.05	0.0007	0.0595	-0.0003	0.0558	-0.0010	0.0517	-0.0015	0.0533	-0.0019	0.0486	-0.0026	0.0468	-0.0006	0.0390
1.15	-0.0005	0.0514	-0.0005	0.0512	-0.0011	0.0492	-0.0010	0.0501	-0.0030	0.0443	-0.0006	0.0451	0.0037	0.0384

Table 4.6: Sensitivity analysis for σ_η with sample size of 2000

	d=0.22		0.26		0.30		0.34		0.38		0.42		0.46	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
σ_η=0.25	-0.0279	0.2587	-0.0333	0.2635	-0.0517	0.2616	-0.0612	0.2551	-0.0657	0.2456	-0.0614	0.2256	-0.0909	0.2228
0.35	0.0061	0.2755	0.0388	0.2532	-0.0111	0.2563	-0.0349	0.2425	-0.0556	0.2313	-0.0409	0.2027	-0.0532	0.1693
0.45	0.0453	0.2756	0.0310	0.2457	0.0064	0.2428	-0.0022	0.2138	-0.0148	0.2026	-0.0236	0.1903	-0.0447	0.1623
0.55	0.0271	0.2408	0.0270	0.2461	0.0031	0.2125	-0.0022	0.2052	-0.0112	0.2004	-0.0123	0.1678	-0.0283	0.1479
0.65	0.0332	0.2155	0.0209	0.2033	0.0086	0.2006	-0.0013	0.1882	0.0075	0.1712	-0.0139	0.1618	-0.0271	0.1443
0.75	0.0358	0.2090	0.0090	0.1932	0.0074	0.1777	0.0264	0.1741	0.0002	0.1685	-0.0088	0.1524	-0.0159	0.1356
0.85	0.0208	0.1871	0.0079	0.1708	0.0073	0.1657	0.0030	0.1621	0.0050	0.1608	-0.0035	0.1287	-0.0066	0.1239
0.95	0.0078	0.1521	0.0187	0.1594	0.0080	0.1471	0.0059	0.1491	0.0078	0.1503	0.0035	0.1379	-0.0057	0.1223
1.05	0.0113	0.1380	0.0121	0.1445	0.0106	0.1407	0.0010	0.1477	0.0013	0.1405	0.0017	0.1280	0.0001	0.1194
1.15	0.0074	0.1395	0.0066	0.1405	0.0018	0.1345	0.0026	0.1343	0.0032	0.1263	0.0015	0.1281	-0.0150	0.1153

Table 4.7: Sensitivity analysis for the memory parameter d with sample size of 4000

	d=0.22		0.26		0.30		0.34		0.38		0.42		0.46	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
σ_η=0.25	0.0641	0.1558	0.0591	0.1641	0.0640	0.1717	0.0593	0.1716	0.0359	0.1439	0.0400	0.1281	0.0446	0.1090
0.35	0.0470	0.1531	0.0559	0.1564	0.0386	0.1480	0.0186	0.1265	0.0069	0.1065	0.0161	0.0921	0.0171	0.0717
0.45	0.0257	0.1385	0.0193	0.1371	0.0018	0.1188	-0.0020	0.0945	-0.0038	0.0835	-0.0008	0.0708	0.0105	0.0589
0.55	0.0102	0.1082	-0.0049	0.0982	0.0007	0.0930	-0.0049	0.0783	-0.0041	0.0728	0.0024	0.0578	0.0062	0.0491
0.65	0.0025	0.0862	-0.0003	0.0745	-0.0002	0.0677	-0.0035	0.0648	0.0001	0.0599	-0.0010	0.0546	0.0049	0.0456
0.75	-0.0013	0.0698	-0.0038	0.0613	-0.0004	0.0548	-0.0006	0.0521	-0.0068	0.0527	-0.0045	0.0473	0.0064	0.0402
0.85	0.0005	0.0507	-0.0021	0.0488	-0.0003	0.0466	-0.0039	0.0449	0.0004	0.0429	-0.0048	0.0437	-0.0016	0.0345
0.95	0.0007	0.0451	-0.0020	0.0438	-0.0053	0.0417	-0.0023	0.0382	-0.0039	0.0395	-0.0060	0.0380	-0.0013	0.0316
1.05	-0.0029	0.0398	-0.0028	0.0390	-0.0012	0.0368	-0.0023	0.0345	0.0002	0.0360	-0.0009	0.0352	-0.0006	0.0295
1.15	-0.0011	0.0370	0.0007	0.0331	-0.0003	0.0332	-0.0033	0.0330	-0.0034	0.0336	-0.0006	0.0325	-0.0021	0.0285

Table 4.8: Sensitivity analysis for σ_η with sample size of 4000

	d=0.22		0.26		0.30		0.34		0.38		0.42		0.46	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
σ_η=0.25	-0.0113	0.2151	0.0012	0.2139	-0.0190	0.2119	-0.0198	0.1903	-0.0213	0.1808	-0.0416	0.1690	-0.0581	0.1527
0.35	0.0373	0.2199	0.0098	0.2091	-0.0233	0.1989	-0.0151	0.1811	-0.0097	0.1639	-0.0326	0.1543	-0.0343	0.1298
0.45	0.0261	0.2107	0.0069	0.1928	0.0148	0.1830	-0.0077	0.1701	-0.0013	0.1513	-0.0123	0.1325	-0.0229	0.1219
0.55	0.0132	0.1780	0.0134	0.1767	0.0044	0.1697	0.0055	0.1500	0.0034	0.1421	-0.0103	0.1181	-0.0137	0.1068
0.65	0.0187	0.1625	0.0043	0.1511	-0.0017	0.1441	0.0070	0.1317	-0.0027	0.1302	0.0027	0.1160	-0.0161	0.1080
0.75	0.0174	0.1447	0.0078	0.1360	-0.0038	0.1308	0.0021	0.1219	0.0108	0.1231	0.0089	0.1151	-0.0204	0.1049
0.85	0.0085	0.1224	0.0058	0.1154	-0.0015	0.1160	0.0086	0.1162	-0.0027	0.1119	0.0102	0.1101	-0.0036	0.0904
0.95	0.0073	0.1097	0.0041	0.1074	0.0125	0.1107	0.0041	0.1007	0.0091	0.1034	0.0122	0.1009	-0.0027	0.0905
1.05	0.0114	0.1095	0.0009	0.1018	0.0022	0.1051	0.0030	0.0987	-0.0038	0.0983	-0.0007	0.0966	-0.0024	0.0905
1.15	0.0053	0.0965	0.0020	0.0912	0.0033	0.0980	0.0049	0.0940	0.0055	0.0951	0.0012	0.0928	-0.0018	0.0866

Table 4.9: Sensitivity analysis for the memory parameter d with sample size of 8000

	d=0.22		0.26		0.30		0.34		0.38		0.42		0.46	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
σ_η=0.25	0.0661	0.1434	0.0644	0.1591	0.0388	0.1480	0.0230	0.1395	0.0148	0.1171	0.0177	0.0925	0.0254	0.0756
0.35	0.0310	0.1340	0.0191	0.1272	0.0030	0.1148	-0.0055	0.0945	-0.0008	0.0831	0.0071	0.0653	0.0095	0.0546
0.45	0.0052	0.1069	-0.0012	0.0907	-0.0075	0.0792	-0.0033	0.0729	-0.0002	0.0627	-0.0022	0.0543	0.0052	0.0440
0.55	-0.0013	0.0745	-0.0043	0.0698	-0.0047	0.0613	-0.0011	0.0549	-0.0048	0.0497	-0.0052	0.0463	0.0006	0.0348
0.65	-0.0037	0.0536	-0.0010	0.0477	-0.0028	0.0483	-0.0047	0.0447	0.0022	0.0416	-0.0039	0.0389	0.0018	0.0336
0.75	-0.0031	0.0448	-0.0016	0.0387	0.0011	0.0381	-0.0041	0.0384	-0.0022	0.0377	-0.0029	0.0334	-0.0004	0.0307
0.85	0.0015	0.0343	-0.0002	0.0341	-0.0014	0.0354	-0.0027	0.0331	-0.0032	0.0309	-0.0025	0.0297	-0.0004	0.0259
0.95	0.0020	0.0314	-0.0002	0.0281	-0.0010	0.0278	-0.0005	0.0260	-0.0025	0.0295	-0.0014	0.0275	-0.0023	0.0246
1.05	0.0017	0.0282	-0.0006	0.0259	-0.0002	0.0265	0.0018	0.0256	-0.0013	0.0251	-0.0007	0.0229	-0.0028	0.0223
1.15	-0.0007	0.0233	-0.0003	0.0237	0.0002	0.0232	-0.0017	0.0234	-0.0005	0.0241	-0.0015	0.0215	-0.0008	0.0218

Table 4.10: Sensitivity analysis for σ_η with sample size of 8000

	d=0.22		0.26		0.30		0.34		0.38		0.42		0.46	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
σ_η=0.25	0.0040	0.1843	-0.0036	0.1786	-0.0005	0.1704	-0.0075	0.1580	-0.0226	0.1398	-0.0354	0.1297	-0.0391	0.1098
0.35	0.0221	0.1800	0.0060	0.1751	0.0095	0.1544	-0.0062	0.1392	-0.0080	0.1299	-0.0173	0.1074	-0.0175	0.0915
0.45	0.0198	0.1566	0.0117	0.1370	0.0157	0.1322	-0.0020	0.1258	-0.0076	0.1143	-0.0014	0.1028	-0.0130	0.0836
0.55	0.0086	0.1307	0.0091	0.1294	0.0092	0.1183	-0.0014	0.1091	0.0041	0.0993	0.0018	0.0931	-0.0069	0.0816
0.65	0.0114	0.1085	-0.0015	0.1021	0.0035	0.1050	0.0047	0.0999	-0.0065	0.0966	0.0046	0.0845	-0.0098	0.0790
0.75	0.0079	0.0953	0.0074	0.0925	0.0011	0.0881	0.0119	0.0924	0.0029	0.0915	0.0021	0.0809	-0.0031	0.0790
0.85	-0.0009	0.0849	0.0006	0.0858	-0.0011	0.0860	0.0076	0.0842	0.0064	0.0849	0.0032	0.0755	-0.0005	0.0703
0.95	0.0025	0.0781	0.0036	0.0743	0.0068	0.0761	0.0006	0.0712	0.0029	0.0811	-0.0001	0.0731	0.0030	0.0679
1.05	0.0019	0.0707	0.0013	0.0703	0.0015	0.0728	-0.0035	0.0736	0.0004	0.0703	0.0025	0.0620	0.0042	0.0641
1.15	0.0032	0.0676	-0.0004	0.0671	0.0025	0.0649	0.0036	0.0716	0.0005	0.0720	0.0001	0.0611	-0.0013	0.0663

Table 4.11: Sensitivity analysis for the memory parameter d with sample size of 16000

	d=0.22		0.26		0.30		0.34		0.38		0.42		0.46	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
σ_η=0.25	0.0343	0.1345	0.0282	0.1330	0.0128	0.1205	-0.0045	0.0982	-0.0005	0.0800	0.0061	0.0710	0.0063	0.0532
0.35	0.0073	0.1037	0.0027	0.0950	-0.0039	0.0797	-0.0097	0.0691	-0.0031	0.0575	-0.0033	0.0512	0.0042	0.0409
0.45	0.0034	0.0608	-0.0020	0.0601	-0.0013	0.0521	-0.0038	0.0479	-0.0068	0.0427	-0.0028	0.0408	0.0021	0.0346
0.55	0.0032	0.0483	0.0010	0.0421	-0.0029	0.0405	-0.0035	0.0376	-0.0020	0.0370	-0.0057	0.0323	-0.0052	0.0295
0.65	-0.0001	0.0358	0.0001	0.0321	0.0000	0.0312	-0.0039	0.0293	-0.0018	0.0296	-0.0017	0.0282	-0.0017	0.0248
0.75	-0.0006	0.0287	0.0006	0.0273	-0.0003	0.0257	-0.0024	0.0257	-0.0005	0.0248	-0.0018	0.0249	-0.0011	0.0235
0.85	-0.0005	0.0245	-0.0022	0.0229	-0.0007	0.0229	-0.0014	0.0217	-0.0011	0.0224	-0.0038	0.0206	-0.0022	0.0194
0.95	-0.0007	0.0212	-0.0006	0.0203	-0.0009	0.0204	-0.0002	0.0195	-0.0001	0.0194	-0.0011	0.0189	-0.0022	0.0182
1.05	-0.0001	0.0184	-0.0012	0.0191	-0.0003	0.0172	-0.0013	0.0182	-0.0006	0.0170	-0.0018	0.0171	-0.0023	0.0168
1.15	-0.0001	0.0175	0.0000	0.0166	-0.0010	0.0155	0.0002	0.0169	-0.0006	0.0151	-0.0015	0.0152	-0.0026	0.0158

Table 4.12: Sensitivity analysis for σ_η with sample size of 16000

	d=0.22		0.26		0.30		0.34		0.38		0.42		0.46	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
σ_η=0.25	0.0225	0.1576	0.0049	0.1447	0.0098	0.1352	0.0020	0.1182	-0.0113	0.1061	-0.0153	0.0940	-0.0141	0.0768
0.35	0.0036	0.1284	0.0074	0.1249	-0.0004	0.1187	0.0116	0.1035	-0.0030	0.0888	-0.0013	0.0827	-0.0108	0.0672
0.45	0.0041	0.1090	0.0103	0.1049	0.0012	0.0971	0.0017	0.0886	0.0058	0.0809	0.0022	0.0758	-0.0088	0.0676
0.55	-0.0011	0.0922	0.0019	0.0836	0.0028	0.0849	0.0036	0.0778	-0.0007	0.0782	0.0063	0.0656	0.0062	0.0621
0.65	0.0004	0.0794	0.0015	0.0753	-0.0021	0.0710	0.0039	0.0700	-0.0019	0.0661	-0.0018	0.0618	-0.0006	0.0582
0.75	0.0020	0.0679	0.0020	0.0658	0.0016	0.0614	0.0049	0.0614	0.0006	0.0610	0.0003	0.0572	0.0016	0.0547
0.85	0.0030	0.0591	0.0032	0.0557	0.0015	0.0601	0.0022	0.0566	0.0037	0.0584	0.0066	0.0510	0.0066	0.0513
0.95	0.0014	0.0570	0.0012	0.0548	0.0046	0.0557	-0.0015	0.0517	-0.0030	0.0550	0.0001	0.0506	0.0026	0.0497
1.05	-0.0001	0.0482	0.0022	0.0502	0.0020	0.0504	0.0033	0.0502	-0.0027	0.0481	0.0002	0.0477	0.0035	0.0478
1.15	0.0030	0.0491	0.0030	0.0462	0.0029	0.0469	0.0001	0.0489	0.0007	0.0444	0.0022	0.0442	0.0036	0.0463

For the estimation of σ_η, for a given value of the memory parameter, RMSE of the estimator of σ_η decreases as the magnitude of σ_η increases. This pattern is persistent across sample sizes and the values of d. For a given value of true σ_η, a more positive value of the true memory parameter is useful in reducing the RMSE of the estimated σ_η. However, this reduction in RMSE becomes almost negligible when the true value of σ_η is larger than 0.85 and sample size increases to 8000 or 16000. In terms of the direction of bias, σ_η tends to be overestimated when the true value of d is smaller than 0.34.

Overall, the estimation of both parameters appear to be more accurate as the magnitudes of d and σ_η increase given the same sample size. In other words, parameter estimation tends to be most difficult when the latent volatility is of low degree of persistence, and the magnitude of the additive noise in the observation equation (4.5) is relatively large. In these cases, increased sample size does little to help to reduce bias and RMSE. The general patterns observed here are largely in line with the simulation results in the previous section.

4.5 Empirical Application

In this section, we examine the degree of fractional integration in the volatility of the daily Yen/USD exchange rate spanning 26 years. The full sample spans the period from December 12, 1983 through April 30, 2009, with 6406 daily observations.[27] Daily return is defined as $100 \times \log(p_t/p_{t-1})$.

4.5.1 Preliminary Data Analysis

The time series plot of log returns is in Figure 4.7. The ACF plots of returns and various approximations to volatility, including (log) squared returns and absolute returns, are given in Figure 4.8. There is no significant autocorrelation in returns, which is consistent with the notion of efficient markets; the returns are approximately mean zero and serially uncorrelated. Meanwhile the evidence for volatility persistence is overwhelming

[27]The raw daily prices are obtained from the web site of Reserve Bank of Australian.

Figure 4.7: Daily returns for the Yen/USD from December 1983 to April 2009

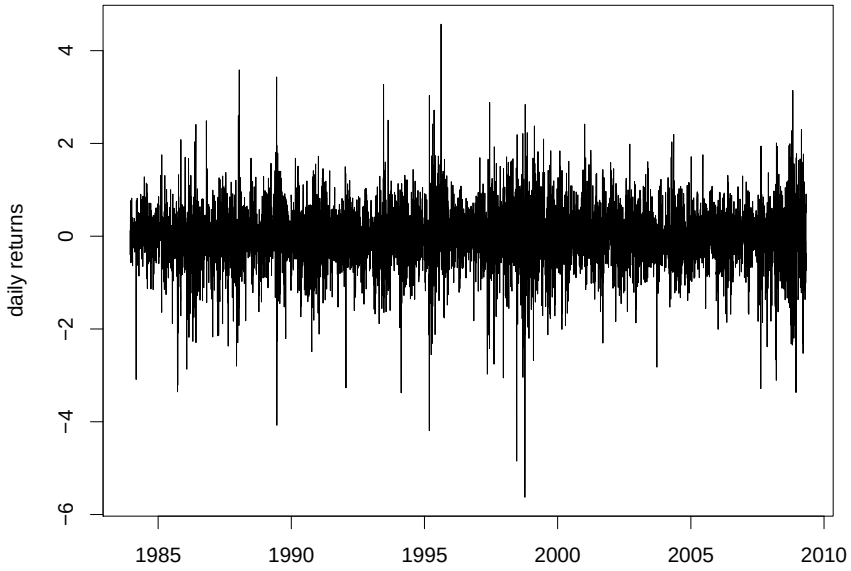
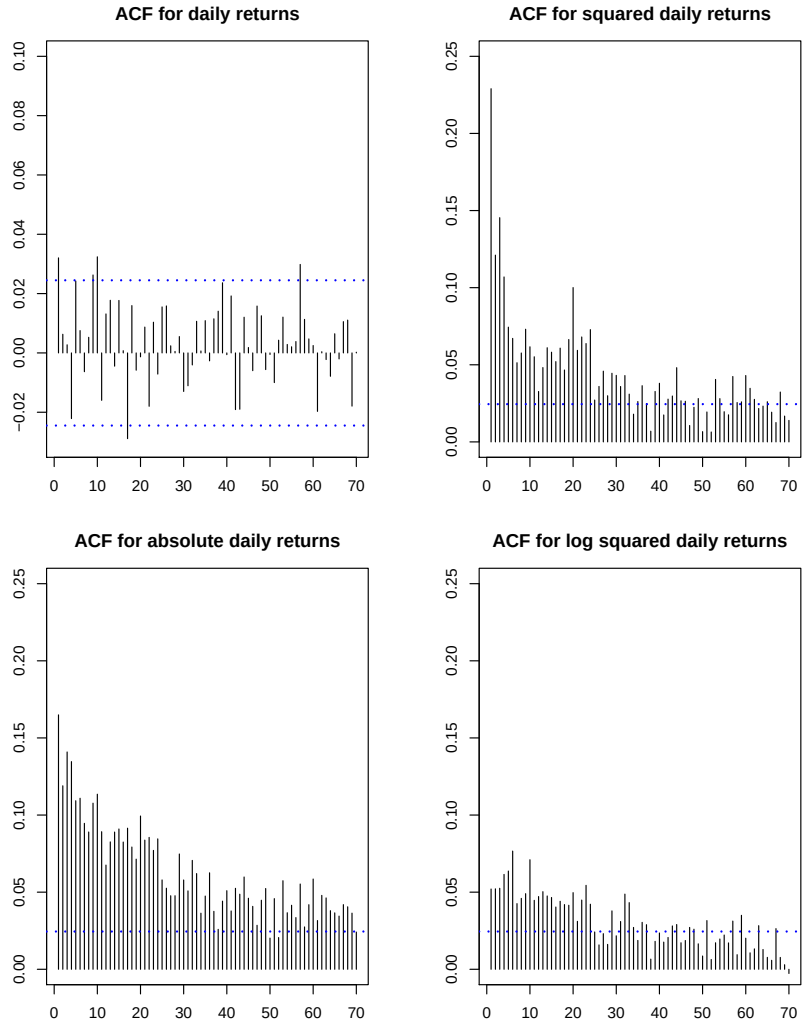


Figure 4.8: Autocorrelation functions for the daily returns and squared daily returns for the Yen/USD from December 1983 to April 2009



The dashed line represents upper or lower 95% confidence interval of autocorrelations for a white noise process.

as indicated by the ACF plots of various volatility approximations. The persistence of absolute returns appear to be the highest.

We also plot the ACF of power transformations of absolute returns following Ding et al. (1993) in Figure 4.9 to further examine the persistence in volatility. The results are consistent with their findings. The power transformation of the absolute returns $|y_t|^r$ also has quite high positive autocorrelations for long lags, which can be characterized as a long memory property. This property appears to be the strongest when r is around one. Autocorrelations appear to decrease fast in the first month or so and then decrease slowly. The transformation becomes noisier as the power increased further beyond one.

4.5.2 Estimation Results

The long memory SV model is fitted to the daily returns of the JPY/USD. Point estimates for the three methods are reported in Table 4.13. The posterior parameter mean is used as point estimator of Bayesian method, of which the MCMC results are based on 40 000 iterations after discarding the first 10 000 iterations to alleviate the effect of initialization. The order of moving average approximation to fractional noise is chosen as 30 for both time domain QML and Bayesian methods. Overall the estimates of time domain QML and Bayesian methods are close. The frequency estimate of d is slightly higher than that of the Bayesian method, while the estimate of σ_η is smaller.

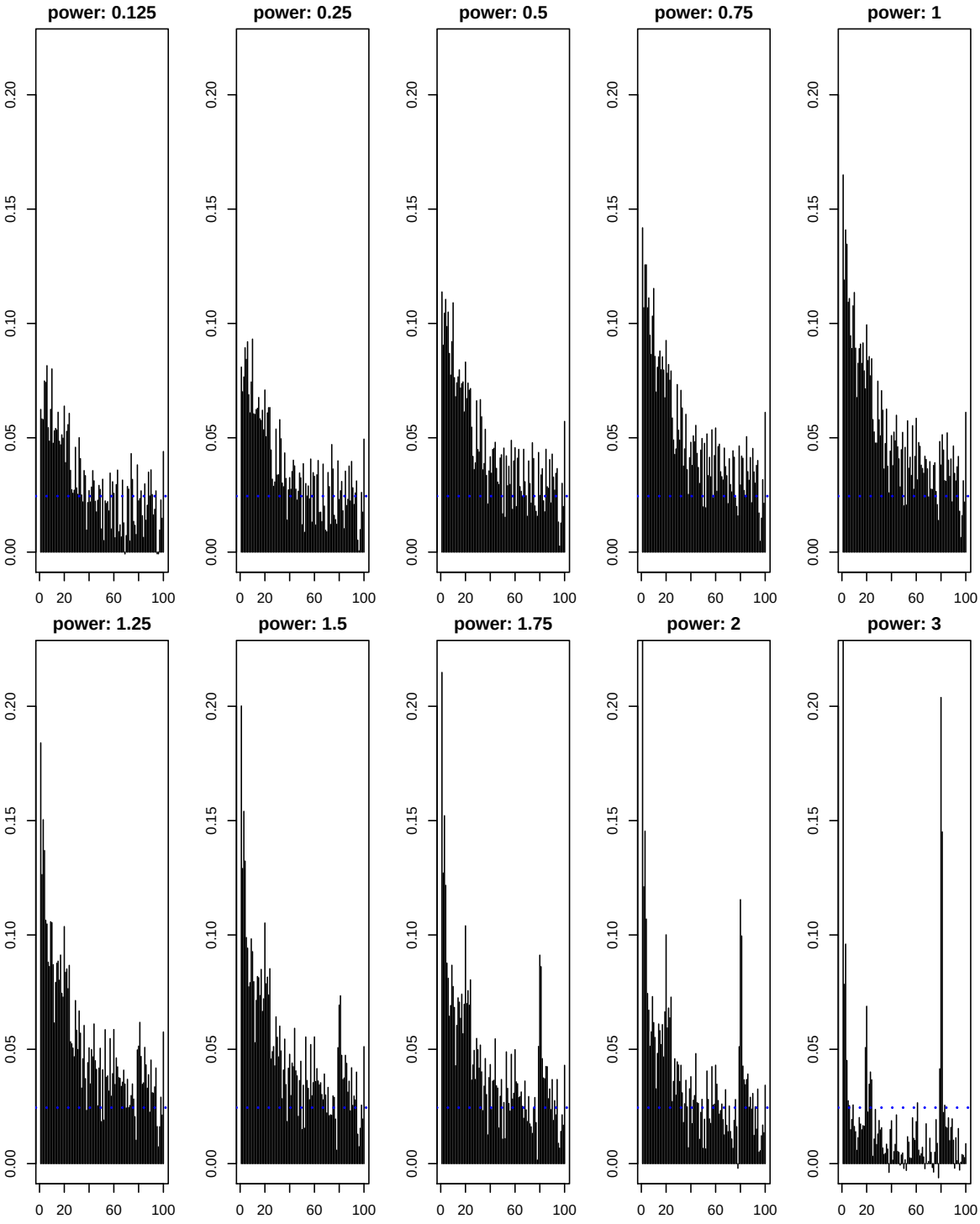
Table 4.13: QMLEs and posterior means of d and σ_η

		QML (frequency domain)	QML(time domain)	Bayesian
Long memory SV	d	0.439 (0.017)	0.391 (0.041)	0.408
	σ_η	0.602 (0.034)	0.763 (0.077)	0.718

Standard errors are provided in brackets, with the corresponding covariance matrix is approximated by the inverse of the negative of the hessian evaluated at QMLEs. In terms of numerical optimization, different algorithms are exercised, including the BFGS, Newron-Raphson and BHHH methods, and little difference is observed. The results reported here are based on the BFGS method. Standard errors are reported for reference as the estimation is a highly nonstandard exercise with parameters being constrained. The asymptotic distribution of QML estimators of long memory SV models is still unknown.

The GPH estimator is estimated from volatility approximations, (log) squared and absolute returns. An appropriate choice of bandwidth of GPH estimation is important

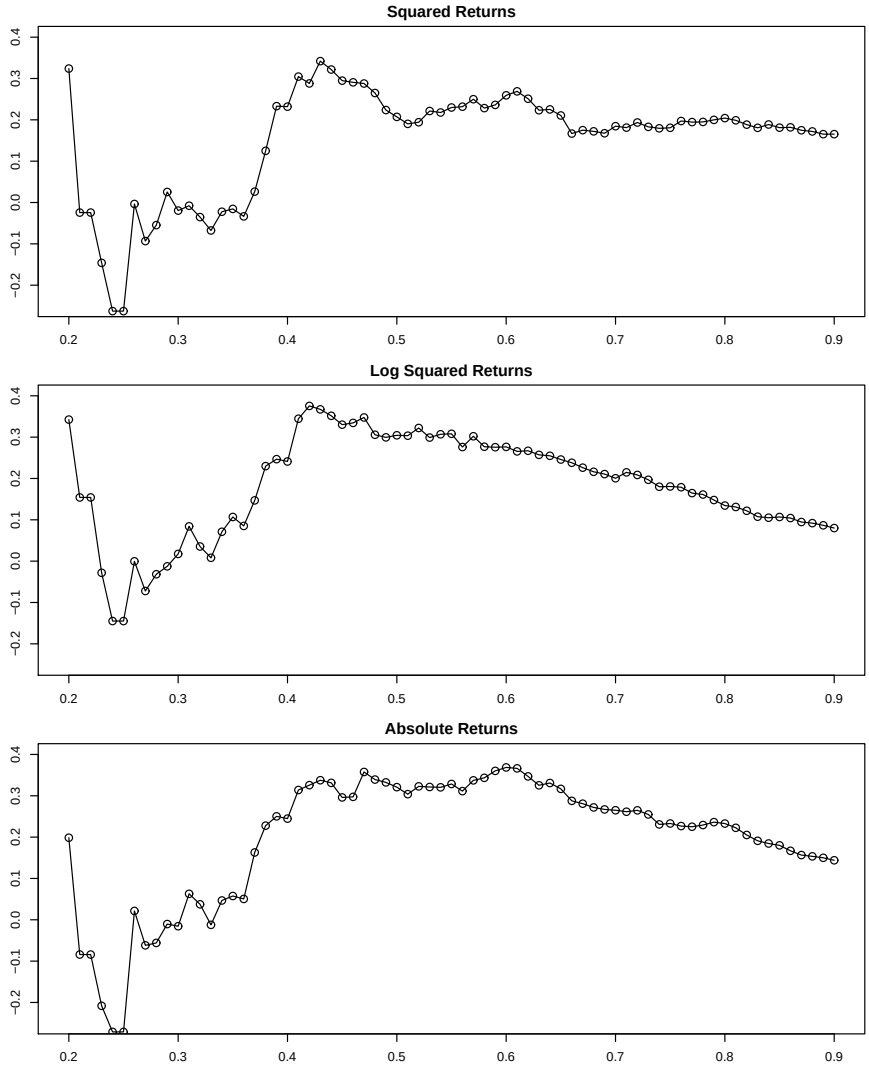
Figure 4.9: ACF of the power transformation of absolute returns ($|y_t|^P$) for the Yen/USD from December 1983 to April 2009



The dashed line represents upper or lower 95% confidence interval of autocorrelations for a white noise process.

as there is no optimal rule available. Figure 4.10 plots the GPH estimates obtained with different values of the bandwidth parameter. The persistence of $\log(y_t^2)$ and $|y_t|$ appears to be closer, and that of ${y_t}^2$ is the lowest. Commonly chosen values of the power r in $|y_t|^r$ are between 0.5 and 0.8, according to this rule, the memory parameter is estimated to be between 0.3 and 0.35. It is clear that estimates of the memory parameter are smaller than the other methods, and this downward bias is consistent with the discussions in our Monte Carlo study (Tables 4.1 and 4.2).

Figure 4.10: GPH estimates of d obtained from transformations of daily returns with varying bandwidths



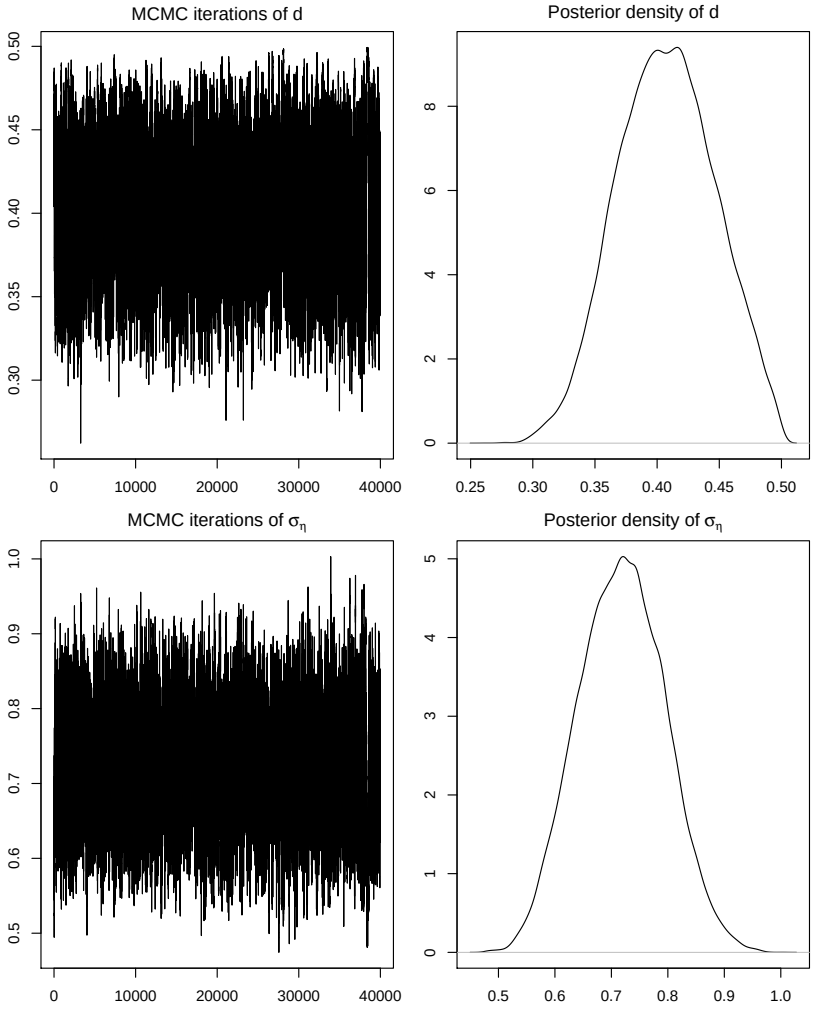
Each point represents a log-periodogram regression (GPH) estimate for memory parameter d for a given bandwidth. The sample period extends from December 1983 to April 2009, for a total of $T = 6406$ observations.

Summary statistics of the posterior density of d and σ_η^2 are reported in Table 4.14, with Figure 4.11 providing the trace plots of MCMC iterations and corresponding posterior density plots. The MCMC chain appears to converge quickly and the acceptance rate is around 0.6. It is clear that the memory parameter is well above 0.3, clear evidence of long memory property in volatility, which is consistent with our preliminary analysis.

Table 4.14: Posterior summaries of d and σ_η of MCMC method

		Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
Long memory SV	d	0.262	0.380	0.408	0.408	0.435	0.499
	σ_η	0.474	0.665	0.719	0.718	0.771	1.003

Figure 4.11: MCMC iterations and posterior density of parameters d and σ_η



Daily returns for the Yen/USD are from December 1983 to April 2009. The MCMC chains are based on 40 000 iterations after discarding the first 10 000 iterations as *burin-in* period. The order of moving average approximation to fractional noise is chosen as 30.

In Chapter 3, we examine the persistence of realized volatility constructed from intraday returns of the YEN/USD from January 2003 to September 2008. The memory parameter was found to be around 0.4, largely consistent with the results in this chapter. As argued by Bollerslev and Wright (2000), high-frequency data may be critically important for reliable inference concerning long-range dependence in volatility. In other words, given a fixed time span, say one year, the detection of long-range dependence is difficult when relying on low-frequency data. So it is expected that an increased time span, say 26 years, is crucial to obtaining reliable inference on the memory parameter, given daily frequency.

Given the estimated parameters, the state-space form of the long memory SV model is used to estimate the latent volatility $\tilde{h_t}$.[28] An estimate of the value of the scale parameter σ in (4.2) is required in order to estimate the conditional variance. We follow Harvey (1998) in constructing

$$\tilde{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \tilde{y_t}^2, \quad \tilde{y_t} = y_t \exp(-\tilde{h_t}/2)$$

so the estimator of the underlying conditional variance is

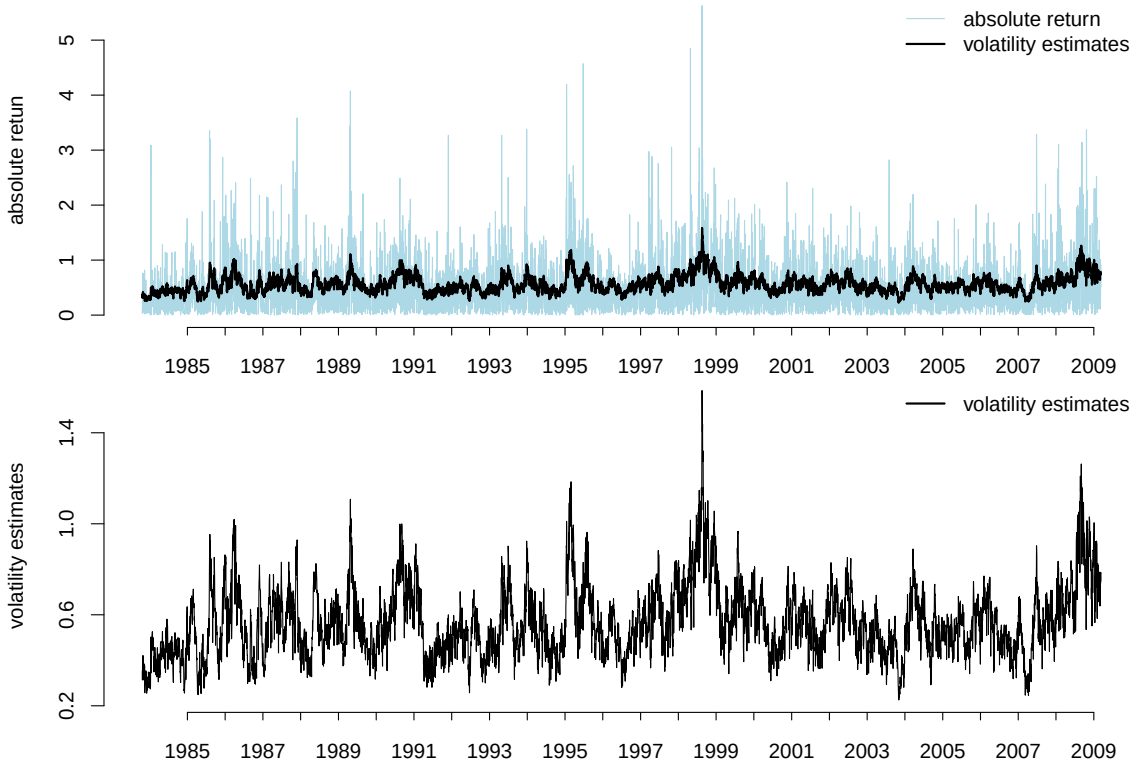
$$\tilde{\sigma_t}^2 = \tilde{\sigma}^2 \exp(\tilde{h_t})$$

We find $\tilde{\sigma}$=0.55 given the posterior mean of d and σ_η. Figure 4.12 plots the daily volatility estimates. The top panel of Figure 4.12 plots the smoothed volatility and absolute returns. The latter quantity is obviously much noisier than the former, although it provides stronger evidence of volatility persistence than other transformations $|y_t|^r$. The smoothed volatility shown in the bottom panel suggests a stationary process in daily volatility, with occasionally high volatility episodes. For example, the largest spike occurred around October 1998,

[28] Volatility estimates can be obtained by *particle filtering*, a method recently proposed in the literature. Its advantage is that it can make use of equation (4.2) directly so that the normality assumption on ξ_t is avoided, but it is complicated and time-consuming although it promises more accurate volatility estimates. The implementation of particle filtering on short memory SV models has been examined by Kim et al. (1998) and Pitt and Shephard (1999); see also Johannes and Polson (2008) and references therein for econometric applications, and Doucet et al. (2001) for a book length treatment in a general discipline.

caused by the Russian financial crisis,[29] and the recent financial turmoil observed at the end of our sample.

Figure 4.12: Smoothed estimates of the JPY/USD volatility and absolute returns



The smoothed volatility is obtained by setting d=0.41, σ_η=0.72 and σ=0.55 which are the posterior mean of Bayesian estimates.

4.6 Conclusions and Extensions

In this chapter, we propose a Bayesian estimator of the long memory SV model, which is based on a state-space representation of the SV model and moving-average approximation to a fractional Gaussian process. When conducting the MCMC algorithm on parameter estimation, we update the memory parameter and volatility-of-volatility parameter in one block by making use of the optimization result from time domain QML estimation. The simulation results indicate that the proposed estimator outperforms clas-

[29] See Maekawa and Xinhong (2009) for the analysis of the Russian financial crisis. Hedge funds liquidate the open positions of Yen on October 1998 resulting in sudden and sharp rises in exchange rate for Yen.

sical methods in terms of root mean squared error when sample size is 2000 and 4000. We conduct an empirical estimation on daily JPY/USD spanning almost three decades, and find clear evidence of long-range dependence in volatility. The degree of persistence is found to be consistent with that of our estimation results obtained in Chapter 3 when realized volatility constructed from intra-day returns is used.

We leave two questions for future research. First, only the basic fractionally integrated process with no short term dynamics is considered for the latent log volatility process in this chapter, it is desirable to introduce more flexibility by using the specification of autoregressive and moving-averaging fractional integration. Another interesting question is to improve the convergence performance of the MCMC method when the persistence in volatility is of low degree and the magnitude of the additive noise in the observation equation (4.5) is relatively large.

Chapter 5

Realized Volatility Modeling with Regime Switching

5.1 Introduction

One of the stylized facts of realized volatility is a high degree of persistence shown as hyperbolically decaying autocorrelation functions, rather than exponentially as in the case of short memory models. A natural choice to capture this dynamic is through fractionally integrated processes, as it is well recognized in the literature that an integrated process $I(d)$, with d being a fraction, can generate autocorrelations with a slowly decaying pattern (see Andersen et al. (2003) for applications of fractional integration models in the estimation realized volatility). The long memory GARCH model of Baillie et al. (1996) and long memory stochastic volatility model of Breidt et al. (1998) are applications of fractional integration models applied to data with daily frequency.

There is a large literature on applying specifications other than fractional integration processes to mimic long memory properties, and some of these applications are superior in intra-sample fit and out-of sample forecasting. Alternative specifications include first, structural breaks and regime switching models, which are in the realm of non-linear time series; and second mixtures of two or three short memory processes, for example, using the

sum of two or more autoregressive moving average (ARMA) processes. For applications of structural breaks and/or regime switching models in this context, see, for example, Liu (2000), Andreou and Ghysels (2002), Franses, Leij and Paap (2002), Granger and Hyung (2004) and Calvet and Fisher (2004). These papers are based on daily squared or absolute returns, and Liu (2000) in particular studies the cause of high persistence in volatility caused by regime switching. Andreou and Ghysels (2002) argue that persistence in volatility may be overstated with the presence of structural breaks. The presence of breaks can also explain long memory in volatility, and Granger and Hyung (2004) argue that, based on their theory and simulation results, it is not easy to distinguish between occasional-break and fractional integration ($I(d)$) models because both can generate slowly-decaying autocorrelations. Calvet and Fisher (2004) document improved performance of long-run volatility forecasting when a Markov-switching process is applied, while Banerjee and Urga (2005) present detailed discussions on the importance of structural breaks and regime switching as sources of long memory. Applications using mixtures of short memory factors, can be found in Gallant et al. (1999), who use a sum of two short memory components to mimic long memory in daily returns in IBM stocks; while Alizadeh et al. (2002) observe that the sum of two autoregressive AR(1) components is able to capture long-range persistence in volatility. Barndorff-Nielsen and Shephard (2002*a*) is based on high-frequency data, where a multiple-factor stochastic volatility model is proposed to capture the slowly decaying autocorrelations in exchange rate realized volatility. This contrasts with the analysis of Alizadeh et al. (2002) where range, i.e. difference between maximum and minimum intraday returns is used as a volatility proxy.

To our knowledge, there have been few applications of the two alternative methods to realized volatility. The only exception is Barndorff-Nielsen and Shephard (2002*a*), who use a sum of short memory processes to mimic long-range dependence. However, there do not appear to be any applications of using regime switching to model realized volatility when interest is in the high degree of persistence in volatility. This motivates our exercise in this chapter. As shown in a series of papers, including Andersen, Bollerslev, Diebold and Labys (2001) and Andersen et al. (2003), realized volatility can provide more accurate estimates of the latent volatility than traditional volatility approximations, such as squared

or absolute returns. It is reasonable to assume that the promising results obtained with daily squared or absolute returns in previous papers would carry over to realized volatility. In addition, it is of practical relevance to compare the three methodologies, especially in terms of forecasting, when applied to realized volatility. We conduct a comparison of the three approaches based on a decade of daily realized volatility of the JPY/USD rate, using criteria of in-sample fit and out-of-sample forecasting. Models examined in this chapter include Markov-switching and three-factor stochastic volatility models. The motivation for the first method is due to Diebold and Inoue (2001) who argue that regime-switching can mimic fractional integration extremely well according to their theoretical and simulation studies; three-factor stochastic volatility model is considered following Barndorff-Nielsen and Shephard (2002*a*).

It is not an easy task to decide which is the best model to use since each of the three approaches can mimic long-range dependence. For example, a long memory hypothesis test with the null of short memory and a regime switching hypothesis test with the null of no regime switching are conducted on the time series examined, and the results clearly reject the null in both tests. So this chapter conducts a recursive estimation exercise in order to evaluate the performance of each model. These estimates indicate that regime switching outperforms fractional integration models, with reference to residual diagnostics and long-term forecasting. This may be due to the flexibility of regime switching models, which allows level and innovation variance (volatility of volatility) to switch between regimes. With a time series spanning a decade, it is likely unduly restrictive to assume volatility is characterized by constant memory, autoregressive and moving average parameters, as in fractional integration models.

Maximum likelihood estimation (MLE) is employed in this chapter. This is because Bayesian estimation is computationally costly as there are several models to be dealt with, and Bayesian method is much more time consuming than MLE especially when recursive estimations are conducted.[1] The next section discusses the methodology, including model specification and estimation, while Section 5.3 provides results of the empirical study of

[1]Filardo and Gordon (1998) discuss Bayesian estimation of regime-switching models. Koop and Potter (1999) provide a treatment of Bayesian model selection for nonlinear models.

daily Yen/USD realized volatility. Conclusions are presented in Section 5.4.

5.2 Methodology

This section is concerned with likelihood evaluation and estimation for each of the models considered. Hypothesis testing of long-memory and identification of regime-switching models are also discussed.

5.2.1 Fractional Integration Process

Let y denote the time series of interest, and y_t be a realization observed on day t with $t = 1, \ldots, T$. A fractional autoregressive and moving-average process ARFIMA(p,d,q) is

$$\phi(B)(1-B)^d(y_t - \mu) = \theta(B)\eta_t \qquad \eta_t \sim N(0, \sigma^2\eta) \tag{5.1}$$

where μ is the unconditional mean of y_t, and η_t is the innovation to the process. The orders of the autoregressive (AR) and moving-average (MA) polynomials are p and q, with $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$. The stationarity and identifiability conditions are $|d|$ <0.5, the roots of $\phi(z)$=0 and $\theta(z)$=0 lie outside the unit circle, and $\theta(z)$ has no roots in common with those of $\phi(z)$. In this chapter, positivity of d is further imposed because autocorrelations in volatility are positive according to empirical evidence. Section 3.2 in Chapter 3 provides further detailes on ARFIMA models.

The key characteristic of a long memory process is about the decaying pattern in the autocorrelation function rather than the magnitude of autocorrelation at a particular lag: see the discussion in Beran (1994). So it might be difficult to identify a long memory property merely based on visual inspection of sample autocorrelations at sufficiently long lags. A battery of tests is desirable prior to fitting data into the ARFIMA models. Some popular long memory tests in the literature are the classical rescaled range test due to Hurst (1951), the modified rescaled range test proposed in Lo (1991) and Giraitis et al. (2003). All these tests are designed with the null hypothesis of short memory against the

alternative of long memory. This chapter applies the last two tests in the empirical analysis largely due to their robustness to the possible mis-specification of short run dynamics. In addition, long-range dependence can be gauged by the estimated long memory parameter d under the assumption that ARFIMA(p,d,q) is the genuine data generating process and the orders of p and q are specified adequately.

5.2.1.1 Estimation and Likelihood Evaluation of ARFIMA Models

The short term dynamics in (5.1), captured by the AR and MA polynomials, combined with the memory parameter allows for a broad class of autocorrelation functions. This rich dynamic is desirable but complicates model estimation since the fractional difference operator $(1-B)^d$ is a binomial expansion of infinite order. This chapter uses the algorithms of Sowell (1992) to evaluate the likelihood function when MLE is conducted. Section 3.3.1 in Chapter 3 discusses some issues related to the exact likelihood evaluation of ARFIMA models. We reproduce some equations for convenience; see Beran (1994) and Baillie (1996) for further details on ARFIMA models and their estimations.

The sample mean is used as an estimator of μ, following Yajima (1988) who argues there is only a small loss in efficiency if a fractionally integrated process is demeaned by its sample mean. Let $\tilde{y}$ denote the demeaned series, and $\beta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, d)$. The autocovariance function of $\tilde{y}$ is $\gamma_i = E[\tilde{y}_t \tilde{y}_{t-i}]$, and

$$\mathbf{\Sigma} = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \ddots & \vdots \\ \gamma_2 & \gamma_1 & \gamma_0 & \ddots & \gamma_2 \\ \vdots & \ddots & \ddots & \ddots & \gamma_1 \\ \gamma_{T-1} & \cdots & \gamma_2 & \gamma_1 & \gamma_0 \end{pmatrix}$$

defines the covariance matrix of the joint distribution of $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_T)'$. The quantity γ_i is a function of β; see Sowell (1992, equations 8-9) for its calculations. Given the

normality of η_t, $\tilde{y} \sim N_T(0, \Sigma)$. The likelihood function of $\tilde{\mathbf{y}}$ is

$$L(\beta, \sigma_\eta^2) = (2\pi)^{-T/2}|\mathbf{\Sigma}|^{-1/2} \exp\{ -\frac{1}{2}\tilde{\mathbf{y}}'\mathbf{\Sigma}^{-1}\tilde{\mathbf{y}}\} \tag{5.2}$$

where $\mathbf{\Sigma}$ is a T×T Toeplitz matrix. A numerical optimization of $L(\cdot)$ is used for MLE. The inversion of $\mathbf{\Sigma}$ could be computationally expensive, especially for a large sample size T.[2]

Let h denote the forecasting horizon of interest, the best linear prediction of $\tilde{y}_{T+h}$ is,[3]

$$\hat{\tilde{y}}_{T+h} = [\gamma_{(T-1+h)} \ \cdots \ \gamma_h]\mathbf{\Sigma}^{-1} \ \tilde{\mathbf{y}}$$

The free *Ox* package *Arfima* of Doornik and Ooms (2006*b*) is used for all the estimation and forecasting of ARFIMA models in this chapter.

5.2.2 Regime Switching Models

It is argued that abrupt change is one of the common features of many economic variables, in particular, financial time series. It is realistic to assume that time series spanning a fairly long period might undergo episodes defined by different regimes. In other words, the assumption that the level of the time series, and/or the magnitude of innovations remain constant over years could be too restrictive. This motivates regime switching models. In general two types of switching models are commonly used in non-linear time series modeling.[4] One is the threshold autoregressive type model due to Tong (1978), where the switching is determined by a threshold variable which is observable. The other type of model treats a regime as unobservable state: the Markov switching model,

[2]Sowell (1992) derives several tricks for recursively computing γ_i, which is further refined by Doornik and Ooms (2003). Doornik and Ooms (2003) and Doornik and Ooms (2006*b*, pp. 22-27) discuss the computational aspects of MLE of ARFIMA models in detail. The GPH estimator is semi-parametric, and the knowledge of p and q are not necessary. Therefore it is more robust to model misspecification than MLE. However, it could be problematic in terms determining the values of the bandwidth necessary for estimation. Another drawback of semi-parametric methods is that out-of-sample forecasting is not readily available, and two-step approach needs to be implemented.

[3]The corresponding theoretical justification is provided by Beran (1994, Ch. 8).

[4]Hamilton (2005) provides an accessible introduction; see Kim and Nelson (1999) and Franses and van Dijk (2000) for text book treatments of regime switching models, with the latter focusing on applications of financial time series.

popularized in econometric applications by Hamilton (1989, 1994) is the most common example. In this chapter, we follow the latter strand mainly due to its popularity and, important, the ability to mimic long range dependence as shown in Diebold and Inoue (2001). Estimation and specification testing of Markov switching models will be discussed in the following sections.

5.2.2.1 Markov Switching Model Specification

Suppose a time series y_t follows an ARMA process, but with the unconditional mean and innovation variance switching between two regimes. Let s_t denote the index of regime which y_t is in at date t, s_t can only take the value of 1 or 2 given two regimes,

$$\begin{cases} \phi(B)(y_t - \mu_1) = \theta(B)\eta_t, \quad \eta_t \sim N(0, \sigma_1^2) & \text{when } s_t = 1 \\ \phi(B)(y_t - \mu_2) = \theta(B)\eta_t, \quad \eta_t \sim N(0, \sigma_2^2) & \text{when } s_t = 2 \end{cases}$$

the shorthand of which is

$$\phi(B)(y_t - \mu_{s_t}) = \theta(B)\eta_t, \quad \eta_t \sim N(0, \sigma_{s_t}^2) \tag{5.3}$$

To complete the model specification, the probability law governing transition from one regime to the other is required. Following Hamilton (1989), the transition probability is

$$Pr(s_t = j | s_{t-1} = i, \ldots, s_1 = k, y_{t-1}, \ldots, y_1) = Pr(s_t = j | s_{t-1} = i) = p_{ij} \tag{5.4}$$

with $i = 1, 2$, $j = 1, 2$ and $k = 1, 2$, which is an Markov chain process with fixed transition probability. In particular,

$$Pr(s_t = 1 | s_{t-1} = 1) = p_{11}$$
$$Pr(s_t = 2 | s_{t-1} = 1) = p_{12}$$
$$Pr(s_t = 1 | s_{t-1} = 2) = p_{21}$$
$$Pr(s_t = 2 | s_{t-1} = 2) = p_{22}$$

where $p_{11} + p_{12} = 1$, $p_{22} + p_{21} = 1$ and p_{11} and p_{22} are non-negative by definition. The specification in (5.4) treats s_t as an unobservable variable, which needs to be inferred from

the observed time series.[5] The probability of changing regime depends on the past only through the regime state in the most recent time period.[6] This setup is analogous to the state space form discussed in Section 4.2.3 of Chapter 4, therefore state filtering and smoothing algorithms are readily applicable.

The specification of Markov switching implies the permanent structural change when $p_{22} = 1$, and also allows for more flexible dynamic by letting $p_{22} < 1$.[7] Of course, there is no reason to preclude the possibility of multiple regimes. We focus on two regimes specification in this chapter, mainly due to its simplicity and to easing the burden of numerical optimization imposed by increased number of parameters as a highly nonlinear model is dealt with.

5.2.2.2 Estimation of Regime Switching Models

Let $\boldsymbol{\omega} = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, p_{11}, p_{22}, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)'$ denote parameters to be estimated, and $\Omega_{t-1} = (y_{t-1}, y_{t-2}, \ldots, y_1)'$ represent the path of the time series up to date $t-1$.

The conditional probability density when regime j is operating is given by

$$f(y_t|s_t = j, \Omega_{t-1}; \boldsymbol{\omega}) \tag{5.5}$$

which is of the form as that of a normal ARMA(p,q) process. The likelihood function of an ARMA(p,q) process is widely discussed in time series analysis books, see, for example, Brockwell and Davis (1991, Ch. 8). To evaluate the likelihood for a single observation y_t, s_t needs to be integrated out as follows since it is unobservable:

$$f(y_t|\Omega_{t-1}; \boldsymbol{\omega}) = \sum_{i=1}^{2} \sum_{j=1}^{2} Pr(s_{t-1} = i|\Omega_{t-1}; \boldsymbol{\omega})\ p_{ij}\ f(y_t|s_t = j, \Omega_{t-1}; \boldsymbol{\omega}) \tag{5.6}$$

[5] It could also be desirable for some applications to assume that switching is driven by exogenous variables, or the lagged variable itself. In the case of volatility modeling, it might be useful to use other economic variables to explain the switching mechanism, given the fact that volatility essentially represents the overall level of information available to the market. We will pursue this in future work.

[6] The assumption of the probability of a regime change depending on the past only through the most recent regime state appears to be restrictive. However, as shown by Hamilton (1989), this is not a particularly restrictive assumption.

[7] See Hamilton (2005) for a detailed discussion.

where $Pr(s_{t-1} = i|\Omega_{t-1}; \boldsymbol{\omega})$ for $i = 1, 2$ is required to complete the above evaluation. This quantity is the filtered probability, representing the inference about the probability of being at regime i at date $t-1$ for $i = 1, 2$, based on the information up to date $t-1$. The filtered probability is obtained via the following iteration from date $t = 1$,

$$Pr(s_t = j|\Omega_t; \boldsymbol{\omega}) = \frac{\sum_{i=1}^{2} Pr(s_{t-1} = i|\Omega_{t-1}; \boldsymbol{\omega})\ p_{ij}\ f(y_t|s_t = j, \Omega_{t-1}; \boldsymbol{\omega})}{f(y_t|\Omega_{t-1}; \boldsymbol{\omega})} \tag{5.7}$$

for $t = 1, \ldots, T$. In order to start the iteration, initialization of $Pr(s_0 = i)$ is necessary. Two types of initial values are commonly used: 0.5, and the unconditional probability of

$$Pr(s_0 = i) = \frac{1 - p_{jj}}{2 - p_{ii} - p_{jj}}$$

In the empirical estimation conducted in this chapter, both options are tried and the results appear to be robust for either choice.

The logarithm likelihood function of the regime switching specification in (5.3) is

$$\log f(y_1, y_2, \ldots, y_T|y_0; \boldsymbol{\omega}) = \sum_{t=1}^{T} \log f(y_t|\Omega_{t-1}; \boldsymbol{\omega}) \tag{5.8}$$

with quantities in (5.6) and (5.7) plugged in. The MLE of $\boldsymbol{\omega}$ are obtained via numerical optimization of (5.8). It is common practice in the estimation of regime switching models to impose local stationarity conditions within each state; we do this in this chapter. It is worth noting that local stationarity is neither sufficient nor necessary condition for global stationarity of the model. See Timmermann (2000) and Stelzer (2009) for moments of Markov switching models. The estimation is a non-standard exercise given that s_t is unobservable. Different starting values of the parameters $\boldsymbol{\omega}$ are worth trying to ensure a global maximum is achieved.[8] Conditional on the MLE of $\boldsymbol{\omega}$, $\hat{\boldsymbol{\omega}}_{MLE}$, one can infer the filtered probability $Pr(s_t = j|\Omega_t; \boldsymbol{\omega})$ for $j = 1, 2$ through (5.7).[9]

Different specifications of autoregressive and moving-average orders, combined with switched level and variance, allow for the model to generate richer dynamics than linear

[8] Hansen (1992) argues that the model produces numerous local optima as it is highly nonlinear.

[9] To obtain the smoothed probability of $Pr(s_t = j|\Omega_T; \boldsymbol{\omega})$ with $\Omega_T = (y_T, y_{T-1}, \ldots, y_1)'$ for $j = 1, 2$, one can refer to Chapter 22 in Hamilton (1989, Ch. 22).

models. Once in the realm of nonlinear models, more options are available. Of course, there is no reason to preclude the specification of a long memory process which allows the memory parameter to switch between high and low-persistence regimes, although this is not pursed here.

The following ARFIMA specification with switching mean and variance is also tried in our empirical exercise.

$$\phi(B)(1-B)^{d_{s_t}}(y_t-\mu_{s_t})=\theta(B)\eta_t, \quad \eta_t\sim N(0,\sigma^2_{s_t}) \qquad \text{or} \tag{5.9}$$

$$\begin{cases} \phi(B)(1-B)^{d_1}(y_t-\mu_1)=\theta(B)\eta_t, & \eta_t\sim N(0,\sigma^2_1) \quad \text{when } s_t=1 \\ \phi(B)(1-B)^{d_2}(y_t-\mu_2)=\theta(B)\eta_t, & \eta_t\sim N(0,\sigma^2_2) \quad \text{when } s_t=2 \end{cases}$$

where the probability law in (5.4) determines the transition from one regime to the other.

5.2.2.3 Forecasting Using Regime Switching Models

Given ML estimates of $\boldsymbol{\omega}$, prediction is obtained in the following fashion. Two quantities might be of interest in terms of forecast: regime forecast, and forecast for observed time series, say, multi-step realized volatility predictions in our example. Suppose we are interested in one-step ahead volatility prediction. In this case, the conditional density $f(y_{\boldsymbol{T+1}}|s_{\boldsymbol{T+1}}=j,\Omega_{\boldsymbol{T}};\boldsymbol{\omega})$ can be a starting point. The condition is upon regime at date $T+1$ so its expectation is readily available as it is a normal ARMA specification. However, the regime is unobservable and it is desirable to have the prediction with uncertainty about regime integrated out as follows:

$$E[y_{\boldsymbol{T+1}}|\Omega_{\boldsymbol{T}};\boldsymbol{\omega}]=\sum_{j=1}^{2}Pr(s_{\boldsymbol{T+1}}=j|\Omega_{\boldsymbol{T}};\boldsymbol{\omega})E[y_{\boldsymbol{T+1}}|s_{\boldsymbol{T+1}}=j,\Omega_{\boldsymbol{T}};\boldsymbol{\omega}]$$

A forecast for the probability of regime j operating is necessary in order to obtain $E[y_{\boldsymbol{T+1}}|\Omega_{\boldsymbol{T}};\boldsymbol{\omega}]$. With specification of two regimes, we have

$$\begin{aligned} Pr(s_{\boldsymbol{T+1}}=1|\Omega_{\boldsymbol{T}};\boldsymbol{\omega}) &= p_{11}Pr(s_{\boldsymbol{T}}=1|\Omega_{\boldsymbol{T}};\boldsymbol{\omega})+p_{21}Pr(s_{\boldsymbol{T}}=2|\Omega_{\boldsymbol{T}};\boldsymbol{\omega}) \\ Pr(s_{\boldsymbol{T+1}}=2|\Omega_{\boldsymbol{T}};\boldsymbol{\omega}) &= p_{12}Pr(s_{\boldsymbol{T}}=1|\Omega_{\boldsymbol{T}};\boldsymbol{\omega})+p_{22}Pr(s_{\boldsymbol{T}}=2|\Omega_{\boldsymbol{T}};\boldsymbol{\omega}) \end{aligned}$$

where the filtered probability $Pr(s_{\boldsymbol{T}} = j|\Omega_{\boldsymbol{T}};\boldsymbol{\omega})$ is available with (5.7). Generalization to multiple-step prediction is straightforward, see Hamilton (1994, p. 694) for details.

5.2.2.4 Regime Identification Testing

Before fitting the data to the regime switching model, it is sensible to test the null hypothesis of one regime against the alternative hypothesis of two regimes. The test result might justify whether there is any regime switching at all, or whether there are in fact no differences between models within each regime. Unfortunately, this is a non-standard testing problem, because the switching probabilities p_{12} and p_{21} are unidentified under the null. Accordingly, regularity conditions for validity of likelihood ratio tests are not met, and the likelihood ratio test statistic is not χ^2 distributed as in the standard case. Therefore standard tests are not applicable any more.[10] We pursue the approach in Hansen (1992), a generalization of the methodology in Davies (1987), to solve the testing problem when unidentified nuisance parameters are present only under the alternative hypothesis.

To illustrate, μ_2 and σ_2^2 are re-parameterized as $\mu_2 = \mu_1 + \mu_d$ and $\sigma_2^2 = \sigma_1^2 + \sigma_d^2$, with $\tau = (\mu_d, \sigma_d^2)'$. The hypothesis takes the form

$$H_0 : \tau = 0 \qquad H_0 : \tau \neq 0$$

with the null being no regime switching, i.e. no difference between regimes. Let γ denote those fully identified parameters $\gamma = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \mu_1, \sigma_1^2)'$, and $\varrho = (p_{12}, p_{21})'$. The test statistic is viewed as a function of the nuisance parameter. According to Hansen (1992), the test requires computing the maximum likelihood estimates of γ for fixed values of $\upsilon = (\tau, \varrho)'$. In particular, at each fixed value of υ, the constrained maximum likelihood estimates of γ, $\hat{\gamma}(\upsilon) = \max\ L_T(\upsilon, \gamma)$, are obtained. Define the standardized likelihood ratio function as $\hat{LR}_T^*(\upsilon) = \hat{LR}_T(\upsilon)/V_T(\upsilon)^{1/2}$, with

$$\hat{LR}_T(\upsilon) \quad = \quad \hat{L}_T\big(\upsilon, \hat{\gamma}(\upsilon)\big) - \hat{L}_T\big(0, \varrho, \hat{\gamma}(0, \varrho)\big)$$

[10] Hansen (1992) and Hamilton (1996, 2005) discuss the testing problem in details. According to Hansen (1992), this identification problem also partly explains the difficulty in numerical optimization.

$$V_T(\upsilon) \quad = \quad \sum_{1}^{T} \left(l_t\big(\upsilon, \hat{\gamma}(\upsilon)\big) - l_t\big(0, \varrho, \hat{\gamma}(0, \varrho)\big) - \frac{1}{T} \hat{LR}_T(\upsilon) \right)^2$$

The maximum of $\hat{LR}^*_T(\upsilon)$ yields the standardized likelihood ratio test statistic

$$\hat{LR}^*_T = \sup \hat{LR}^*_T(\upsilon)$$

the asymptotic distribution of which is obtained by simulation as suggest by Hansen (1992), who also provides further details of simulation and theoretical justifications for the test.

5.2.3 Multi-factor Stochastic Volatility Model

As discussed in Section 5.1, long range dependence can be also captured by mixing two or more short-memory processes. This section describes the methodology we use to model realized volatility under this framework. Our main reference is Barndorff-Nielsen and Shephard (2002*a*), the contributions of which include adjusting for the measurement error in realized volatility, and capturing the long-range dependence by a sum of stochastic volatility (SV) processes. Section 5.2.3.1 focuses on the first contribution, with the second contribution being the topic of Section 5.2.3.2.

5.2.3.1 Single-factor Specification

Barndorff-Nielsen and Shephard (2002*a*) is set up under continuous-time SV models. The consistence of realized volatility as a volatility estimator is based on the assumption that the underlying volatility process evolves continuously over time; see Section 2.5 in Chapter 2 for a review of realized volatility. Follow the specification in Barndorff-Nielsen and Shephard (2002*a*, p. 253), the log-price $p^*(t)$ is the solution to the stochastic differential equation (SDE)

$$\mathrm{d}p^*(t) = \{\mu + \beta\sigma^2(t)\}\mathrm{d}t + \sigma(t)\mathrm{d}w(t)$$

where $\sigma^2(t)$ is the *instantaneous* or *spot volatility*, μ is drift term and β is risk premium. The spot volatility is assumed to be stationary and independent with the standard Brow-

nian motion $w(t)$. For asset returns over an interval of $\Delta > 0$, denoted as r_t, we have

$$r_t|\sigma_t^2 \sim N(\mu\Delta + \beta\sigma_t^2, \sigma_t^2).$$

which is in line with the Mixture-of-Distributions hypothesis reviewed in Section 2.2 of Chapter 2. The term σ_t^2 is called *actual volatility*, which is the discretized *integrated volatility* over an interval of interests, say daily. Realized volatility is a consistent estimator of σ_t^2, but not free of measurement error. In other words, as argued by Barndorff-Nielsen and Shephard (2002*a*), treating the observed daily realized volatility equivalently as σ_t^2 might face the *errors-in-variables* problem. Accordingly, they propose to decompose the observed daily realized volatility $\{y\}_t$ into actual volatility and volatility error,

$$\{y\}_t = \sigma_t^2 + u_t \tag{5.10}$$

where t is of daily frequency, and u_t is *realized volatility error*, which is a white noise sequence and uncorrelated with σ_t^2. The variance of u_t, the magnitude of realized volatility error, is determined by the parameters specifying the underlying continuous SV process; see Barndorff-Nielsen and Shephard (2002*a*) for the derivations of the asymptotic distribution of u_t.

To illustrate, suppose the underlying SV process is the constant elasticity of variance (CEV) process which is the solution to the SDE

$$\mathrm{d}\sigma^2(t) = -\lambda\{\sigma^2(t) - \xi\}\mathrm{d}t + \omega\sigma(t)^{\zeta}\mathrm{d}b(\lambda t), \qquad \zeta \in [1, 2] \tag{5.11}$$

where $b(t)$ is standard Brownian motion and uncorrelated with $w(t)$. The parameters specifying the process include $(\xi, \lambda, \omega^2)'$. Given this specification, the autocorrelation function of the volatility process is determined by λ. Therefore, λ captures the persistence in the volatility process. Barndorff-Nielsen and Shephard (2002*a*) refer to λ as the memory parameter (Barndorff-Nielsen and Shephard (2002*a*, p. 261)). Its magnitude is not directly comparable to the long memory parameter d discussed in Section 5.2.1. According to Barndorff-Nielsen and Shephard (2002*a*, equation 9), the variance of realized volatility

error is

$$var(u_t) = 2M\Big[\frac{2w^2}{\lambda^2}\Big(\exp(-\lambda M^{-1}) - 1 + \lambda M^{-1}\Big) + \frac{\xi^2}{M^2}\Big] \tag{5.12}$$

with M being the number of intra-day returns used to constructed realized volatility. For example, if daily realized volatility is constructed by five-minute returns, then $M = 288$.

Under a rather general specification of the continuous-time SV model, the autocovariance function of σ_t^2 in (5.10) is shown to be of the same form as that of an ARMA(1,1) process. For the specification in (5.11), based on the characteristics of autocovariance functions of an ARMA(1,1) process, it can be shown that the autoregressive coefficient Φ is a function of $\exp(-\lambda)$. The moving average coefficient Θ is the solution to the following equation and to be determined numerically,

$$\frac{2(\Phi - 1 + \lambda)}{(1-\Phi)^2} - \frac{1 - \Phi^2 + (\Theta + \Phi)^2}{(1-\Phi^2)(\Theta + \Phi) + \Phi(\Phi + \Theta)^2} = 0, \quad \text{with } \Phi = \exp(-\lambda) \tag{5.13}$$

Putting all these together, realized volatility $\{y\}_t$ can treated as an ARMA(1,1) process with additive noise according to (5.10). This treatment naturally leads to the general framework of state space form: see Brockwell and Davis (1991) for the state space form of an ARMA process with additive noise. The variance of the additive noise is given by (5.12); the autoregressive and moving average roots to the ARMA(1,1) process are given in (5.13); parameters to be estimated are $(\xi, \lambda, \omega^2)'$. The Kalman filter delivers the best linear unbiased estimator of actual volatility given asymptotic normality of realized volatility error (Barndorff-Nielsen and Shephard (2002*a*, section 3)). It is a routine exercise in state space modeling to conduct state filtering and smoothing in order to estimate the actual volatility, conditional on the quasi MLE of the parameters. Section 4.2.3 in Chapter 4 discusses the algorithms for state filtering and likelihood evaluation of state space models.

5.2.3.2 Multi-factor Specification

The single-factor specification addresses measurement error in realized volatility. In order to capture the long-range dependence in actual volatility, the single-factor speci-

fication can be extended to multi-factor SV models, by assuming $\sigma^2(t)$ is a sum of Z independent CEV processes. Each CEV process is of the form as in (5.11), and has individual parameter λ_z for $z = 1, \ldots, Z$, which controls the autocorrelation function of each factor. Accordingly, the autocorrelation function of spot volatility can have components which are a mix of quickly and slowly decaying components. This is desirable as it can generate long-range dependence. Barndorff-Nielsen and Shephard (2002*a*, section 2.3) explain how the mixing can generate long-range dependence.

To illustrate, suppose $Z = 2$, the parameters to be estimated are $(\xi, \lambda_1, \lambda_2, \omega^2, \varpi_1, \varpi_2)'$, where ϖ_z is the weight of each component and $\varpi_1 + \varpi_2 = 1$ to ensure identification of parameters. Equations (5.12) and (5.13) are, accordingly, modified as

$$var(u_t) = 2M\Big[2w^2 \sum_{z=1}^{Z} \frac{\varpi_z}{\lambda_z^2}\Big(\exp(-\lambda_z M^{-1}) - 1 + \lambda_z M^{-1}\Big) + \frac{\xi^2}{M^2}\Big] \tag{5.14}$$

and

$$\Phi = \frac{\sum_{z=1}^{Z} \varpi_m \Big[1 - \exp(-\lambda_z)\Big]^2 \exp(-\lambda_z)\lambda_z^{-2}}{\sum_{z=1}^{Z} \varpi_z \Big[1 - \exp(-\lambda_z)\Big]^2 \lambda_z^{-2}} \tag{5.15}$$

$$\frac{\sum_{z=1}^{Z} 2\varpi_z \Big[\exp(-\lambda_z) - 1 + \lambda_z\Big]\lambda_z^{-2}}{\sum_{z=1}^{Z} \varpi_z \Big[1 - \exp(-\lambda_z)\Big]^2 \lambda_z^{-2}} - \frac{1 - \Phi^2 + (\Theta + \Phi)^2}{(1 - \Phi^2)(\Theta + \Phi) + \Phi(\Phi + \Theta)^2} = 0 \tag{5.16}$$

with the moving average root Θ, solution to the last equation, determined numerically. Expressions in (5.14)-(5.16) make use of the result that the autocovariance function of a sum of independent components is the sum of autocovariances of the terms in the sum.

MLE is used to estimate the parameters $(\xi, \lambda_1, \lambda_2, \ldots, \lambda_Z, \omega^2, \varpi_1, \varpi_2, \ldots, \varpi_{Z-1})'$. Volatility forecasts are obtained conditional on the point parameter estimates. Actual volatility σ_t^2 is treated as a state variable, and the algorithm for forecasting the state variable is delivered by Kalman filter.

5.3 Empirical Analysis

In this section, we conduct an empirical analysis of the Yen/USD daily realized volatility over a decade. We focus on the long-range dependence property of the observed time series, and try to address the following question: given that the three methods considered are all able to mimic slowly decaying autocorrelations theoretically, then which one is able to provide a better description of the data from a practical point of view. Our main object is not to determine what the true underlying mechanism (if it exists) is, but to examine and compare the performances of the three alternatives in terms of in-sample fit and out-of sample forecast. From a theoretical point of view, it is not easy to answer a question about whether it is a Markov switching process that approximates fractional integrated process, or the other way around. The following sections detail how we proceed.

5.3.1 Preliminary Data Analysis

The construction of realized volatility is first described, followed by discussion on the unconditional distributional characteristics of the series to be analyzed. Hypothesis testing of short memory against long memory process, and testing about the presence of multiple regimes, is also conducted.

5.3.1.1 Realized Volatility Construction

The realized volatility used in this chapter is constructed by the raw tick-by-tick *over-the-counter* (OTC) quotes of the Yen/USD that appeared on Reuter's FXFX page. The sample examined covers the period from January 7, 1996 to August 6, 2009. Each quote consists of *bid* and *ask* prices and the times at which they are recorded to the nearest millisecond. Raw data is not free of erroneous recordings, or errors due to other sources, such as technical fault of the electronic recording system. The most reliable way to remove these errors might be through visual inspection with each dubious quote to be examined carefully, however, this is almost infeasible as the number of ticks to be processed is overwhelming once intra-day data is dealt with. Therefore, a formal approach,

algorithms which can be implemented automatically by software, is desirable. We apply the filtering algorithm proposed by Dacorogna et al. (1993) to mitigate the effects of errors. The filtering of raw intra-day financial data is a growing research area; see Dacorogna et al. (2001) for a book-length treatment.

In this chapter, daily realized volatility is constructed from five-minute returns with returns defined as differences in logarithm prices. The knowledge of asset prices sampled at five-minute intervals is required. We follow the convention by defining logarithm price as the midpoint of the logarithm of *ask* and *bid* prices. The regularized price is taken as the last available price from the interval. One possible limitation here is that prices are not exactly regularly sampled. However, we argue that the effect of irregularity might be minor given the high liquidity of the Yen/USD with the average time lapse between two consecutive quotes being measured in seconds. The alternative approach, constructing price for each five-minute interval by linearly interpolating from the last price from this interval and the first price from next interval, could introduce spurious predictability. The difference between the two approaches is negligible if quotes are available every a few seconds, which is the case in our example.

Daily realized volatility is the sum of squared intra-day returns. There is no clear consensus on what is the optimal sampling frequency (see Section 2.5.2 in Chapter 2). The consistency of realized volatility as an estimate of actual volatility relies on the assumption that intra-day returns can be sampled at very fine intervals. However, it is more common to use five-minute intervals for a number of reasons. First, some assets are not liquid enough to be sampled or observed frequently even if the underlying price might evolve continuously. Second, market microstructure noise introduces extra source of dynamics precluding the use of intra-day returns at very fine intervals even if the trading might be very thick. The market for the Yen/USD is fairly active, for example, there are 82 399 quotes of *bid* and *ask* prices available on a single day of October 6, 2008. So a tradeoff is necessary to determine a sampling frequency, which makes the constructed realized volatility robust to microstructure noise and, at the same time, returns are sampled as fine as possible. A variety of frequencies have been tried in the literature, and the most popular choice is 5 minute. In the work of Andersen, Bollerslev, Diebold and Labys (2001);

Andersen et al. (2003), a 30 minute frequency is also tried. We tried both options, and do not observe an obvious disparity between the two resultant times series. The first option is chosen in this chapter.[11]

Foreign exchange is traded 24 hours per day world-wide. A common observation in trading patterns is that trading activity in terms of volume and number of quotes diminish significantly during weekends and holidays. On these thin trading days, most of the five-minute returns are close to zero, as a result, their sum of squares is of very small magnitude. This will cause numerical problem especially when the logarithm of realized volatility is modeled.[12] To eliminate weekend and holiday effects, the period from 21:00 GMT Friday to 21:00 GMT Sunday, and holidays (including fixed and moving holidays)[13] are excluded from our sample. We also remove those days with data holes manifesting as consecutive zero five-minute returns.[14] These leave us with total 3259 daily realized volatility, with the first 2900 observations treated as training set and the period from March 4, 2008 to August 6, 2009 as rolling sample for recursive estimations and evaluates forecasting performance. Sample size is important in examining the long-memory property of time series, since it is the behavior of autocorrelations at long lags that matters. It is difficult to capture this slowly-decaying in ACF plot with small samples.

[11]The filtering method proposed by Corsi et al. (2001) is promising. It might be interesting to see whether the results obtained here is robust to the way realized volatility being constructed. Another method available in the literature is to use *volatility signature plot* to determine the optimal frequency (Andersen, Bollerslev, Diebold and Labys (2001) and Andersen et al. (2003)).

[12]This is similar to the *inlier* problem of discrete-time SV model: see Breidt and Carriquiry (1996) and Shephard (1996). SV models make use of log squared returns, so an almost zero return will cause numerical problem in estimation. SV models are commonly used on daily or weekly frequency, and it is rare for the observed prices to be exactly the same from the beginning to the end of a trading day or a week.

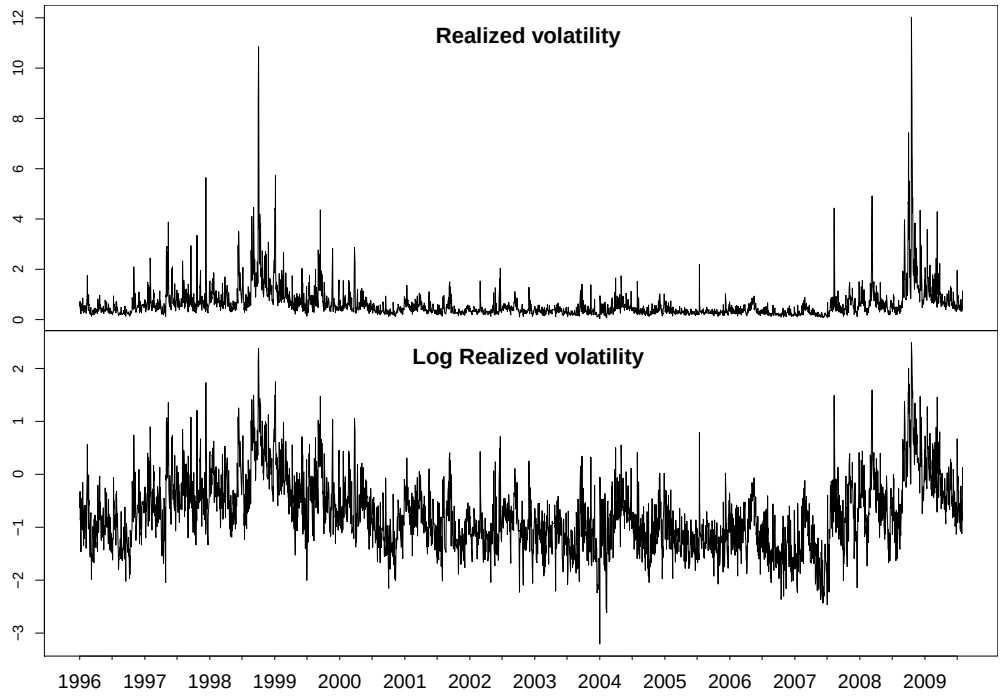
[13]These holidays include Christmas (December 21-26), New Year (December 31 - January 2), Good Friday, Easter Monday, Memorial Day, Independence Day, Labor Day, Thanksgiving and the day after Thanksgiving. We plot the raw data for each of these days, and make the decision to remove the corresponding day or not based on whether there is long consecutive missing quotes, i.e. whether the trading is too thin.

[14]By defining normal trading days as with data holes, we refer them as those normal trading days with the 10 largest numbers of consecutive missing quotes. The choice of 10 is clearly *ad hoc*, but there is no universally acceptable rule in the literature.

5.3.1.2 The Distributional Characteristics of Realized Volatility

The time sequence plot for realized volatility is given in Figure 5.1. Realized volatility appears to be particularly volatile during the Asian financial crisis from 1997 to 1998, and the global financial crisis beginning in 2008. There are two extreme spikes presented in the plot, the first is caused by the Russian financial crisis[15] and the other is due to the 2009 Global Financial Crisis. The spikes appear to be much smaller on the logarithm scale, where a few large negative values are present due to the logarithmic transformation.

Figure 5.1: Daily realized volatility for the Yen/USD (07/01/1996 -06/08/2009)

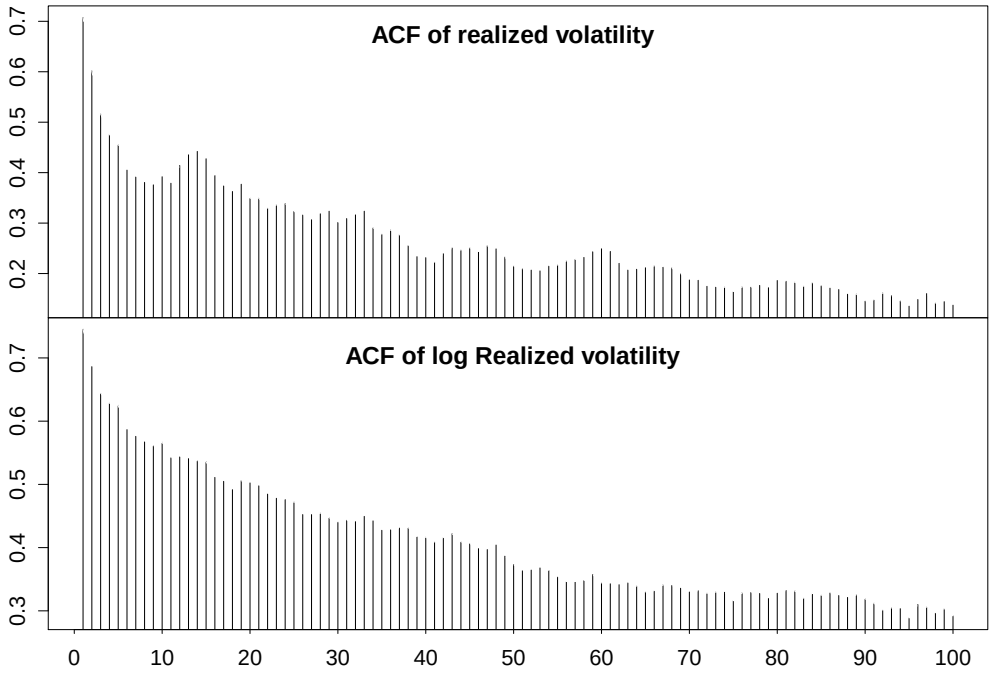


The ACF plot (Figure 5.2) implies substantial positive autocorrelations in realized volatility. Autocorrelations decay to zero at a slow rate, with the first 100 autocorrelations all being positive. This is as expected, since realized volatility is considered to be a fairly accurate estimate of actual volatility, and the literature in both GARCH and stochastic volatility models has well documented the high persistence in volatility based on daily returns. The autocorrelation function of logarithm realized volatility shows higher auto-

[15]See Maekawa and Xinhong (2009) for analysis of the Russian financial crisis. Hedge funds liquidated their open positions of the Yen on October 1998 resulting in sudden and sharp rises in the exchange rate for Yen (Dollar and Euro depreciation). The crisis covers three days of October 7-9, 1998 and the months from September to October being particularly volatile.

correlation and a smoother pattern compared with that of realized volatility itself. This might be explained by the presence of positive spikes since outliers will reduce autocorrelations in observed time series. The first autocorrelation of realized volatility is around 0.7, largely consistent with the values reported in the literature (normally around 0.60-0.65). Accordingly, autocorrelation between consecutive actual volatilities might be even higher because noise can mask persistence and realized volatility is still not free of measurement error. Besides the slowly-decaying pattern, local peaks can be observed at autocorrelations of seasonal lags, such as 5, 10, 15 and 20, clearly indicating a minor calendar effect and variation in volatility across the five days of the week.[16]

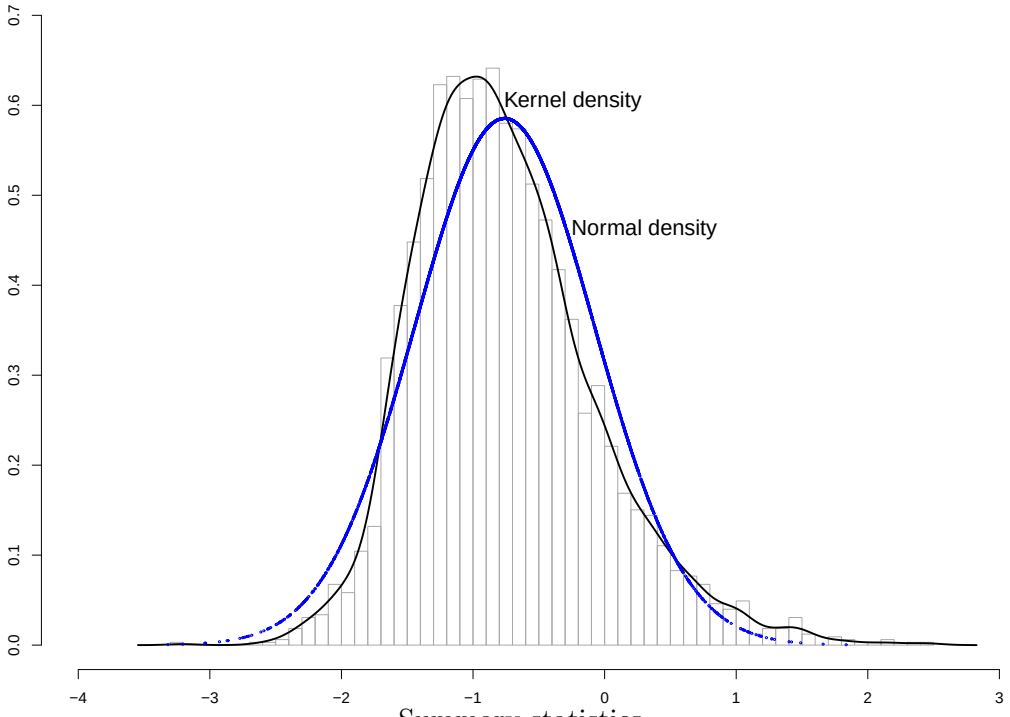
Figure 5.2: ACF of daily realized volatility for the Yen/USD (07/01/1996 -06/08/2009)



The density plot and summary statistics of log realized volatility $\log(RV)$ are provided in Figure 5.3. Taking the log of RV produces a distribution that is approximately normal. However, disparity still exists with the values of skewness and kurtosis indicating a slightly skewed and leptokurtic distribution of $\log(RV)$, which is also indicated by the difference

[16]This is consistent with the observation in Taylor and Xu (1997), where they show that average realized volatility increases from Monday to Friday, presumably reflecting the timing of macroeconomic announcements.

Figure 5.3: Kernel density and summary statistics of logarithm realized volatility



Summary statistics

	Mean	Median	Max	Min	Std.Dev	Skewness	Kurtosis	Jarque-Bera[a]
RV	0.619	0.429	12.023	0.040	0.686	6.150	66.95	575919.5
$\log(RV)$	-0.760	-0.847	2.487	-3.207	0.682	0.717	4.018	420.0

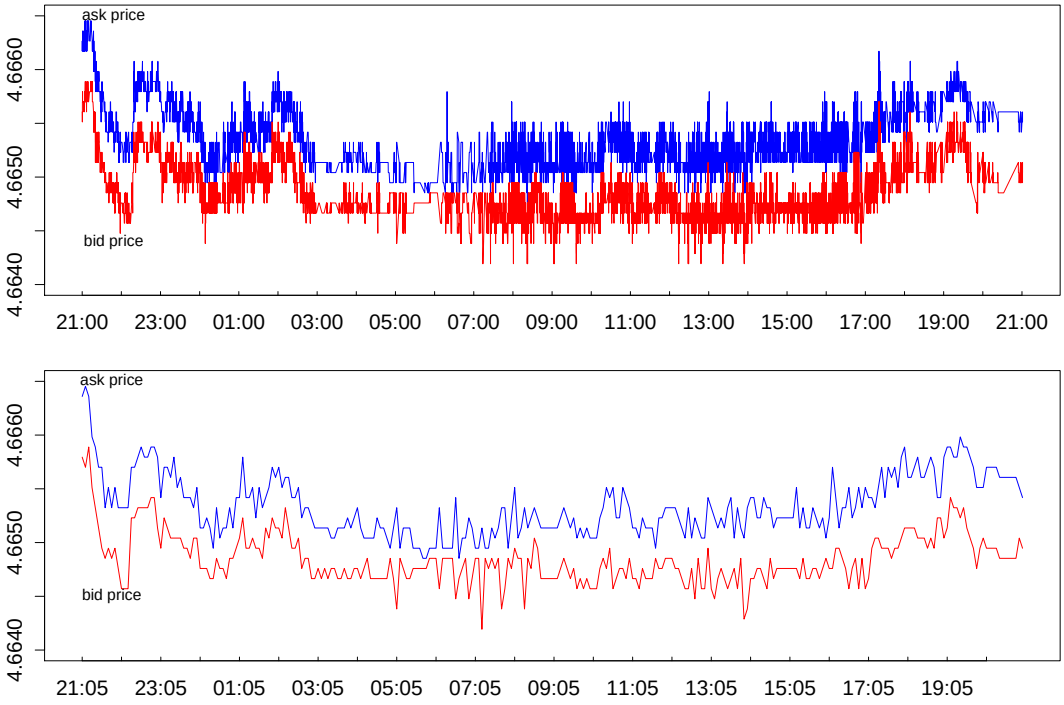
[a] JarqueBera is a goodness-of-fit measure of departure from normality.

between the estimated kernel density and normal density plots. This casts some doubt on the log normality assumption of volatility, which might affect the statistical inference on ARFIMA parameter estimates when MLE is conducted on $\log(RV)$.

Minimum Realized Volatility The minimum daily realized volatility of our sample is 0.04, which is the estimate of actual volatility over a 24-hour period beginning from GMT 21:00 on January 7, 2004. This quantity is fairly close to zero, yielding a very large negative value for $\log(RV)$. So we plot the raw data for a close examination. The plot in the top panel of Figure 5.4 is the filtered *ask* and *bid* prices (in logarithm) observed over that 24 hour period, made up of 9 907 quotes. The bottom panel plots the prices observed at five-minute intervals. Overall, the trading activity appears tranquil during the 24 hours, implying price movements of fairly small magnitudes. A visual inspection of the tick-by-tick prices indicates little evidence of abnormality, such as holes in data

feed revealed by constant prices over several hours. As a result, the corresponding daily realized volatility is not treated as an anomaly and is kept in our sample.

Figure 5.4: Yen/USD log prices over 24-hour period from 21:00 07/01/2004 (GMT)

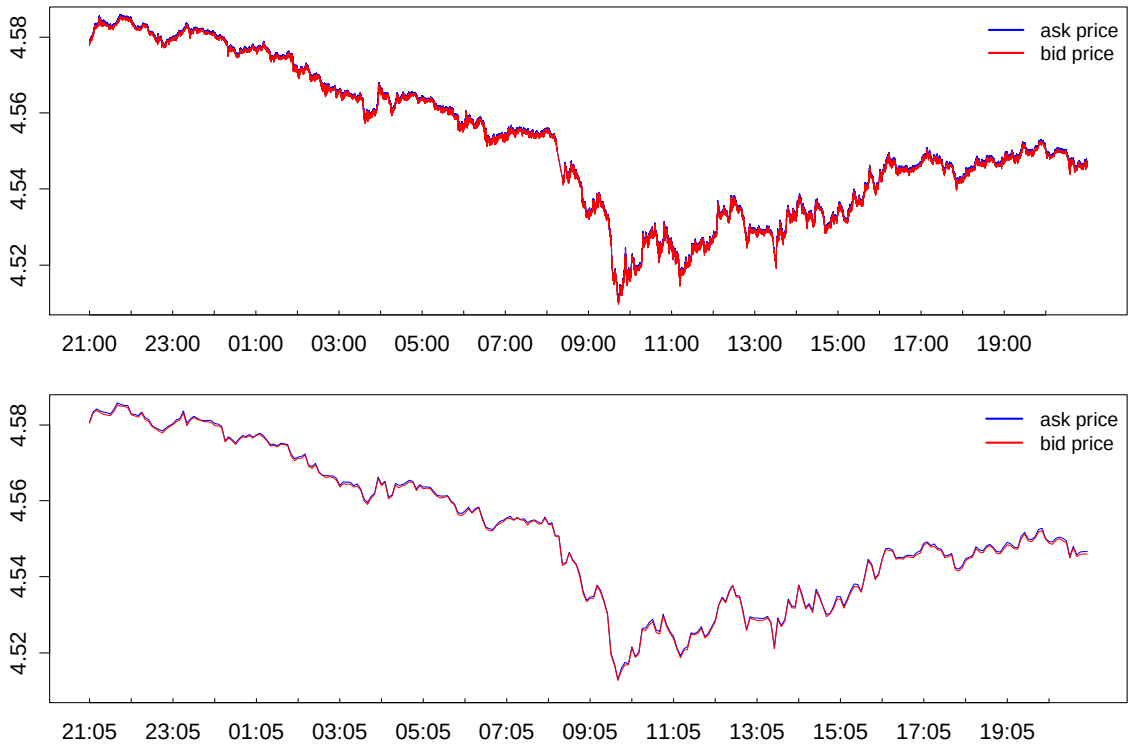


The top panel is the tick-by-tick *ask* and *bid* prices. The bottom panel is the every five-minute *ask* and *bid* prices. All the prices are in logarithm.

Maximum Realized Volatility At the other extreme, the maximum volatility is observed on a day early in the financial crisis which began in 2008. The corresponding tick-by-tick prices, including 72 915 ticks, and regularized prices are provided in Figure 5.5. This was an active trading day, with a much larger price than that when the minimum realized volatility was observed (Figure 5.4). Compared with the magnitude of prices, the bid-ask spread is very small, as the lines for ask prices and bid prices are fairly close in the plot. Overall little evidence of abnormality is present, hence the corresponding realized volatility is kept in our sample.

We discuss these two extreme cases, because it is important to combine the filter algorithm (Section 5.3.1.1) and visual inspection when realized volatility is constructed. It is well recognized in the literature that intra-day data might be contaminated by errors.

Figure 5.5: Yen/USD log prices over 24-hour period from 21:00 23/10/2008 (GMT)



The top panel is the tick-by-tick *ask* and *bid* prices. The bottom panel is the every five-minute *ask* and *bid* prices. All the prices are in logarithm.

To design a universally accepted filtering rule (algorithm) so that millions of raw asset prices can be processed automatically prior to constructing realized volatility is difficult in practice. The quality of the constructed realized volatility is expected to play a role in the statistical inference of latent volatility as the former is treated as a consistent estimator of the latter. When realized volatility is modeled on its original scale, these influential observations are referred as the spikes in the top panel in Figure 5.1. A large magnitude in realized volatility indicates a high degree of price movements. If the filter was too strong, the resultant asset price path might be smoother, which reduces the magnitudes of realized volatility by washing out part of the dynamics in price movements. On the other hand, a very weak filter will fail to detect erroneous recordings. To determine a suitable tradeoff is not easy. When volatility is modeled on logarithmic scale, a volatility fairly close to zero will transform into extreme negative values. A small observed realized volatility is an indication of low market activity, but it is also likely to be caused by missing quotes due to technical faults. Therefore visual inspection is often needed. Nevertheless, visual

inspection of tick-by-tick prices on a daily basis could be time-consuming if a long period of time series is examined.

Long Memory Testing As part of the preliminary analysis, modified rescaled range tests proposed in Lo (1991) and Giraitis et al. (2003) are conducted on RV and $\log(RV)$ respectively. The two tests both have a null hypothesis of a short memory process, with an alternative hypothesis of a stationary long memory ARFIMA(p,d,q) process. Test results are tabulated in Table 5.1. The test of Giraitis et al. (2003) is more robust to model misspecification such as departure from linearity. Overall, the two tests results are consistent: the null is clearly rejected especially in the test of Giraitis et al. (2003).

Table 5.1: Modified rescaled range tests of long memory for RV and $\log(RV)$

	RV		$\log(RV)$	
	Lo Test	Giraitis Test	Lo Test	Giraitis Test
K_1	1.942^{**}	0.332^{**}	2.288^{**}	0.453^{**}
K_2	1.571	0.217^{**}	1.764^{**}	0.270^{**}
K_3	1.398	0.172^{*}	1.530	0.203^{**}

[a] **: significance at 5%. *: significance at 10%.

[b] The test statistics will differ when different values of number of lags K included for calculation of covariances are chosen, we follow Lo (1991) and try the following values $K_1 = 90$, $K_2 = 180$ and $K_3 = 270$. The optimum choice of K is not available, the situation is similar as the choice of bandwidth when non-parametric estimation of memory parameter d is conducted. So different values of K should be experimented with.

[c] The critical values of Lo (1991) and Giraitis et al. (2003) tests are: 1.747 and 0.187 at 5% significance level, and 1.620 and 0.152 at 10% significance level, respectively.

Regime Identification Testing We conduct the test to see whether the data suggest the existence of more than one regime, prior to estimating two regimes, following the approach discussed in Section 5.2.2.4. The order of autoregressive and moving average terms in the regime switching model (5.3) are both chosen as 1. The choice of ARMA(1,1) is mainly for reasons of parsimony; there is no reason to preclude other settings. The grids for the range of values each parameter takes in computing the test statistic are

Grid for transition probabilities p_{11} and p_{11} : 0.1 to 0.8 in steps of 0.1 (eight gridpoints)

Grid for regime mean difference μ_d : 0.1 to 2.1 in steps of 0.4 (six gridpoints)

Grid for regime variance difference σ_d^2 : 0.05 to 0.4 in steps of 0.05 (eight gridpoints)

corresponding to a partition of the space for $\upsilon = (p_{11}, p_{22}, \mu_d, \sigma_d^2)'$ into 3072 grid points. Monte carlo simulation is used to obtain the critical value of $\hat{LR}_T^*(\upsilon)$, and the critical values are 2.777, 3.098 and 3.599 at the significance levels of 10%, 5% and 1% respectively. The standardized LR test statistics $\hat{LR}_T^*(\upsilon)$ is 8.598, being significant at 1% level. Hence, the assumption of one regime is clearly rejected.

According to these test results, there is evidence in favor of both AFIMA and regime switching models when compared to a stationary null model. Recursive estimations are conducted in the next section to see which specification is able to provide better fit and forecasting.

5.3.2 Recursive Estimations

Recursive estimations are conducted in this section in the following fashion: the first observation, which is on January 7, 1996, is fixed, and the sample is updated with each new observation becoming available. The rolling period examined has end dates ranging from March 4, 2008 to August 6, 2009, resulting in 360 recursive estimations for each method. Summaries of estimation results are discussed first, followed by comparisons of in-sample fit.

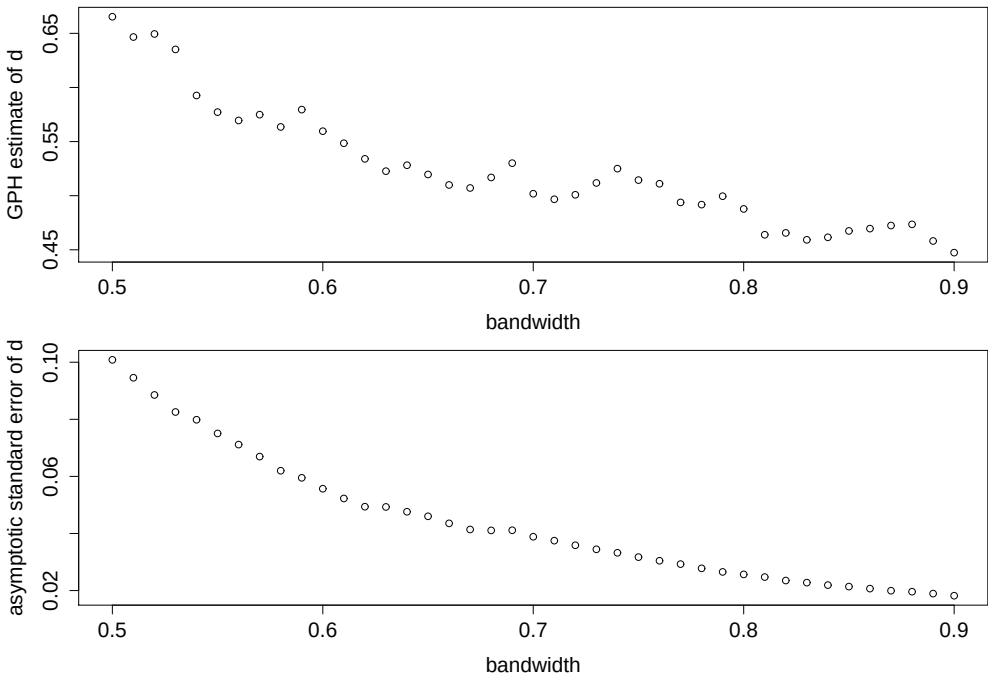
5.3.2.1 Estimation Results for ARFIMA Models

ARFIMA models (5.1) are estimated on $\log(RV)$, where the choice of orders of p and q is important because memory parameter estimates could be misleading if short-term dynamics are mis-specified. A range of specifications of p and q are tried, however, the discussion that follows is restricted to ARFIMA(1,d,1), ARFIMA(5,d,0), ARFIMA(3,d,3) and ARFIMA(0,d,0). The four specifications are chosen based on the following logic: ARFIMA(1,d,1) is a parsimonious specification, and the choice of ARFIMA(5,d,0) is largely due to seasonality with five day periodicity implied in the ACF plot. The other two specifications represent the cases of possible over-parameterization and a simple assumption of little short-term dynamics in volatility, respectively. In additional models not

reported here we vary the orders of p and q in the range of (0,1,2,3).[17] The general comments on these estimation results are: point estimates of memory parameter are sensitive to the orders of p and q, but in terms of analysis of residuals and out-of sample forecasting, no significant differences are observed.

Figure 5.6 plots the GPH estimate of the memory parameter d when modeling $\log(RV)$ for the full sample period when different values of bandwidth are chosen, with point estimates and corresponding asymptotic standard errors in the top and bottom panels, respectively. It is clear that GPH estimator is sensitive to the choice of bandwidth. A common choice of bandwidth is the one which provides a relatively small standard error. Based on this rule, d is estimated around 0.45, in the range of typical values documented in the literature.

Figure 5.6: GPH estimates of memory parameter d for log volatility of the full sample



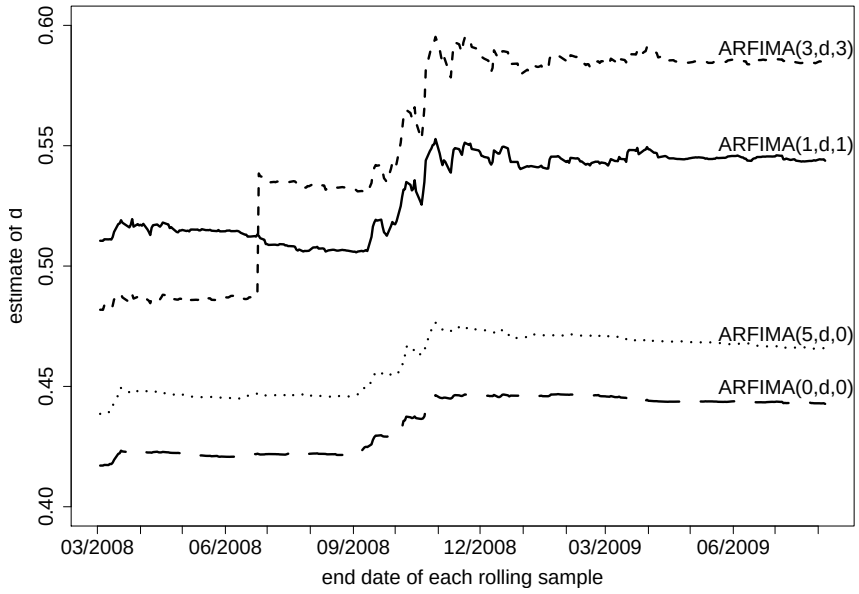
Point estimates of d are plotted in the top panel with corresponding asymptotic standard errors plotted in the bottom panel. Horizontal axis represents the values of bandwidth chosen for GPH estimators. The full sample is from 07/01/1996 to 06/08/2009.

Figure 5.7 shows the recursive estimates of d for the selected ARFIMA models; the estimates do vary with the specification of short-term dynamics. Point estimates of d for

[17]The estimation results for this wide range of values of p and q are available on request.

the ARFIMA(3,d,3) and ARFIMA(1,d,1) models are out of the stationary region ($d > 0.5$), especially for samples ending after late 2008, when the financial turmoil began. The degree of persistence in volatility is pushed up since the crisis, with all the d estimates increasing ever since. The ARFIMA(5,d,0) and ARFIMA(0,d,0) specifications provide similar d estimates to that of the GPH estimator when the bandwidth is around 0.8.

Figure 5.7: Recursive estimates of d for selected ARFIMA(p,d,q) models



Maximum likelihood estimate is on the vertical axis. The horizontal axis represents the end date of each rolling sample.

5.3.2.2 Estimation Results for Regime Switching Models

Two types of regime switching models are estimated: short memory regime switching (equation 5.3) and long memory regime switching (equation 5.9) models. We are primarily concerned with the possibility of using short memory regime switching to mimic the long memory property of time series. So the discussion in this section is on short memory regime switching, with long memory regime switching to be discussed later when we evaluate out-of-sample forecasting.

In terms of the short memory regime-switching models, two specifications of (5.3) are tried on $\log(RV)$: ARMA(1,1) and ARMA(5,1), with both allowing the level and variance of innovations to switch across the 2 regimes. For the specification of regime switching

ARMA(5,1), autoregressive coefficients at lags 2,3,4 are dropped after they are found to be insignificant. The inclusion of lag 5 is to capture the mild seasonality observed in Figure 5.2. Of course, there is no reason to preclude a specification which allows for autoregressive and moving-average coefficients to switch between regimes. We did estimate such models, but the results were inferior to the specifications reported here. Estimation of regime switching models can be sensitive to initialization as the model is highly nonlinear, so a range of initial values are tried and the estimates which provide the maximum likelihood are chosen.[18]

Figure 5.8 plots recursive estimates of the regime switching ARMA(1,1) model, where μ_i and σ_i denote mean and standard deviation for each regime with $i = 1, 2$, and ϕ_1 and θ_1 denote the autoregressive and moving-average coefficients at lag 1. The first regime is characterized by a negative level and relatively small magnitude of innovation, which represents the tranquil state. The second regime is characterized by a positive level and large magnitude of innovation, representing a relatively volatile state. Autoregressive roots are outside the unit circle, indicating local stationarity of each regime.[19] Depending on the sample, the probability of a tranquil state being followed by another tranquil state is around $\hat{p}_{11} = 0.92$, so the typical persistency of this episode is $1/(1 - \hat{p}_{11}) = 12.5$ days. The estimated probability of a volatile state being followed by another volatile state is around $\hat{p}_{22} = 0.65$, indicating average persistency of this episode is $1/(1 - \hat{p}_{22}) = 3$ days. According to the equations in Franses and van Dijk (2000, p. 82), the unconditional probability $Pr(s_t = i)$ for the process being in each regime is

$$\begin{aligned} \hat{Pr}(s_t = 1) &= \frac{1 - \hat{p}_{22}}{2 - \hat{p}_{11} - \hat{p}_{22}} = \frac{1 - 0.65}{2 - 0.92 - 0.65} = 0.81 \\ \hat{Pr}(s_t = 2) &= \frac{1 - \hat{p}_{11}}{2 - \hat{p}_{11} - \hat{p}_{22}} = \frac{1 - 0.92}{2 - 0.92 - 0.65} = 0.19 \end{aligned}$$

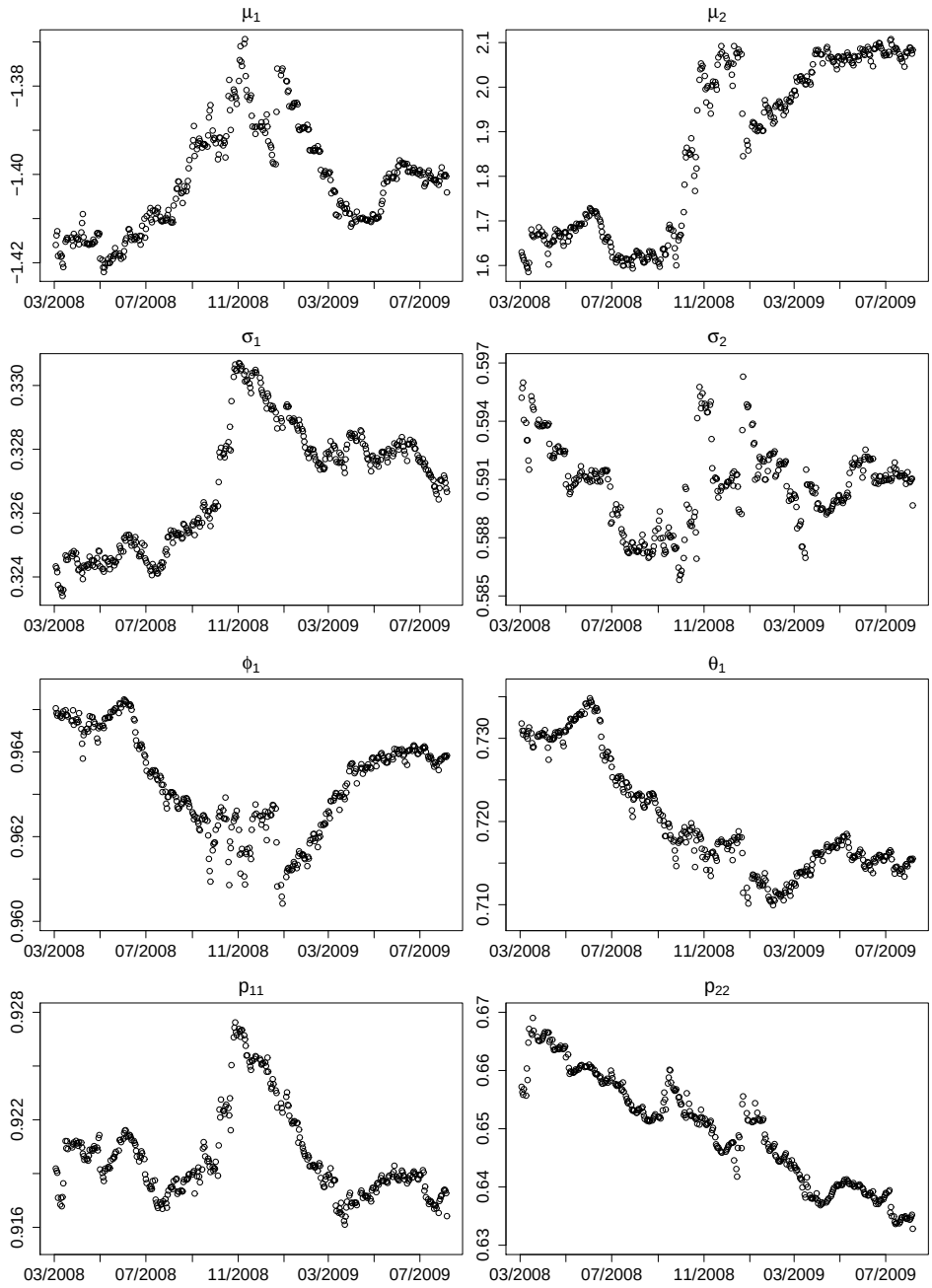
therefore most of the time the process is in the tranquil state. The mean of the tranquil state is estimated to be around -1.4, corresponding roughly to 0.25 in RV as the regime switching models are estimated on $\log(RV)$. The mean of the other state is estimated to

[18] The time series modeling package *Time Series Modeling 4.30* developed by James Davidson is used in this chapter to conduct the recursive estimation and forecasting of regime switching models.

[19] Local stationarity of each regime is not necessary condition for global stationarity: see the stationary conditions of regime switching models discussed by Francq and Zakoïan (2001).

be around 1.8, corresponding roughly to 6.05 in RV. Compared with 0.62, which is the unconditional mean of RV reported in Figure 5.3, the latter state is "noisy".

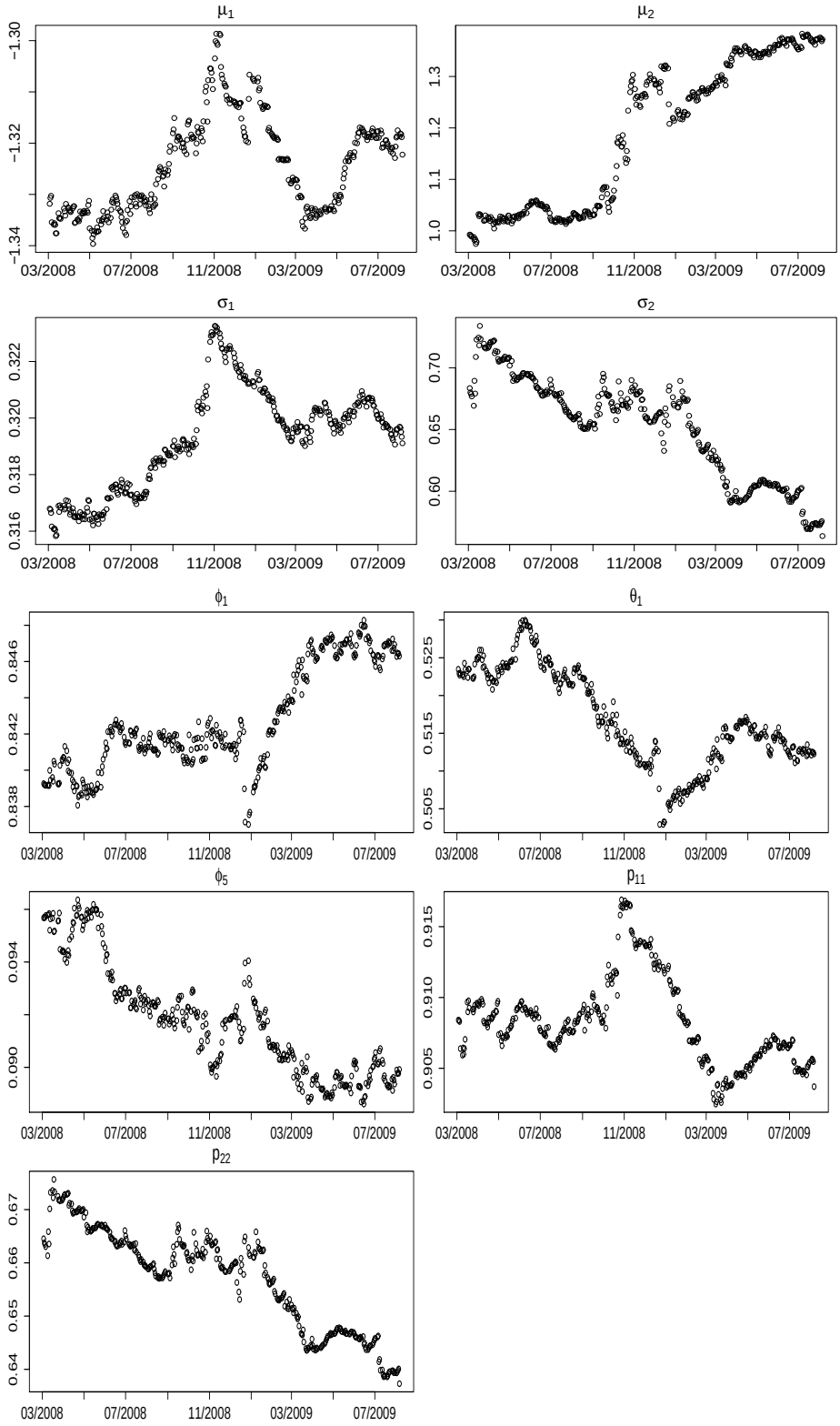
Figure 5.8: Recursive parameter estimates of regime switching ARMA(1,1) Model



Overall, the financial turmoil that began at the end of 2008 has impacted the stability of parameter estimates. A close examination of the magnitude of parameter estimates suggests instability is not very problematic in this case, since the range of the estimates of

each parameter is reasonably small except that of μ_2, the level of the second regime. The spikes due to the financial crisis have clearly pushed up the estimates of μ_2.

Figure 5.9: Recursive parameter estimates of regime switching ARMA(5,1) Model



Autoregressive roots at lag 2, 3, 4 are restricted to zero.

Figure 5.9 plots the recursive estimates for the regime switching ARMA(5,1) model. The parameter estimates are largely close to those in Figure 5.8. The autoregressive coefficient at lag 5 is around 0.1.

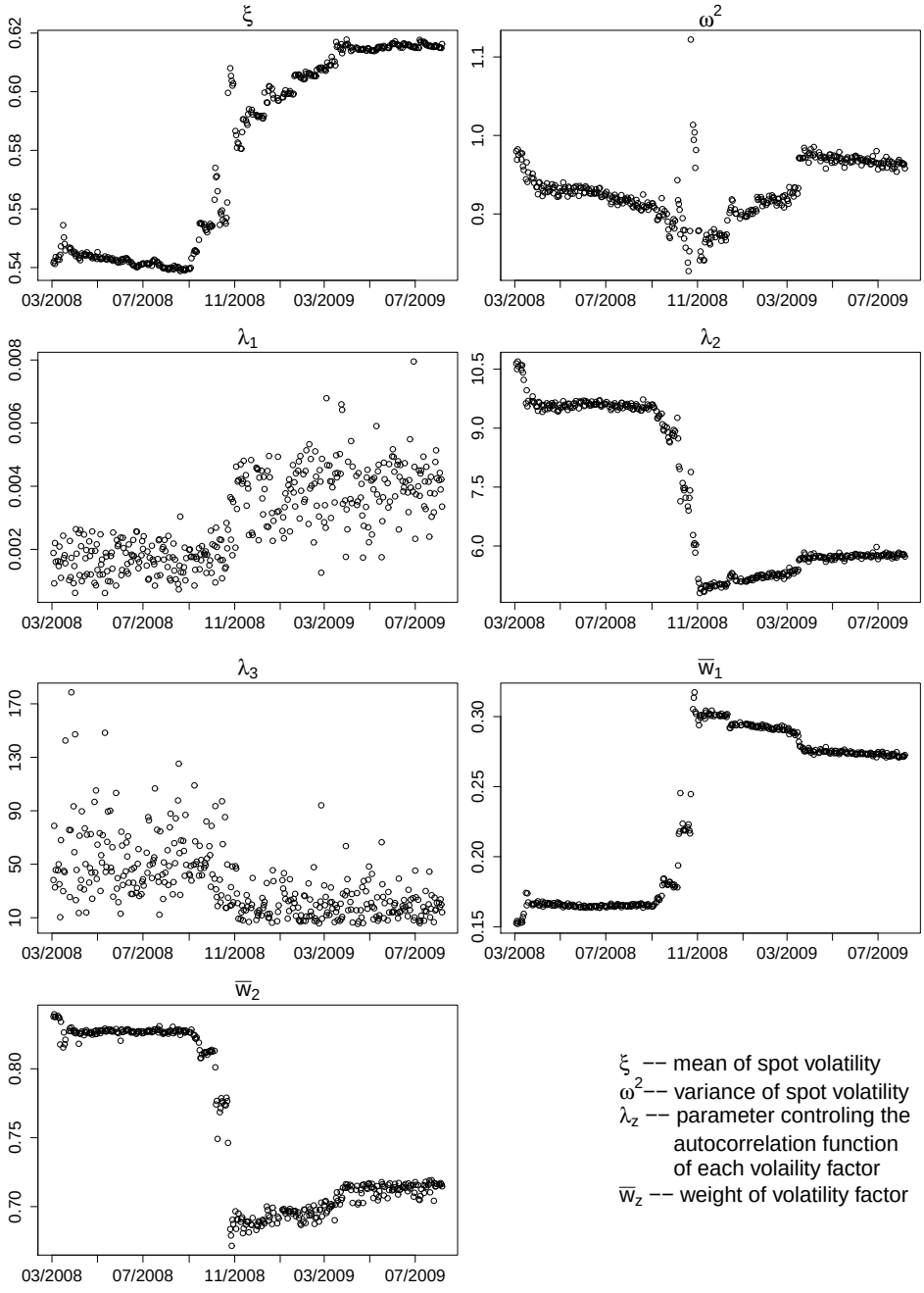
5.3.2.3 Estimation Results for a Multi-factor Stochastic Volatility Model

The three factor SV model is estimated according to the method discussed in Section 5.2.3.[20] Figure 5.10 plots the recursive estimates of all seven parameters, where λ_1, λ_2, and λ_3 denotes the parameter which determines the autocorrelation function of each volatility factor, with corresponding weights being ϖ_1, ϖ_2, and $\varpi_3 = 1 - \varpi_1 - \varpi_2$, respectively. Mean and variance of spot volatility are denoted as ξ and ω^2. Numerical estimation is not easy for this exercise, as revealed by the numerical optimizations being sensitive to initial values, taking a long time to converge and only weak convergence is available for some rolling samples. A seven-dimensional grid is used for initialization of the first rolling sample, with 800 sets of initial values being experimented with. The set of parameter estimates which give the largest value for the likelihood is chosen. The MLE of the first rolling sample is used to initialize the estimation of the second rolling sample, and so on.

Since the end of 2008, the parameters estimates show a clear pattern of instability. In general, the instability appears to be of larger degree than that the cases of ARFIMA and regime switching models (Figures 5.7, 5.8 and 5.9). For example, in Figure 5.8 the recursive parameter estimates of the regime switching ARMA(1,1) model do vary depending on samples, but the variation in general appears to be less. On the other hand, it appears that the estimation of λ_3 (Figure 5.10), which determines the autocorrelation function of the third volatility factor, is difficult, as there is a large variation of the recursive estimates. This might suggest the three factor SV model is an inferior description of the time series examined. In particular, it is less capable of incorporating the impact of the financial crisis than the regime switching and ARFIMA models.

[20] Alizadeh et al. (2002) documents a successful application of using the sum of short-memory processes to mimic the long-memory property of volatility. They suggest using two volatility factors, with one representing the persistence dynamics and the other representing the transient dynamics. They argue that the highly persistent factor can contribute to autocorrelations in volatility, and the less persistent one can do a good job in capturing volatility in volatility.

Figure 5.10: Parameter recursive estimates of a three-factor stochastic volatility model



The estimated values of ξ and ω^2, the mean and variance of the spot volatility are comparable to those reported in Barndorff-Nielsen and Shephard (2002*a*, table 3).[21] Pa-

[21] Barndorff-Nielsen and Shephard (2002*a*) used the US dollar-German Deutschmark series covering the 10-year period from 1 December 1986, until 30 November 1996. Every 5 min this series records the most recent quote to appear on the Reuters screen. The estimated parameters for a three-factor SV model based on realized volatility computed using 10-min returns are $\xi = 0.508$, $\omega^2 = 4.79$, $\lambda_1 = 0.0331$, $\lambda_2 = 0.973$, $\lambda_3 = 268$, $\varpi_1 = 0.0183$ and $\varpi_2 = 0.0180$. See their table 3 for the results with varying frequency used to construct realized volatility, and varying number of volatility factors.

rameter λ_z specifies the autocorrelation function of the zth volatility factor. According to equation (5.11), which assumes the underlying SV process is a constant elasticity of variance (CEV) process, the autoregressive coefficient of actual volatility is a function of $\exp(-\lambda)$. The estimated λ_1 implies a highly persistent volatility component with an autoregressive coefficient close to one. For the other two components, the estimated autoregressive coefficients are virtually zero implying two rapidly decaying components. In particular, $\exp(\hat{\lambda}_1) = 0.9940$, indicating a volatility factor which has quite a large degree of memory, with $\exp(\hat{\lambda}_2) = 0.0025$ and $\exp(\hat{\lambda}_3) \approx 0$. The second volatility component is estimated to have the largest weight, with the third component getting very little weight.

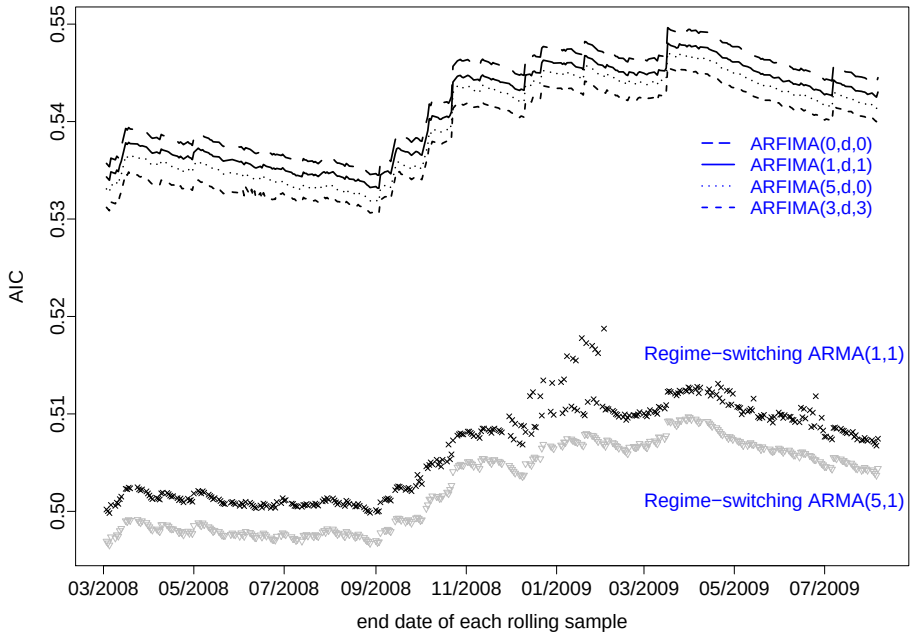
5.3.2.4 Goodness of Fit

We focus on the comparisons between regime switching and ARFIMA models for the moment. The three factor SV model is of state space form, decomposing realized volatility into actual volatility and measurement error. Two types of residuals are obtained from the estimation: one is from the measurement equation, and the other is from the state equation. The three-factor SV model is estimated on realized volatility rather than the log transformation. Hence the corresponding residuals are not directly comparable to those of the regime switching and ARFIMA models. Accordingly, the evaluation of the three factor SV model is only focused on the forecasting performance.

We choose AIC as a measure a goodness of fit. Figure 5.11 plots the AIC based on our recursive estimations, where smaller value of AIC indicates better goodness of fit. Overall, regime switching models appear to be a better description of $\log(RV)$ than ARFIMA models. In addition, models of the same type tend to provide comparable in-sample fit, while the values of AIC distinguish clearly between the two types of model. The financial crisis adversely impact the in-sample fit of each model considered. However, regime switching models produce smaller AIC all through our rolling samples. In particular, the specification of regime switching ARMA(5,1) is consistently the best.

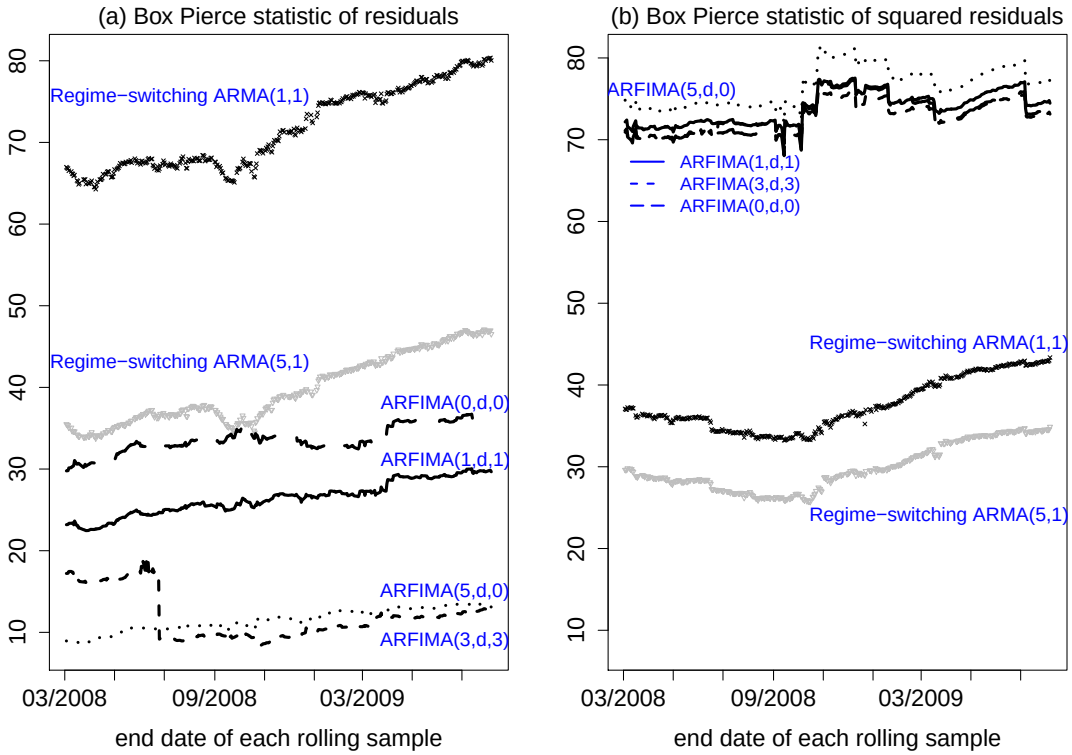
Diagnostic tests are conducted on the residuals from each rolling sample, including

Figure 5.11: AIC of rolling samples



We define that the smaller the AIC, the better the goodness of fit.

Figure 5.12: Serial dependence in residuals and squared residuals of rolling samples



testing for serial correlation on residuals and squared residuals, and a normality test.[22] Figure 5.12 examines the serial dependence in residuals and squared residuals, where Box-Pierce statistics are recorded for each rolling sample of each model examined. The null hypothesis of Box-Pierce test is serially uncorrelated (white noise) series. The statistics reported here are obtained at lag 12.[23] The test statistic is

$$Q(l) = T \sum_{g=1}^{l} \gamma_g^2$$

where l is the order of lag to use for the test, T is the number of observations, and γ_g is the gth autocorrelation. The test statistic is asymptotically distributed as a χ^2 with $l-p-q$ degrees of freedom when applied to the residuals of an ARMA(p,q) model. If the residuals are not based on the results of ARMA models, then under the null hypothesis, $Q(l)$ is asymptotically chi-squared with degrees of freedom equal to l. The critical values of a χ^2 distribution with 12 degrees of freedom is 18.55, 21.03 and 26.22 at the significance levels of 10%, 5% and 1%, respectively.

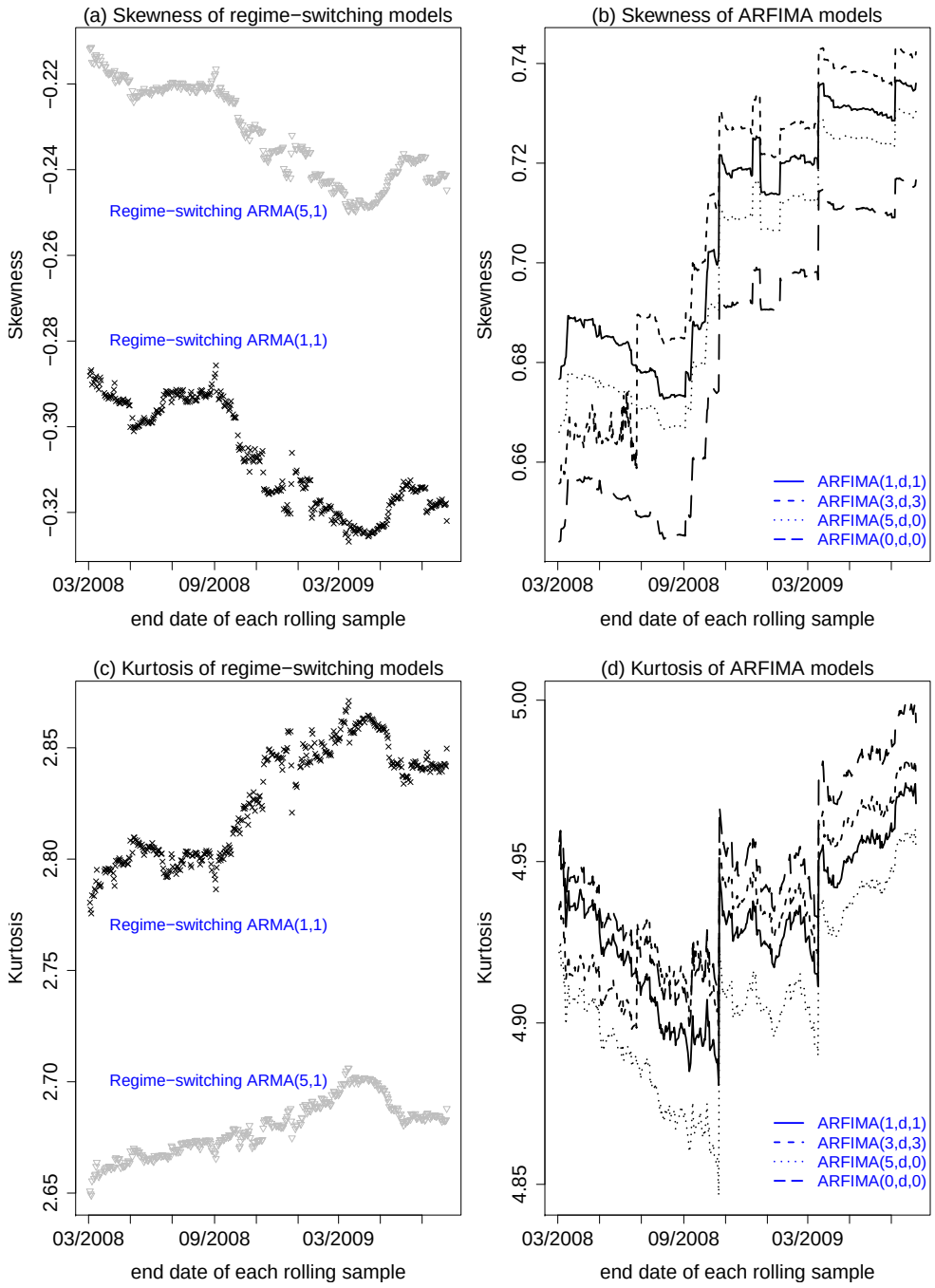
The larger the statistic is, the higher the degree of serial dependence left over in residuals or squared residuals. ARFIMA(3,d,3) and ARFIMA(5,d,0) show the lowest degree of serial dependence in the residual levels (Figure 5.12 Panel (a)), while the memory parameter estimate of ARFIMA(3,d,3) exceeds the stationarity boundary (Figure 5.7). The performances of ARFIMA(0,d,0) and regime switching ARMA(5,1) are similar, with regime switching ARMA(1,1) having the highest degree of serial dependence in residuals. These indicates the relative weakness of the two regime switching models in removing serial correlation in residuals. In terms of serial dependence in squared residuals (Figure 5.12 Panel (b)), regime switching models are better than ARFIMA models. Examination of serial dependence in squared residuals helps to detect and measure the degree of non-linearity unaccounted for by the model used. The inferior performance of the ARFIMA

[22] Hamilton (1996) discusses specifically diagnostic tests of Markov-switching models. In particular, tests for residual autocorrelation, heteroscedasticity, regime mis-specification, and omitted explanatory variables are developed. These tests make heavy use of score vectors, and to conduct them for 360 rolling samples could be computationally costly. Therefore, they are not pursued here.

[23] There remains the practical problem of choosing the order of lag to use for the test. With a small lag chosen, the test may not detect serial correlation at high-order lags. On the other hand, with too large a lag, the test may have low power because the significant correlation at one lag may be diluted by insignificant correlations at other lags.

model indicates its weakness on this aspect compared with regime switching.

Figure 5.13: Skewness and kurtosis of the residuals of rolling samples



With the normality test, residual Jarque-Bera statistics are calculated for each sample. Neither the regime switching nor the fractional integration specifications are able to generate Gaussianity of residuals. The magnitude of Jarque-Bera statistics ranges from from 35 to 60 of the former and from 650 to 900 of the latter, indicating very clear evi-

dence against the null hypothesis of normally distributed error terms. The deviation from normality assumption is much worse of the ARFIMA models compared with that of the regime switching models. The Jarque-Bera test statistic measures the difference of the skewness and kurtosis of the series with those from a normal distribution. The expected values of skewness and kurtosis of a normal distribution are zero and 3. Figure 5.13 plots the residual skewness and kurtosis recorded for each rolling sample of each model examined. In general, the statistics from the same type of models are comparable. The residuals skewness and kurtosis of the regime switching models deviate much less from the expected values than those of their ARFIMA counterparts. The excessive kurtosis of the ARFIMA models (Figure 5.13 Panel (d)) suggests that they are less capable in dealing with extreme volatility values.

Based on these observations, the ARFIMA models in general perform worse than the regime switching models except when we examine serial correlations in residual levels. The performance of each model worsens once the sample includes the global financial crisis, as suggested by the increased magnitudes of all the statistics examined. In the next section, we will examine whether the superiority of regime switching will generate improved out-of-sample forecasting.

5.3.3 Forecasting Performance

For each model, twenty-step ahead predictions, given maximum likelihood estimates, are generated for each rolling sample. With recursive estimations conducted, twenty time series of volatility forecasts and corresponding forecast errors are obtained.[24] An important choice when comparing forecasts is the selection of forecast horizons and we choose to examine each of the 20 horizons.

The discussion in this section is focused on forecasting $\log(RV)$ rather than RV for the following reasons. It is clear from Figure 5.1 that the rolling sample period takes place

[24]Since volatility is not observable, construction of its forecast error will depend on the measure used to approximate volatility. We follow the convention by treating realized volatility as volatility approximation as it provides unbiased estimates of actual volatility. Another commonly used approximation before the advent of realized volatility is squared daily returns, despite the fact that it is a very noisy approximation.

in a very volatile period. There are a few spikes in realized volatility, which will generate influential forecasting errors and mask the forecasting performance significantly, while evaluation based on the logarithm might be more informative.[25] Secondly, both regime switching and fractional integration models are estimated on $\log(RV)$, and back transformation by simply exponentiating the logarithms will introduce bias.[26] Mean squared forecast error (MSFE) is used as our accuracy measure mainly due to its popularity and simplicity. Of course, there is no reason to preclude other accuracy measures, or other loss functions such as symmetric or asymmetric loss functions. This is an interesting topic since forecasting performance is expected to be associated with the choice of loss functions. Since volatility is largely treated as an input in financial instruments pricing, for example options and futures, forecast evaluations under a more practical framework rather than from a purely statistical point of view could be of more empirical relevance. However, this is not pursued here. There is a large literature in forecasting evaluation; see Diebold and Lopez (1996) and Newbold and Harvey (2001), and references therein. For recent applications in volatility forecasting, see Patton and Sheppard (2009).

Tabulations of MSFE for the rolling sample period for all the models estimated are reported in Table 5.2, with Figure 5.14 plotting the MSFE of some of the models. Forecast horizons evaluated range from one to twenty-step ahead. The results show that the ranking of same type of models are similar across forecast horizons. Regime switching models provide the smallest MSFE at almost all the horizons evaluated, the relative performance of which improves as the forecast horizon becomes longer.[27] No great difference is observed in terms of short term forecasts, especially for one-step ahead forecasting. Comparing within the same types of model, the specifications with more parameters tend to perform better, however, the improvement is relatively minor. Of the three types of model considered, regime switching and ARFIMA outperform the single and three-factor SV models.

The ranking of forecasting accuracy is consistent with that of AIC. Better in-sample-fit provides improved out-of-sample forecasts, especially at long-term horizons such as

[25] The volatility of volatility appears to increase as the level of volatility increases, therefore forecast evaluation measured in logarithms is able to gauge the performance in terms of percentage.

[26] Bias is introduced due to the fact that $E\big[f(X)\big] \neq f\big(E[X]\big)$ if $f(X)$ is non-linear.

[27] The application in Calvet and Fisher (2004) suggests regime-switching model is able to improve long-run volatility forecasting.

Table 5.2: Mean squared forecast error of $\log(RV)$ using rolling samples

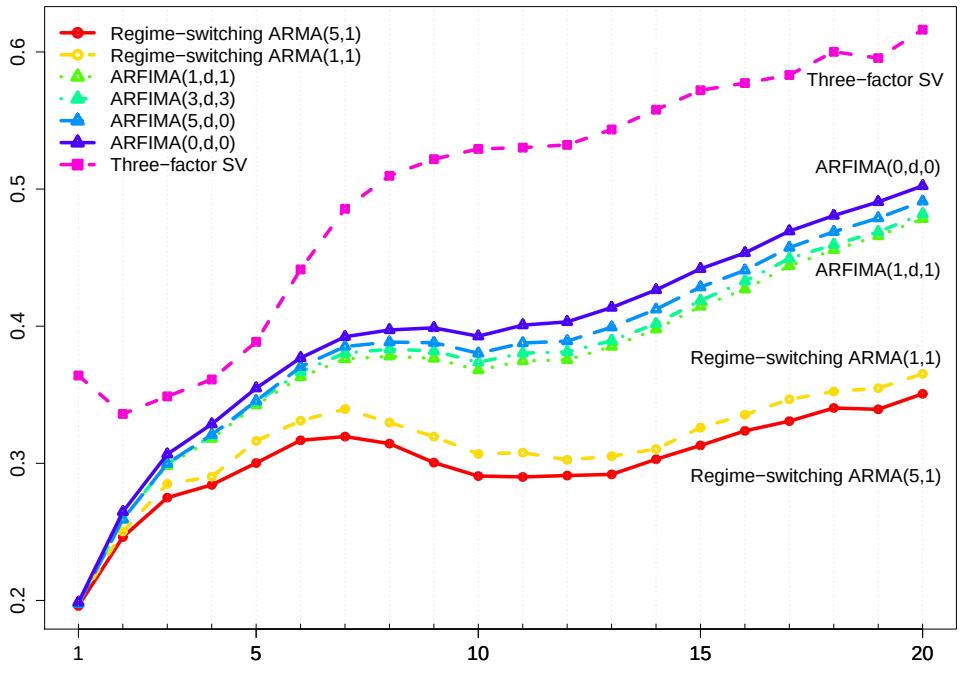
Forecasting horizons		1	5	10	15	20
Models						
Fractional integration	ARFIMA(0,d,0)	0.199	0.355	0.393	0.442	0.502
	ARFIMA(5,d,0)	0.198	0.345	0.380	0.428	0.491
	ARFIMA(3,d,3)	0.198	0.344	0.374	0.419	0.482
	ARFIMA(1,d,1)	0.198	0.342	0.368	0.414	0.478
Regime switching	RS-ARMA(5,1)	0.196	0.300	0.291	0.313	0.351
	RS-ARMA(1,1)	0.199	0.316	0.307	0.326	0.365
	RS-ARFIMA(0,d,0)[a]	0.197	0.324	0.353	0.394	0.440
	RS-ARFIMA(5,d,0)[b]	0.198	0.319	0.354	0.396	0.442
Stochastic volatility	One-factor	0.318	0.681	0.679	0.674	0.680
	Three-factor	0.364	0.389	0.529	0.572	0.616

[a]The specification of RS-ARFIMA(0,d,0) is to assume $\log(RV)$ switching within two regimes, with each regime controlled by a ARFIMA(0,d,0) process, but the level, innovation variance and memory parameter of one process are different from those of another.

[b]The specification of RS-ARFIMA(5,d,0) is similar to that of RS-ARFIMA(0,d,0), but with autoregressive root at lag 5 being estimated for each regime.

[c]Mean squared forecast error (MSFE) is obtained from the rolling sample period of March 4, 2008 to August 6, 2009. Given daily data are examined, the horizons of one-step, five-step and twenty-step represent daily, weekly and monthly, respectively.

Figure 5.14: Mean squared forecast error of $\log(RV)$ using rolling samples



[a]MSFE is plotted on the vertical axis, with forecasting horizons on the horizontal axis.

[b]Compared with Table 5.2, the MSFE of one-factor SV and regime switching ARFIMA models are not plotted in the graph, which is mainly to make the graph easy to read.

monthly. As to fraction integration with different choice of autoregressive and moving average orders, the MSFE of the parsimonious model, say, ARFIMA(1,d,1) is relatively close to but worse than that of the more complicated ARFIMA(3,d,3) specification. This is largely in line with the empirical results shown in Chapter 3, where ARFIMA(p,d,q) model is used to predict volatility, with Bayesian model averaging applied to integrate out uncertainty in model specifications. The empirical results indicated that the weighted average model does not show significant improvement over fractional integration models with different orders of p and q, and forecasts provided by ARFIMA(p,d,q) with different orders of p and q are comparable.

The MSFE of the three factor SV model is the second largest at almost all the horizons examined only one-factor SV is worse. One possible explanation might be the difficulty in numerical estimation of the SV model. Numerical optimization has problem in converging as only weak convergence is attained but strong convergence for some rolling samples even when a range of initial parameters are experimented. The difficulty in parameter estimation has a negative impact on the out-of-sample forecast. This also casts some doubt on the asymptotic results obtained in Barndorff-Nielsen and Shephard (2002a). The performance of the one factor SV model is the worst. This model assumes spot volatility is controlled by one volatility component. In the case of the three factor SV model, it is the sum of three volatility components, each with a different autocorrelation function, by which the long memory property in volatility is captured. The one factor specification lacks such a mechanism, despite the fact that numerical optimization of the three factor model is problematic for some rolling samples.

If the better performance of regime switching and fractional integration can be explained by the finding that $\log(RV)$ is approximately Gaussian, as identified by Andersen, Bollerslev, Diebold and Labys (2001) and implied in the empirical distribution of our sample (Figure 5.3), then how do we explain the improved in-sample fit and relatively smaller MSFE of regime switching over those of ARFIMA models? We follow the argument in Alizadeh et al. (2002): it is desirable for a volatility model to be able to deal with two crucial aspects, i.e. autocorrelations in volatility and volatility of volatility. Both specifications are equally able to describe the slowly-decaying autocorrelations observed in

volatility, as justified by Diebold and Inoue (2001) and other empirical examples discussed in our introduction to this chapter. Then the relative performance of either specification is largely determined by its ability to capture the volatility of volatility. Regime switching specifications allow the volatility of volatility to switch between two regimes. One regime is characterized by high innovation variance, and the other has low innovation variance. As such, more flexibility is allowed in describing volatility dynamics. Given a sample period of over a decade, it is a realistic possibility that volatility evolution is controlled by two (or more) regimes, while the assumption of fractional integration lacks such flexibility. It is in this sense that regime switching is a better description of $\log(RV)$ and more accurate forecasts might be expected. Specifications of more than two regimes may even offer improved AIC and long-run forecasts, however the gain might be limited given the patterns observed in Figures 5.11 and 5.14, where the performance of parsimonious models is comparable to those more complicated specifications for the same type of model.

Let us briefly discuss other models considered. Long memory regime switching models are also estimated.[28] Nevertheless, they are not the main focus of our exercise here as we are concerned with the possibility of using short memory regime switching to mimic long memory property. The specification of regime switching ARFIMA(p,d,q) assumes each regime is controlled by an ARFIMA(p,d,q) process, and the level, innovation variance and memory parameter of each regime are different. This specification nests the single ARFIMA(p,d,q) model. Table 5.2 shows that the MSFE of regime switching ARFIMA(p,d,q) is smaller than the single ARFIMA(p,d,q), but larger than that of regime switching ARMA(p,q). The under-performance of long memory regime switching compared with short memory switching may be due to the increased complexity in likelihood evaluation in the former case. This would likely reduce the accuracy of parameter estimation, and hence adversely affect out-of-sample forecasts by this class of models. In other words, adding an ARFIMA specification to regime switching is not necessary, and a short memory regime switching is sufficient to capture the long memory property observed in the time series.

[28] The recursive estimation results for the long memory regime switching models are available upon request.

5.4 Conclusions and Extensions

A common observation regarding volatility is the slowly decaying autocorrelations, especially with high-frequency data. Realized volatility modeling often involves models of a fractional integration process. However, it is well documented in the literature, with theoretical and empirical support, that alternative specifications, such as regime switching or a sum of short-memory processes, are equally able to mimic the long-range dependence. This chapter conducts a recursive estimation of realized volatility of the JPY/USD rate, with the aim of assessing the abilities of these alternative specifications in modeling and forecasting volatility. Daily realized volatility of the JPY/USD rate during the period 1996-2009 is examined.

The main result is that a short memory regime switching model, which allows the level and innovation variance to switch within regimes, appears to provide a better description of the series being examined than ARFIMA(p,d,q) models with fixed mean and variance. The evidence includes improved in-sample fit, better-behaved residuals, and increased accuracy of long-run volatility forecasting. Short term forecasts of each model are similar, except for the three-factor stochastic volatility model which performs poorly. This result may be due to estimation difficulties. The recent financial turmoil adversely affects stability in estimation of each model, although the superior performance of the regime switching models is observed well before the global financial crisis.

An important difference between fractional integration and regime switching models is in the treatment of the mean. Long term predictions of a time series would rely on the accuracy of the level estimates. With a time series spanning a decade, including occasional volatile periods, fractional integration models lack the flexibility to incorporate abrupt changes in the volatility level. The improved forecasting performance of regime switching models suggests the specifications of model components, such as the mean and variance, are important in describing the long term characteristics of time series.[29]

Within each regime, the volatility evolution is specified by a short memory process.

[29] This is consistent with the comments made by Franses and van Dijk (2000, p. 76) on the important role of intercepts in nonlinear time series models.

By allowing the evolution to switch between two short memory processes, not only is a process with slowly decaying autocorrelations better described, but also better long term prediction is obtained. This is a helpful finding, because it suggests that the long memory property, revealed as slowly decaying autocorrelations of time series, could be effectively modeled by methods which avoid the use of long memory specifications, such as fractionally integrated processes. Or possibly, that what appears to be long memory is not: it is actually short memory with regime switching, and it is with the regime switching mechanism being ignored that gives the appearance of long memory.

Two questions are suggested for future research. First, it is assumed here that switching of regimes is controlled by a Markov process, a purely statistical mechanism. Hamilton (1989) argues that this assumption is not unduly restrictive, although as argued by Schwert (1989), an important question in volatility modeling is why volatility changes over time. In this regard, a switching mechanism explained by other factors, for example, the variables of macroeconomic volatility, financial leverage and trading volume, examined by Schwert (1989), appears to be of greater interest than a purely statistical process. Secondly, there may be value in combining forecasts from different models to improve prediction accuracy.

Chapter 6

Conclusion

This thesis has been concerned with volatility modeling given the fact that volatility is an important input to pricing various financial derivatives. Our treatment is exclusively univariate. Three exercises involving analysis of the JPY/USD exchange rate are conducted, with two of these making use of intra-day data, and the other using daily data. The exercises are concerned especially with the high degree of persistence in volatility, one of the important stylized facts of assets returns.

6.1 Main Findings

The main findings of our exercises can be summarized as follows. First, Bayesian model averaging is helpful in improving forecast accuracy of volatility at long-term horizons, according to the criterion of mean square forecast error. At the daily horizon, the prediction provided by Bayesian model averaging is out-performed only by the single best ARFIMA model. A simulation study conducted also indicates the risk reduction expected from Bayesian model averaging. In terms of the statistical inference of the memory parameter of volatility, the memory parameter estimated from the overall average model indicates significant evidence of long-range dependence in the daily volatility of the JPY/USD exchange rate from January 2003 to August 2008. We argue that the parameter estimate from the overall average model is robust in the sense that by employing a model derived

as the average across a range of long memory and short memory models, part of the uncertainty of model selection is avoided. The risk of making inference based on a single model is addressed in a coherent way. The posterior model probability is used as model weight, which is the byproduct of Bayesian estimation, and therefore the implementation of Bayesian averaging is feasible.

Secondly, a Bayesian method is able to estimate a long memory stochastic volatility model better than the classical method. Our Bayesian estimation is based on the state space form of stochastic volatility models, and the fractional integration is approximated by a moving-average process with high orders. This treatment facilitates the implementation of MCMC simulation. According to a Monte Carlo simulation study conducted with a range of true parameters, the proposed Bayesian estimator, when the posterior mean is used as a point estimator, tends to outperform the classical counterpart by producing smaller root mean square error, with sample sizes from 2000 to 4000. The simulation study also indicates that the estimation of long memory SV models is in general not an easy task when the persistence of latent volatility is of low degree and the innovation to the volatility process is weak. The Bayesian method is used to estimate daily returns for the JPY/USD rate over almost 30 years. Clear evidence of long-range dependence in volatility is observed, with the degree of persistence consistent with that when realized volatility is used.

Finally, short memory regime switching models are shown to be a better description of daily realized volatility than the long memory ARFIMA models. We consider a stationary ARMA model which allows the mean and variance of volatility to switch within regimes. A recursive estimation on the daily JPY/USD from 1996 to 2009 is conducted. The results show that by allowing a short memory specification switching within two regimes, in-sample fit and out-of-sample forecast can be both improved. While the short term forecasts of the models being compared are similar, the long-run volatility forecasts produced by regime switching models are more accurate. This is a useful finding, because it enhances the options to capture the slowly decaying autocorrelations of time series, other than the long memory ARFIMA models. This raises an interesting question. Given the long memory property observed, i.e. slowly-decaying autocorrelations, what is the underlying

data generating process? Long memory models, or switched short memory models? Is it possible that volatility does not really evolve with a long memory data generating process, but rather short memory properties, but non-constant mean and variance?

6.2 Limitations and Extensions

There are several possible extensions to the research conducted in this thesis. The Bayesian model averaging exercise in Chapter 3 considers only ARFIMA and ARMA models, without incorporating the effects of jumps which commonly are observed in high-frequency data (see for example, Barndorff-Nielsen and Shephard (2004*b*, 2006)). It would be useful to deal with this feature in the model averaging exercise, and averaging across different models may affect the outcome.

The long memory SV model examined in Chapter 4 is of the basic form. A possible extension is to specify the latent log volatility as an autoregressive moving-averaging fractional integration, thus allowing for more flexibility. Our sensitivity study indicates that estimation is difficult when both the memory and volatility innovation parameters are of relatively small magnitudes. This is evident from the low convergence rate of the MCMC algorithm, and improving the MCMC convergence is of practical relevance. Alternatively, other Bayesian methods, such as the importance sampling approach, may be employed.

The regime switching specified in Chapter 5 assumes the switching is determined by a Markov process, which is a purely statistical mechanism. It would be helpful to experiment with other switching mechanisms (e.g. an economic mechanism), in the quest to discover the underlying driving force of varying volatility. Also, the empirical results in Chapter 5 indicate the forecasting behavior of various volatility models could be improved. In particular, forecasts derived from combinations of alternative models may improve prediction accuracy, by reducing forecast errors resulting from the use of a single model.

To conclude, this thesis has explored a number of interesting phenomena and possible models and suggested possible fruitful directions for improved modeling of volatility, but it is clear that there is still much left to learn.

Bibliography

Aït-Sahalia, Y., Mykland, P. A. and Zhang, L. (2005), 'How often to sample a continuous-time process in the presence of market microstructure noise', *Review of Financial Studies* **18**, 351–416.

Alizadeh, S., Brandt, M. W. and Diebold, F. X. (2002), 'Range-based estimation of stochastic volatility models', *Journal of Finance* **57**, 1047–1091.

Andersen, T., Bollerslev, T., Christoffersen, P. and Diebold, F. (2006), Volatility and correlation forecasting, *in* G. Elliot, C. W. J. Granger and A. Timmermann, eds, 'Handbook of economic forecasting', Elsevier North Holland, London.

Andersen, T. G. (1994), 'Stochastic autoregressive volatility: a framework for volatility modeling', *Mathematical Finance* **4**, 75–102.

Andersen, T. G. (1996), 'Return volatility and trading volume: an information flow interpretation of stochastic volatility', *Journal of Finance* **51**, 169–204.

Andersen, T. G. and Benzoni, L. (2009), Realized volatility, *in* T. G. Andersen, R. A. Davis, J. Kreiss and M. T., eds, 'Handbook of Financial Time Series', Springer, Berlin; London.

Andersen, T. G. and Bollerslev, T. (1997*a*), 'Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns', *Journal of Finance* **52**, 975–1005.

Andersen, T. G. and Bollerslev, T. (1997*b*), 'Intraday periodicity and volatility persistence in financial markets', *Journal of Empirical Finance* **4**, 115–158.

Andersen, T. G. and Bollerslev, T. (1998*a*), 'Answering the skeptics: yes, standard volatility models do provide accurate forecasts', *International Economic Review* **39**, 885–905.

Andersen, T. G. and Bollerslev, T. (1998*b*), 'Deutsche Mark-Dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies', *Journal of Finance* **53**, 219–265.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F. and Diebold, F. X. (2007), Practical volatility and correlation modeling for financial market risk management, *in* M. Carey and R. Stulz, eds, 'The Risks of Financial Institutions', University of Chicago Press and National Bureau of Economic Research, Chicago, pp. 512–548 (with discussion).

Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2007), 'roughing it up: including jump components in the measurement, modeling and forecasting of return volatility', *Review of Economics and Statistics* **89**, 701–720.

Andersen, T. G., Bollerslev, T. and Diebold, F. X. (2010), Parametric and nonparametric volatility measurement, *in* Y. Aït-Sahalia and L. P. Hansen, eds, 'Handbook of financial econometrics', North-Holland/Elsevier, Amsterdam; Boston.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001), 'The distribution of realized stock return volatility', *Journal of Financial Economics* **61**, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2000*a*), 'Exchange rate returns standardized by realized volatility are (nearly) Gaussian', *Multinational Finance Journal* **4**, 159–179.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2000*b*), 'Great realizations', *Risk* **13**, 105–108.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001), 'The distribution of realized exchange rate volatility', *Journal of the American Statistical Association* **92**, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003), 'Modelling and forecasting realized volatility', *Econometrica* **71**, 579–625.

Andersen, T. G. and Sørensen, B. E. (1996), 'GMM estimation of a stochastic volatility model: a Monte Carlo study', *Journal of Business and Economic Statistics* **13**, 328–352.

Andreou, E. and Ghysels, E. (2002), 'Detecting multiple breaks in financial market volatility dynamics', *Journal of Applied Econometrics* **17**, 579–600.

Andreou, E. and Ghysels, E. (2009), Structural breaks in financial time series, *in* T. G. Andersen, R. A. Davis, J. Kreiss and M. T., eds, 'Handbook of Financial Time Series', Springer, Berlin; London.

Andrews, D. W. K. and Sun, Y. (2004), 'Adaptive local polynomial Whittle estimation of long-range dependence', *Econometrica* **72**, 569–614.

Arteche, J. (2004), 'Gaussian semi-parametric estimation in long memory in stochastic volatility and single plus noise models', *Journal of Econometrics* **119**, 131–154.

Back, K. (1991), 'Asset pricing for general processes', *Journal of Mathematical Economics* **20**, 371–395.

Baillie, R. T. (1996), 'Long memory processes and fractional integration in econometrics', *Journal of Econometrics* **73**, 5–59.

Baillie, R. T. and Bollerslev, T. (1989), 'The message in daily exchange rates: A conditional-variance tale', *Journal of Business and Economic Statistics* **7**, 297–305.

Baillie, R. T. and Bollerslev, T. (1990), 'Intra-day and inter-market volatility in foreign exchange rates', *Review of Economic Studies* **58**, 565–585.

Baillie, R. T., Bollerslev, T. and Mikkelsen, H. O. (1996), 'Fractionally integrated generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **74**, 3–30.

Bandi, F. M. and Russell, J. R. (2008), 'Microstructure noise, realized variance, and optimal sampling', *Review of Economic Studies* **75**, 339–369.

Banerjee, A. and Urga, G. (2005), 'Modeling structure breaks, long memory and stock

market volatility: an overview', *Journal of Econometrics* **129**, 1–34.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008), 'Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise', *Econometrica* **76**, 1481–1536.

Barndorff-Nielsen, O. E. and Shephard, N. (2001), 'Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics', *Statistical Society: Series B* **63**, 167 – 241.

Barndorff-Nielsen, O. E. and Shephard, N. (2002*a*), 'Econometric analysis of realized volatility and its use in estimating stochastic volatility models', *Journal of the Royal Statistical Society B* **64**, 253–280.

Barndorff-Nielsen, O. E. and Shephard, N. (2002*b*), 'Estimating quadratic variation using realized variance', *Journal of Applied Econometrics* **17**, 457–477.

Barndorff-Nielsen, O. E. and Shephard, N. (2004*a*), 'Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics', *Econometrica* **72**, 885–925.

Barndorff-Nielsen, O. E. and Shephard, N. (2004*b*), 'Power and bipower variation with stochastic jumps (with discussion)', *Journal of Financial Econometrics* **2**, 1–48.

Barndorff-Nielsen, O. E. and Shephard, N. (2006), 'Econometrics of testing for jumps in financial economics using bipower variation', *Journal of Financial Econometrics* **4 (1)**, 1–30.

Barndorff-Nielsen, O. E. and Shephard, N. (2007), Variation, jumps and high frequency data in financial econometrics, *in* R. Blundell, P. Torsten and W. K. Newey, eds, 'Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress', Econometric Society Monograph, Cambridge University Press.

Bates, D. S. (1995), Testing option pricing models, Technical report, Wharton School, University of Pennsylvania.

Beran, J. (1994), *Statistics for Long-Memory Processes*, Chapman & Hall, New York.

Black, F. (1976), Studies of stock market volatility changes, *in* 'Proceedings of the American Statistical Association, Business and Economic Statistics Section', pp. 177–181.

Black, F. and Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**, 637–654.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroscedasticity', *Journal of Econometrics* **31**, 307–327.

Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992), 'ARCH modeling in finance: a review of the theory and empirical evidence', *Journal of Econometrics* **52**, 5–59.

Bollerslev, T., Engle, R. F. and Nelson, D. B. (1994), ARCH models, *in* R. F. Engle and D. L. McFaddde, eds, 'Handbook of Econometrics', Vol. IV, Elsevier Science, Amsterdam, chapter 49.

Bollerslev, T. and Ghysels, E. (1996), 'Periodic autoregressive conditional heteroscedasticity', *Journal of Business and Economic Statistics* **14**, 139–151.

Bollerslev, T. and Mikkelsen, H. O. (1996), 'Modeling and pricing long memory in stock market volatility', *Journal of Econometrics* **73**, 151–184.

Bollerslev, T. and Mikkelsen, H. O. (1999), 'Long-term equity anticipation securities and stock market volatility dynamics', *Journal of Econometrics* **92**, 75–99.

Bollerslev, T. and Wooldridge, J. M. (1992), 'Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances', *Econometric Reviews* **11**, 143 – 172.

Bollerslev, T. and Wright, J. H. (2000), 'Semiparametric estimation of long-memory volatility dependencies: the role of high-frequency data', *Journal of Econometrics* **98**, 81–106.

Bollerslev, T. and Wright, J. H. (2001), 'High-frequency data, frequency domain inference, and volatility forecasting', *Review of Economics and Statistics* **83**, 596–602.

Brandt, M. W. and Diebold, F. X. (2006), 'A no-arbitrage approach to range-based estimation of return covariances and correlations', *Journal of Business* pp. 61–74.

Breidt, F. J. and Carriquiry, A. L. (1996), Improved quasi-maximum likelihood estimation for stochastic volatility models, Technical report, Department of Statistics. Iowa State University.

Breidt, F. J., Crato, N. and deLima, P. (1998), 'The detection and estimation of long memory in stochastic volatility', *Journal of Econometrics* **83**, 325–348.

Brock, W., Lakonishok, J. and LeBaron, B. (1992), 'Simple technical trading rules and the stochastic properties of stock returns', *Journal of Finance* **47**, 1731–1764.

Brockwell, P. J. and Davis, R. A. (1991), *Time Series: Theory and Methods*, 2 edn, Springer-Verlag, New York.

Calvet, L. E. and Fisher, A. J. (2004), 'How to forecast long-run volatility: regime switching and the estimation of multifractal processes', *Journal of Financial Econometrics* **2**, 49–83.

Campbell, J. Y. and Hentschel, L. (1992), 'No news is good news: An asymmetric model of changing volatility in', *Journal of Financial Economics* **31**, 281–318.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997), *The econometrics of Financial Markets*, Princeton University Press, Princeton.

Carter, C. and Kohn, R. (1994), 'On gibbs sampling for state-space models', *Biometrika* **81**, 541–553.

Chan, N. H. and Giovanni, P. (2000), 'Long memory stochastic volatility: a Bayesian approch', *Communications in Statistics - Theory and Method* **29: 5**, 1367–1378.

Chan, N. H. and Palma, W. (1998), 'State space modelling of lone-memory process', *Annals of Statistics* **26**, 719–740.

Chernov, M., Gallant, A. R., Ghysels, E. and Tauchen, G. (2003), 'Alternative models for stock price dynamics', *Journal of Econometrics* **116**, 225–257.

Cheung, Y.-W. and Diebold, F. X. (1993), 'On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean', *Journal of*

Econometrics **62**, 301–306.

Chib, S. and Greenberg, E. (1995), 'Understanding the Metropolis-Hastings algorithm', *The American Statistician* **49**, 327–335.

Clark, P. K. (1973), 'A subordinated stochastic process model with fixed variance for speculative prices', *Econometrica* **41**, 135–156.

Comte, F. and Renault, E. (1998), 'Long memory in continuous-time stochastic volatility models', *Mathematical Finance* **8**, 291–323.

Corsi, F. (2009), 'A simple approximate long-memory model of realized volatility', *Journal of Financial Econometrics* **7**, 1–23.

Corsi, F., Zumbach, G., Müller, U. A. and Dacorogna, M. (2001), 'Consistent high-precision volatility from high-frequency data', *Economic Notes* **30**, 183–204.

Cowles, M. K. and Carlin, B. P. (1996), 'Markov chain Monte Carlo convergence diagnostics: a comparative review', *Journal of the American Statistical Association* **91**, 883–904.

Dacorogna, M. M., Gençay, R., Müller, U. A., Olsen, R. B. and Pictet, O. V. (2001), *An Introduction to High-Frequency Finance*, Academic Press, San Diego; London.

Dacorogna, M. M., Müller, U. A., Nagler, R. J., Olsen, R. B. and Pictet, O. V. (1993), 'A geographical model for the daily and weekly seasonal volatility in the foreign exchange market', *Journal of International Money and Finance* **12**, 413–438.

Danielsson, J. (1994), 'Stochastic volatility in asset prices estimation with simulated maximum likelihood', *Journal of Econometrics* **64**, 375–400.

Danielsson, J. and Richard, J. F. (1993), 'Accelerated Gaussian importance sampler with application to dynamic latent variable models', *Journal of Applied Econometrics* **8**, 153–173.

Dassios, A. (1995), Asymptotic expressions for approximations to stochastic variance models, mimeo, London School of Economics.

Davidian, M. and Carroll, R. J. (1987), 'Variance function estimation', *Journal of the*

American Statistical Association **82**, 1079–1091.

Davies, R. B. (1987), 'Hypothesis testing when a nuisance parameter is present only under the alternative', *Biometrika* **74 (1)**, 33–43.

de Jong, P. and Shephard, N. (1995), 'The simulation smoother for time series models', *Biometrika* **82**, 339–350.

Deo, R. S. and Hurvich, C. M. (2001), 'On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models', *Econometric Theory* **17**, 686–710.

Deo, R. S. and Hurvich, C. M. (2003), Estimation of long memory in volatility, *in* P. Doukhan, O. G. George and M. S. Taqqu, eds, 'Theory and Applications of Long-Range Dependence', MA: Birkhauser, Boston.

Diebold, F. X. (1988), *Empirical modeling of exchange rate dynamics*, Springer-Verlag, Berlin; New York.

Diebold, F. X. and Inoue, A. (2001), 'Long memory and regime switching', *Journal of Econometrics* **105**, 131–159.

Diebold, F. X. and Lopez, J. A. (1996), Forecast evaluation and combination, *in* G. S. Maddala and C. R. Rao, eds, 'Handbook of Statistics', Vol. 14, Elsevier, Amsterdam.

Ding, Z. and Granger, C. W. J. (1996), 'Modeling volatility persistence of speculative returns: A new approach', *Journal of Econometrics* **73**, 185–215.

Ding, Z., Granger, C. W. J. and Engle, R. F. (1993), 'A long memory property of stock market returns and a new model', *Journal of Empirical Finance* **1**, 83–106.

Doornik, J. A. and Ooms, M. (2003), 'Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models', *Computational Statistics & Data Analysis* **42**, 333–348.

Doornik, J. A. and Ooms, M. (2004), 'Inference and forecasting for ARFIMA models, with an application to US and UK inflation', *Studies in Nonlinear Dynamics and Economet-*

rics **8 (2)**, Article 14.

Doornik, J. A. and Ooms, M. (2006*a*), *A package for estimating, forecasting, and simulating Arfima Models: Arfima package 1.04 for Ox*, Arfima Package Manual.

Doornik, Jurgen, A. and Ooms, M. (2006*b*), *A package for estimating, forecasting and simulating Arfima Models: Arfima package* 1.04 *for Ox*, Nuffield College, Oxford OX1 1NF, UK, Free University, Amsterdam, The Netherlands.

Doucet, A., de Freitas, N. and Gordon, N., eds (2001), *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York.

Drost, F. C. and Nijman, T. E. (1993), 'Temporal aggregation of GARCH processes', *Econometrica* **61**, 909–927.

Durbin, J. and Koopman, S. J. (1997), 'Monte Carlo maximum likelihood estimation for non-Gaussian state space models', *Biometrika* **84 (3)**, 669–684.

Durbin, J. and Koopman, S. J. (2001*a*), *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.

Durbin, J. and Koopman, S. J. (2001*b*), *Time series analysis by state space methods*, Oxford University Press, Oxford.

Elerian, O., Chib, S. and Shephard, N. (2001), 'Likelihood inference for discretely observed non-linear diffusions', *Econometrica* **69**, 959–993.

Engle, R. F. (1982), 'Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation', *Econometrica* **50**, 987–1007.

Engle, R. F. and Bollerslev, T. (1986), 'Modeling the persistence of conditional variances', *Econometric Reviews* **1**, 1–50.

Engle, R. F., ed. (1995), ARCH*: selected readings*, Oxford University Press, Oxford; New York.

Engle, R. F. and Lee, G. G. J. (1999), A long-run and short-run component model of stock return volatility, *in* R. E. Engle and H. White, eds, 'Cointegration, causality, and

forecasting: a festschrift in honor of Clive W. J. Granger', Oxford University Press, Oxford.

Epps, T. W. and Epps, M. L. (1976), 'The stochastic dependence of security price changes and transaction volumes: implications for the mixture-of-distributions hypothesis', *Econometrica* **44**, 305–321.

Eraker, B. (2001), 'Markov chain Monte Carlo analysis of diffusion models with application to finance', *Journal of Business and Economic Statistics* **19**, 177–191.

Fama, E. F. (1963), 'Mandelbrot and the stable paretian hypothesis', *Journal of Business* **36**, 420–429.

Fama, E. F. (1965), 'The behavior of stock market prices', *Journal of Business* **38**, 34–105.

Ferraz, R. O. and Hotta, L. K. (2007), 'Quasi-Maximum likelihood estimation of long-memory stochastic volatility models', *Brazilian Review of Econometrics* **27**, 225–233.

Filardo, A. J. and Gordon, S. F. (1998), 'Business cycle durations', *Journal of Econometrics* **85**, 99–123.

Fox, R. and Taqqu, M. S. (1986), 'Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series', *The Annals of Statistics* **14**, 517–532.

Francq, C. and Zakoïan, J.-M. (2001), 'Stationarity of multivariate Markov-switching ARMA models', *Journal of Econometrics* **102**, 339–364.

Franses, P. H., Leij, M. V. D. and Paap, D. (2002), 'Modeling and forecasting level shifts in absolute returns', *Journal of Applied Econometrics* **17**, 601–616.

Franses, P. H. and van Dijk, D. (2000), *Non-linear Time Series Models in Empirical Finance*, Cambridge University Press, New York.

French, K. R., Schwert, G. W. and Stambaugh, R. F. (1987), 'Expected stock returns and volatility', *Journal of Financial Economics* **19**, 3–29.

Frühwirth-Schnatter, S. (1994), 'Data augumentation and dynamic linear models', *Journal*

of Time Series Analysis **15**, 183–202.

Fuller, W. A. (1996), *Introduction to Statistical Time Series*, 2 edn, Wiley, New York.

Gallant, A. R., Hsu, C. T. and Tauchen, G. (1999), 'Using daily range data to calibrate volatility diffusions and extract the forward integrated variance', *Review of Economics and Statistics* **81**, 617–631.

Gallant, A. R. and Tauchen, G. (1996), 'Which moments to match?', *Econometric Theory* **12**, 657–681.

Gallant, A. R. and Tauchen, G. E. (1998), 'Reprojecting partially observed systems with application to interest rate diffusions', *Journal of the American Statistical Association* **93**, 10–24.

Gamerman, D. and Lopes, H. F. (2006), *Markov Chain Monte Carlo : Stochastic Simulation for Bayesian Inference*, Taylor & Francis, Boca Raton.

Garman, M. B. and Klass, M. J. (1980), 'On the estimation of security price volatilities from historical data', *Journal of Business* **53**, 67–78.

Gelfand, A. E. and Smith, A. F. M. (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association* **85**, 398–409.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004), *Bayesian data analysis*, Chapman & Hall, Boca Raton.

Gelman, A. and Rubin, D. B. (1992), 'Inference from iterative simulation using multiple sequences', *Statistical Science* **7**, 457–472.

Geweke, J. (1989), 'Bayesian inference in econometric models using Monte Carlo integration', *Econometrica* **57**, 1317–1339.

Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to calculating posterior moments, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds, 'Bayesian Statistics 4', Oxford University Press, New York, pp. 169–193.

Geweke, J. (1999), 'Using simulation methods for Bayesian econometric models: inference,

development and communication (with discussion and rejoinder)', *Econometric Reviews* **18**, 1–126.

Geweke, J. (2005), *Contemporary Bayesian econometric and statistics*, John Wiley & Sons, Hoboken.

Geweke, J. and Porter-Hudak (1983), 'The estimation and application of long time series models', *Journal of Time Series Analysis* **4**, 221–238.

Geweke, J. and Whiteman, C. (2006), Bayesian forecasting, *in* R. Elliott, C. Granger and A. Timmermann, eds, 'Handbook of Economic Forecasting', Vol. 1, Elsevier North Holland, Amsterdam, chapter 1, pp. 3–80.

Ghysels, E., Harvey, A. and Renault, E. (1996), Stochastic volatility, *in* S. Maddala and C. R. Rao, eds, 'Handbook of Statistics 14, Statistical Methods in Finance', North-Holland, Amsterdam, pp. 119–191.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J., eds (1996), *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London; Melbourne.

Giraitis, L., Kokoszka, P., Leipus, R. and Teyssière, G. (2003), 'Rescaled variance and related tests for long memory in volatility and levels', *Journal of Econometrics* **112**, 265–294.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *Journal of Finance* **48**, 1779–1801.

Gourieroux, C. and Jasiak, J. (2001), *Financial econometrics: problems, models, and methods*, Princeton University Press, Princeton, N.J.; Chichester.

Granger, C. W. J. (1966), 'The typical spectral shape of an economic variable', *Econometrica* **34**, 150–161.

Granger, C. W. J. (1980), 'Long memory relationships and the aggregation of dynamic models', *Journal of Econometrics* **14**, 227–238.

Granger, C. W. J. and Hyung, N. (2004), 'Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns', *Journal of Empirical Finance* **11**, 399–421.

Granger, C. W. J. and Joyeux, R. (1980), 'An introduction to long-memory time series models and fractional differencing', *Journal of Time Series Analysis* **1**, 15–29.

Hamilton, J. D. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica* **57**, 357–384.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton.

Hamilton, J. D. (1996), 'Specification testing in Markov-switching time-series models', *Journal of Econometrics* **70**, 127–157.

Hamilton, J. D. (2005), Regime switching models. Department of Economics, University of San Diego, California.

Hansen, B. E. (1992), 'The likelihood ratio test under nonstandard conditions: testing the Markov-switching model of GNP', *Journal of Applied Econometrics* **7**, S61–S82.

Hansen, L. P. and Scheinkman, J. A. (1995), 'Back to the future: Generating moment implications for continuous-time Markov processes', *Econometrica* **63**, 767–804.

Hansen, P. R. and Lunde, A. (2005), 'A realized variance for the whole day based on intermittent high-frequency data', *Journal of Financial Econometrics* **3**, 525–554.

Hansen, P. R. and Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business and Economic Statistics* **24**, 127–161.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Harvey, A. C. (1998), Long memory in stochastic volatility, *in* J. Knight and E. Satchell, eds, 'Forecasting Volatility in Financial Markets', Butterworth-Haineman, London.

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

Harvey, C. R. and Huang, R. D. (1991), 'Volatility in the foreign currency futures market', *Review of Financial Studies* **4**, 543–569.

Hastings, W. K. (1970), 'Monte Carlo sampling methods and using Markov chains and their applications', *Biometrika* **57**, 97–109.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999), 'Bayesian model averaging: a tutorial', *Statistical Science* **14 (4)**, 382–417.

Hosking, J. R. M. (1981), *Biometrika* **68**, 165–176.

Hsieh, D. A. (1991), 'Chaos and nonlinear dynamics: Application to financial markets', *Journal of Finance* **46**, 1839–1877.

Hsu, N.-J. and Breidt, F. J. (2003), 'Bayesian analysis of fractionally integrated ARMA with additive noise', *Journal of forecasting* **22**, 491–514.

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

Hurst, H. (1951), 'Long term storage capacity of reservoirs', *Transactions of the American Society of Civil Engineers* **116**, 770–799.

Hyung, N., Poon, S.-H. and Granger, C. W. J. (2006), The source of long memory in financial market volatility, Working paper, University of California at San Diego.

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

Jensen, M. J. (2000), Bayesian inference of long-memory dependence in volatility via wavelets. Department of Economics, University of Missouri.

Johannes, M. and Polson, N. (2008), Particle filtering, *in* T. G. Andersen, R. A. Davis, J. Kreiss and M. T., eds, 'Handbook of Financial Time Series', Springer-Verlag, Berlin.

Jones, C. S. (1998), Bayesian estimation of continuous-time finance models, Working paper, Simon Graduate School of Business, University of Rochester.

Jungbacker, B. and Koopman, S. J. (2007), 'Monte Carlo estimation for non-Gaussian state space models', *Biometrika* **94 (4)**, 827–839.

Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems', *Journal of Basic Engineering, Transactions ASMA, Series D* **82**, 35–45.

Kim, C. J. and Nelson, C. R. (1999), *State-space models with regime switching : classical and Gibbs-sampling approaches with applications*, MIT Press, Cambridge, Mass.

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

Knight, J. and Satchell, S., eds (2007), *Forecasting volatility in the financial markets*, Butterworth-Heinemann, Oxford.

Koop, G. (2003), *Bayesian econometrics*, J. Wiley, Chichester; Hoboken, N.J.

Koop, G., Ley, E., Osiewalski, J. and Steel, M. F. J. (1997), 'Bayesian analysis of long memory and persistence using ARFIMA models', *Journal of Econometrics* **76**, 149–169.

Koop, G. and Potter, S. M. (1999), 'Bayes factors and nonlinearity: evidence from economic time series', *Journal of Econometrics* **88**, 251–281.

Koopman, S. J., Shephard, N. and Doornik, J. A. (1998), 'Statistical algorithms for models in state space using Ssf-Pack 2.2', *Econometrics Journal* **1**, 1–55.

Liu, M. (2000), 'Modeling long memory in stock market volatility', *Journal of Econometrics* **99**, 139–171.

Lo, A. W. (1991), 'Long term memory in stock market prices', *Econometrica* **59 (5)**, 1279–1313.

Maekawa, K. and Xinhong, L. (2009), Long memory in the realized volatility of returns on the Yen/US$ exchange rate during the three financial crises. Hiroshima University of Economics, Working Paper.

Mandelbrot, B. (1963), 'The variations of certain speculative prices', *Journal of Business* **36**, 394–419.

McAleer, M. and Medeiros, M. C. (2008), 'Realized volatility: A review', *Econometric Reviews* **27**, 10–45.

Meddahi, N. (2002), 'A theoretical comparison between integrated and realized volatility', *Journal of Applied Econometrics* **17**, 479–508.

Meddahi, N. and Renault, E. (2004), 'Temporal aggregation of volatility models', *Journal of Econometrics* **119**, 355–379.

Melino, A. and Turnbull, S. M. (1990), 'Pricing foreign currency options with stochastic volatility', *Journal of Econometrics* **45**, 239–265.

Merton, R. C. (1969), 'Lifetime portfolio selection under uncertainty: The continuous-time case', *Review of Economics and Statistics* **51**, 247–257.

Merton, R. C. (1980), 'On estimating the expected return on the market: an exploratory investigation', *Journal of Financial Economics* **8**, 323–361.

Morgan, J. P. (1997), *RiskMetrics, Technical Documents*, 4th edn, New York.

Müller, P. (1991), A generic approach to posterior integration and Gibbs sampling, Technical report, Department of Statistics, Purdue University.

Müller, U. A., Dacorogna, M. M., Olsen, R. B., Pictet, O. V., Schwarz, M. and Morgenegg, C. (1990), 'Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis', *Journal of Banking and Finance* **14**, 1189–1208.

Nelson, D. B. (1990), 'ARCH models as diffusion approximations', *Journal of Econometrics* **45**, 7–38.

Nelson, D. B. (1991), 'Conditional heteroskedasticity in asset returns: a new approach', *Econometrica* **59**, 347–370.

Nelson, D. B. (1992), 'Filtering and forecasting with misspecified ARCH models I: getting the right variance with the wrong model', *Journal of Econometrics* **52**, 61–90.

Nelson, N. B. and Foster, D. P. (1994), 'Asymptotic filtering theory for univariate ARCH

models', *Econometrica* **62**, 1–41.

Newbold, P. and Harvey, D. I. (2001), Forecast combination and encompassing, *in* M. P. Clements and D. F. Henry, eds, 'A Companion to Economic Forecasting', Blackwells, Oxford.

Officer, R. (1973), 'The variability of the market factor of the New York Stock Exchange', *Journal of Business* **46**, 434453.

Pai, J. S. and Ravishanker, N. (1998), 'Bayesian analysis of autoregressive fractionally integrated moving-average processes', *Journal of Time Series Analysis* **19 (1)**, 99–112.

Palma, W. (2007), *Long-memory time series: theory and methods*, John Wiley & Sons, Hoboken.

Parkinson, M. (1980), 'The extreme value method for estimating the variance of the rate of return', *Journal of Business* **53**, 61–65.

Patton, A. J. and Sheppard, K. (2009), Evaluating volatility and correlation forecasts, *in* T. G. Andersen, R. A. Davis, J. Kreiss and M. T., eds, 'Handbook of Financial Time Series', Springer, Berlin; London.

Pérez, A. and Ruiz, E. (2001), 'Finite sample properties of a QML estimator of stochastic volatility models with long memory', *Economic Letters* **70**, 157–164.

Pitt, M. K. and Shephard, N. (1999), 'Filtering via simulation: Auxiliary particle filters', *Journal of the American Statistical Association* **94**, 590–599.

Poon, S. and Granger, C. W. J. (2003), 'Forecasting volatility in financial markets: A review', *Journal of Economic Literature* **41**, 478539.

Poon, S. and Taylor, S. J. (1992), 'Stock returns and volatility: An empirical study of the UK stock market', *Journal of Banking and Finance* **16**, 37–59.

Praetz, P. D. (1972), 'The distribution of share price changes', *Journal of Business* **45**, 49–55.

Protter, P. E. (1990), *Stochastic integration and differential equations: a new approach*,

Springer-Verlag, Berlin; New York.

Ripley, B. D. (1987), *Stochastic Simulation*, Wiley, New York.

Roberts, G. O., Gelman, A. and Gilks, W. R. (1997), 'Weak convergence and optimal scaling of random walk metropolis algorithms', *The Annals of Applied Probability* **7**, 110–120.

Roberts, G. O. and Stramer, O. (2001), 'On inference for partially observed nonlinear diffusion models using the Hastings-Metropolis algorithm', *Biometrika* **88**, 603–621.

Robinson, P. M. (1994), 'Efficient tests of nonstationary hypotheses', *Journal of the American Statistical Association* **89**, 1420–1437.

Robinson, P. M. (2001), 'The memory of stochastic volatility models', *Journal of Econometrics* **101**, 195–218.

Sandmann, G. and Koopman, S. J. (1998), 'Estimation of stochastic volatility models via Monte Carlo maximum likelihood', *Journal of Econometrics* **87**, 271–301.

Schwert, G. W. (1989), 'Why does stock volatility change over time?', *Journal of Finance* **44**, 1115–1153.

Schwert, G. W. (1990), 'Stock market volatility', *Financial Analysts Journal* **46**, 23–34.

Shephard, N. (1996), Statistical aspects of ARCH and stochastic volatility models, *in* D. R. Cox, D. V. Hinkley and O. E. Barndorff-Nielsen, eds, 'Time Series Models in Econometrics, Finance and Other Fields', Chapman & Hall, London, pp. 1–67.

Shephard, N. and Andersen, T. G. (2009), Stochastic volatility: origins and overview, *in* T. G. Andersen, R. A. Davis, J. Kreiss and M. T., eds, 'Handbook of Financial Time Series', Springer, Berlin; London.

Shephard, N., ed. (2005), *Stochastic Volatility: Selected Readings*, Oxford University Press, Oxford; New York.

Sowell, F. (1992), 'Maximum likelihood estimation of stationary univariate fractionally integrated models', *Journal of Econometrics* **53**, 165–188.

Stelzer, R. (2009), 'On markov-switching ARMA processes-stationarity, existence of moments, and geometric ergodicity', *Econometric Theory* **25**, 43–62.

Sun, Y. and Phillips, P. C. B. (2004), 'Nonlinear log-periodogram regression for perturbed fractional processes', *Journal of Econometrics* **115**, 355–389.

Tanner, T. and Wong, W. (1987), 'The calculation of posterior distributions by data argumentation', *Journal of the American Statistical Association* **82**, 528–549.

Tauchen, G. E. and Pitts, M. (1983), 'The price variability-volume relationship on speculative markets', *Econometrica* **51**, 485–505.

Taylor, S. J. (1982), Financial returns modeled by the product of two stochastic processes - a study of daily sugar prices 1961-79, *in* O. D. Anderson, ed., 'Time Series Analysis : Theory and Practice 1', North-Holland, Amsterdam, pp. 203–226.

Taylor, S. J. (1986), *Modeling financial time series*, John Wiley, Chichester.

Taylor, S. J. (1994), 'Modeling stochastic volatility: A review and comparative study', *Math. Finance* **4**, 183–204.

Taylor, S. J. (2005), *Asset Price Dynamics, Volatility, and Prediction*, Princeton University Press, New Jersey.

Taylor, S. J. and Xu, X. (1997), 'The incremental volatility information in one million foreign exchange quotations', *Journal of Empirical Finance* **4**, 317–340.

Thomakos, D. D. and Wang, T. (2003), 'Realized volatility in the futures markets', *Journal of Empirical Finance* **10**, 321–353.

Tiao, G. C. and Tsay, R. S. (1994), 'Some advances in non-linear and adaptive modeling in time-series', *Journal of Forecasting* **13**, 109–131.

Timmermann, A. (2000), 'Moments of Markov switching models', *Journal of Econometrics* **96**, 75–111.

Tong, H. (1978), On a threshold model, *in* C. Chen, ed., 'Pattern Recognition and Signal Precessing', Sijthoff & Noordhoff, Amsterdam.

Whittle, P. (1953), 'Estimation and information in stationary time series', *Arkiv för Matematik* **2**, 423–434.

Xu, L., Liu, C. and Nie, G. (2006), 'Bayesian estimation and the application of long memory stochastic volatility models', *Statistical Methodology* **3**, 483–489.

Yajima, Y. (1988), 'On estimation of a regression model with long-memory stationary errors', *Annals of Statistics* **16**, 791–807.

Zakoian, J. M. and Rabemananjara, R. (1993), 'Threshold ARCH models and asymmetries in volatility', *Journal of Applied Econometrics* **8**, 31–49.

Zellner, A. and Min, C.-K. (1995), 'Gibbs sampler convergence criteria', *Journal of the American Statistical Association* **90**, 921–927.

Zhou, B. (1996), 'High-frequency data and volatility in foreign exchange rates', *Journal of Business and Economic Statistics* **14**, 45–52.

Zivot, E. (2005), Analysis of high frequency financial data: Models, methods and software. University of Washington.