Essays in Nonstationary Financial Time Series

Manabu Asai

Essays in Nonstationary Financial Time Series

Manabu Asai

Doctoral Program in Policy and Planning Sciences, University of Tsukuba

Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Economics

Abstract

The main purposes of this dissertation are: (1) to carry out empirical investigations of nonstationary financial time series based on available econometric techniques, and then (2) to develop time series techniques to test for the nonstationarity of variance.

Chapter 1 investigates the relationship between macroeconomic variables, such as the industrial production index, interest rate and inflation rate, and the stock market, using Toda and Yamamoto's (1995) vector autoregressions (VAR) specification. The major findings are: (1) macroeconomic variables do Granger cause the stock market variable, where the reverse causality is ambiguous. (2) The lagged stock market variable affects its own current value but its impact tends to diminish in the long–run. Based upon the findings including the above two, we draw a policy implication that the PKO by the Japanese government would not work, but appropriate macroeconomic policies would benefit not only the real side of the macroeconomy but also the stock market.

Chapter 2 compares in- and out-of-sample forecasting performances of the Heath, Jarrow, and Morton (HJM) (1992) model, a short rate model that nests popular spot rate models, ARIMA, VAR, and a prediction method based on the SABL, using the daily three month Euroyen interest futures data. Predictive performances of the forecasts from the HJM model resemble to those of the other models, except for the *SABL*. SABL's performance is sometimes excellent, but occasionally very poor. It seems that such contrasting differences in the *SABL* performance are due to non-liner trend com-

ponents' changing behavior.

Chapter 3 derives a log-GARCH representation of a class of stochastic volatility (SV) models, including the ARMA-SV model, and analyzes the finite sample properties of a Quasi-Maximum Likelihood (QML) estimator. The Monte Carlo results indicate that their finite sample properties are (1) superior to those of the Generalized Method of Moments (GMM) estimator and those of the QML estimator based on the Kalman filter, and (2) close to those of the nonlinear filtering maximum likelihood estimator, which is a computationally intensive method. Chapter 3 provides a testing procedure of a unit root in log-volatility, and presents an empirical example using daily observations on the yen/dollar exchange rate data. Chapter 3 also develops a method of analyzing ARMA(p,q)-SV regression error models using the Markov chain Monte Carlo technique in a Bayesian framework.

Acknowledgments

I owe much to a great number of people who have given me encouragement and support. First of all, I would like to express my deepest gratitude to my academic advisor, Professor Tsunemasa Shiba. He instructed me patiently, offered freely support and encouragement during the course of the study. Moreover, he gave me both materially and mentally excellent research environments.

I would like to thank the faculty of the Institute of Policy and Planning Sciences for their thoughtful criticisms, suggestions, and encouragement. Particularly I wish to mention the following names; Tokuo Iwaisako, Kazuo Kishimoto, and Hitoshi Takehara.

I wish to thank Professors Yoshihisa Baba of Soka University and Tatsuya Kubokawa of the University of Tokyo. Professor Baba guided me in statistics and time series analysis for three years when I was an undergraduate student. Professor Kubokawa advised me how to study Bayesian statistics. They gave me interesting and useful comments for the earlier version of chapter 3.

I am also grateful to Professors Hideki Nakamura of Fukushima University, Teruo Nakatsuma of Hitotsubashi University, Yoshikatsu Tatamitani of the University of Tsukuba, and Akio Nishiura of Soka University, and Toshinori Takayama of the MTB Investment Technology Institute Co., Ltd. for their kindness and friendship, and for useful discussions with them. I also want to thank Toshiyuki Takahashi, an official of Shizuoka prefecture, who is a co-author of the paper on which chapter 2 is based. I was very impressed by his activities while his staying at Tsukuba.

I appreciate financial support for three years from research fellowships of the Japan Society for the Promotion of Science.

Finally, I am pleased to dedicate this work to Dr. Daisaku Ikeda, the founder of Soka University and Soka Schools, for his education about ten years.

Tsukuba

July 1998

Contents

Abbreviations

ADF	augmented Dickey-Fuller
AIC	Akaike information criterion
AR	autoregressive
ARIMA	autoregressive integrated moving average
ARMA	autoregressive moving average
ARCH	autoregressive conditional heteroskedasticity
CD	convergence diagnostic
CIR	Cox, Ingersoll, and Ross
CR	call rate
diag	diagonal matrix
DGP	data generating process
EX	yen/dollar exchange rate
FEVD	forecast error variance decomposition
GARCH	generalized autoregressive conditional heteroskedasticity
GED	generalized error distribution
GMM	generalized method of moments
GSR	generalized continuous short rate model
HJM	Heath, Jarrow, and Morton
$I(d)$	integrated of order d
I_S	indicator function of the set S
IF	importance function
IG	inverted gamma distribution

iid(μ,σ^2)	independently and identically distributed with mean μ and variance σ^2
IPI	industrial production index
IRF	impulse response function
LGARCH-n	log-GARCH approach based on standard normal distribution
LGARCH-t	log-GARCH approach based on standardized t distribution
LN	lognormal distribution
LRT	likelihood ratio statistic
MA	moving average
MCMC	Markov chain Monte Carlo
MM	method of moments
ML	maximum likelihood
NFML	nonlinear filtering maximum likelihood
$N_i(\mu, \Sigma)$	i-variate normal distribution with mean vector μ and covariance matrix Σ
NID(μ,σ^2)	normally and independently distributed with mean μ and variance σ^2
NSE	numerical standard error
PKO	price keeping operation
ME	macroeconomy
QML	quasi-maximum likelihood
RF	remainder function
RMSE	root mean squared error
RMSFE	root mean squared forecast error
SBIC	Schwarz-Bayes information criterion
SM	stock market
SML	simulated maximum likelihood

SV	stochastic volatility
TIBOR	Tokyo interbank offered rate
TIFFE	Tokyo international financial futures exchange
T–Y	Toda and Yamamoto
Uni	uniform distribution
VAR	vector autoregression
VECM	vector error correction model
WN(μ,σ^2)	white noise process with mean μ and variance σ^2
WPI	whole sale price index
WT	Wald statistic

Introduction

The main purposes of this dissertation are: (1) to carry out empirical investigations of nonstationary financial time series based on available econometric techniques, and then (2) to develop time series techniques to test for the nonstationarity of variance. This dissertation consists of three chapters, which correspond to three published papers of the author.

It is well known that many economic time series and financial data possess unit roots[1]. For a unit root process, stationarity can be achieved by a simple differencing. A problem arises if one attempts to fit a regression relating the levels of independent time series, since such an attempt often produces a high value of R^2, combined with a low Durbin-Watson value. The Monte Carlo simulations conducted by Granger and Newbold (1974) suggest that a high value of R^2 or $\bar{R}^2$, combined with a low Durbin-Watson value, is no indication of a true relationship, although such interpretations were made in many applied econometric papers. Granger and Newbold (1974) referred to this phenomenon as the "spurious regression." Phillips (1986) provides an analytical study of linear regressions involving the levels of economic time series. He develops an asymptotic theory for regressions that relate quite general integrated random processes. He showed that the usual t and F ratio test statistics do not possess limiting distributions such as the normal nor the χ^2 distributions. His theoretical results explain many of the simulation findings of Granger and Newbold (1974). The presence of nonstationarity in the time series, therefore, makes it impossible to interpret regression results in the usual way, and thus the conventional econometric inference procedures break down.

Tests for the presence of unit roots are required before estimating econometric models. When a researcher tests a unit root, he has to pay attention to the distribution of the test statistic. Fuller (1976) showed that the asymptotic distributions and rates of convergence for the estimated coefficients of unit root processes differ from those of stationary processes, and thus the conventional t test is invalid. The asymptotic distributions of unit root processes may be described in terms of functionals on Brownian motion. After the work of Dickey and Fuller (1979), a number of approaches to test for unit roots have been proposed; see, *e.g.*, Hamilton (1994, pp.475–543).

A particular class of vector unit root processes is known as a co-integrated process. Once a researcher finds that variables of his interest all possess unit roots, he has to investigate whether a co-integration relationship among variables exists or not. Co-integrated processes were implicit in the "error-correction" models advocated by Davidson *et al.* (1978). A formal development of the key concepts did not come, however, until the work of Granger (1983) and Engle and Granger (1987). For more details, see Hamilton (1994, Chapters 17–20), among others.

Let us look at the nonstationarity from a different angle. Brownian motion is a continuous time version of a random walk process, and is widely applied to term structure models. Models of the term structure are used for pricing and hedging fixed-income securities. In the term structure framework, movements of interest rates or forward rates are modeled in terms of the Ito process. The Ito process contains a stochastic differential equation that is derived by the standard Brownian motion. If a researcher want to estimate parameters of a continuous term structure model, then he needs its

discrete time approximation. Such an approximation becomes an integrated process, and thus nonstationary.

Integrated processes are used not only to model levels of financial time series but also to model variances or volatilities. After the seminal work of Engle (1982), autoregressive conditional heteroskedasticity (ARCH) models have been widely used to model time-varying volatility and the persistence of shocks to volatility. One member of the family of ARCH processes, GARCH(1,1), since its introduction by Bollerslev (1986), has been especially popular in econometric modeling. As an alternative to the GARCH, we may set up a model containing an unobservable volatility component. We take logarithm of the volatility and model it in a linear stochastic process, such as an autoregression. Models of this kind are called the stochastic volatility (SV) models. While conditional volatility in an integrated GARCH model is strict stationary as shown by Nelson (1990), an unobservable volatility in a random walk SV model is nonstationary. Although testing procedures for integration in GARCH models are discussed in Engle and Bollerslev (1986), no testing procedure for random walk SV processes has been proposed.

This dissertation deals with the following three aspects of nonstationarity in economic and financial times series that call for careful attention by the researchers: (1) regression techniques for integrated variables, (2) integrated processes derived as the discrete time approximation of stochastic differential equations for term structure models, and (3) tests for unit roots in time-varying volatility.

An earlier version of Chapter 1 is a paper, Asai and Shiba (1996). After the incredible boom years that started in 1987 came to halt in 1991, the

Japanese stock market has been experiencing down turn since then. One of the dominant theses for this down turn, is the fundamentalist viewpoint. It says that it is the weakness in the Japanese macroeconomy that is to be blamed, and hence not the negative 'bubble' or speculative move that is the culprit of the bear market. The Japanese government, on the other hand, has maintained that by the price keeping operation (PKO) in the stock market, they may put a stop to the market's recession. The two contrasting views on the Japanese stock market, then, need to be empirically tested so that the government can choose appropriate measures to remedy the depressed stock market. Chapter 1, therefore, investigates the relationship between macroeconomic variables, such as the industrial production index, interest rate and inflation rate, and the stock market, using Toda and Yamamoto's (1995) vector autoregressions (VAR) specification. The major findings are: (1) macroeconomic variables do Granger cause the stock market variable, where the reverse causality is ambiguous. (2) The lagged stock market variable affects its own current value but its impact tends to diminish in the long–run. Based upon the findings including the above two, we draw a policy implication that the PKO by the Japanese government would not work, but appropriate macroeconomic policies would benefit not only the real side of the macroeconomy but also the stock market.

A paper, Asai, Takahashi, and Shiba (1998), has been revised and extended so that it can be included as Chapter 2 in this dissertation. In their celebrated paper, Heath, Jarrow, and Morton (HJM) (1992) proposed a continuous time instantaneous forward rate model that gives a general and complete description for variations of term structures of interest rates. In the

paper, they also derived a set of necessary and sufficient conditions for the existence of an equivalent martingale measure under which all discount bonds and associated contingent claims are priced or valued. Although the HJM model is not meant to be used to forecast interest rates, it is used for pricing bonds in many financial institutions, in practice. Hence, there is a need to compare the performances of various different interest rate forecasting schemes including the HJM model. In- and out-of-sample forecasting performances of the HJM model, a short rate model that nests popular spot rate models[2], ARIMA, VAR, and a prediction method based on the *SABL*[3], have been compared using the daily three month Euroyen interest futures data. Predictive performances of the forecasts from the HJM model resemble to those of the other models, except for the *SABL*. *SABL*'s performance is sometimes excellent, but occasionally very poor. It seems that such contrasting differences in the *SABL* performance are due to non-liner trend components' changing behavior.

Chapter 3 is based upon a paper, Asai (1998). A wide variety of the GARCH class models and the SV models are available for modeling changes in asset return variance or volatility over time. The log-GARCH models are the logarithmic extension of the GARCH models. While the GARCH models are popular and easily estimated, it is well recognized that the SV models are not easy to estimate. It should be noted that the SV models are more general and therefore preferable in several respects, compared to the GARCH models. Chapter 3 derives a log-GARCH representation of a class of SV models, including the ARMA-SV model, and analyzes the finite sample properties of a Quasi-Maximum Likelihood (QML) estimator.

The Monte Carlo results indicate that their finite sample properties are (1) superior to those of the Generalized Method of Moments (GMM) estimator and those of the QML estimator based on the Kalman filter, and (2) close to those of nonlinear filtering maximum likelihood (NFML) estimator, which is a computationally intensive method. Chapter 3 also provides a testing procedure of a unit root in log-volatility, and presents an empirical example using daily observations on the yen/dollar exchange rate data.

Monte Carlo simulations in Chapter 3 suggest that some computer intensive methods performs better compared to the log-GARCH approach when researchers can afford to neglect computational costs. In this sense, appendix 3.G develops a method of analyzing ARMA(p,q)-SV regression error models using the Markov chain Monte Carlo technique in a Bayesian framework. An empirical study that uses the Bayesian MCMC technique on the daily yen/dollar exchange rate is also provided in the appendix.

Footnotes

*1 For example, Takeuchi (1991) has found that many Japanese macroeconomic time series possess unit roots. Baba (1995) and Yoshida and Rasche (1990), among others, empirically tested and found unit roots in Japanese financial data.

*2 These models include the Vasicek model, the Cox-Ingersoll-Ross class models, the Merton model, the Dothan model, the Brennan-Schwartz model, the constant elasticity of variance process, and the geometric Brownian motion model.

*3 This is a popular seasonal decomposition method, found in such statistics software as the *S-PLUS*.

Chapter 1
Relationship between the Stock Market and the Macroeconomy: VAR Investigation

1.1 Introduction

What are the relevant factors that would explain the variation in stock prices and returns? In recent papers such as Keim and Stambaugh (1986), Campbell and Shiller (1988), Fama and French (1988) and Cutler, Poterba, and Summers (1991), the authors report that variables such as dividend yield and price–earning ratio explain 25 percent or more of the variation in stock returns. Further, Balvers, Cosimano, and MacDonald (1990), Schwert (1990), and Fama (1990) present evidence that economic indicators such as industrial production index also have power predicting stock returns.

Stock price has been considered as the leading indicator of macroeconomy. For instance, if we use the dividend discount model, the fundamental value of the stocks may be regarded as a function of the past macroeconomic variables since they constitute information sets to generate flow of expected future income[1]. A typical view that says that the stock market has been the most sensitive indicator of the business cycle, can be found *e.g.,* in Siegel (1991). Cochrane (1991), on the other hand, has constructed and tested a model that GNP growth, among other variables, explains variations in stock returns. Although, he tested and found evidence in favor of the reverse relationship too.

Kitasaka (1995) used near-VAR system, which is the VAR system restricting some parameters to be zero, to test the causality among the Japanese stock markets and the macroeconomic variables of Japan and U.S. His result indicates that no causality running from the Japanese macroeconomic variables to the stock market. His work is, however, open to criticism since he seems to have misused his test statistics. The tests he used are valid when the variables in the VAR are stationary, whereas his VAR system is nonstationary.

In this chapter we investigate the relationship between the stock market (**SM**) and the macroeconomy (**ME**) in Japan. After incredible boom years that started in 1987 came to a halt in 1991, Japanese stock market has been experiencing tremendous down turn ever since then. The "bubble" years are over and there is slight recovery in sight, as of this writing. One of the dominant theses for this down turn, is the fundamentalist viewpoint. It says that it is the weakness in the Japanese macroeconomy that is to be blamed, and hence not the negative "bubble" or speculative move that is the culprit of the bear market. This view, in turn, implies that the direction of the stock market can only be reversed by appropriate economic policies that would strengthen the macroeconomy.

The Japanese government, on the other hand, has maintained that by the price keeping operation (so-called P.K.O.) in the stock market, they may put a stop to the market's recession[2]. This view puts credentials to stock market's self–mending power, to put itself into the recovery path by its own power with help from the P.K.O.

The two contrasting views on the Japanese stock market, then, need to

be empirically tested so that the government can choose appropriate measures to remedy the depressed stock market, Let us list the following several hypotheses that need to be investigated in this chapter:

1. Is there a causal relationship between the stock market and the macroeconomy, causality going from **ME** to **SM** (*i.e.* ME$\longrightarrow$SM)? This tests the fundamentalists' view.

2. The reverse of (1), *i.e.,* **ME**$\longleftarrow$**SM**? This is a view that stock market's performance precedes that of the macroeconomy, since **SM** uses predictions of the future macroeconomic variables. In other words, the market's performance can be considered as signaling the future performance of **ME**.

3. Can **SM** be regarded as exogenous? In other words, does the past **SM** variable have explanatory power in explaining the current **SM** performance[3]? If this is supported, then government's positive intervention into **SM** may be justified. A further question we want to ask is: is the relationship a short–run or long–term phenomenon? This is an interesting question to ask, since if it is only a short–run phenomenon, then the above stated government's intervention would only have a passing effect, while if it is a long–run phenomenon then its effect may not be visible in the short–run but eventually be felt in the long–run.

There has been a great deal of discussions empirically as to what, if any, relationship there is between the stock market and the macroeconomy, in particular the real side. It is well known that the traditional IS-LM framework is not really a suitable theoretical tool for answering such a question.

Among the attempts to establish a theoretical link between the stock market and the macroeconomy, Blanchard's (1981) model seems to be worth looking into, because the model is based on a traditional framework and it incorporates Tobin's (1969) q-theory. Let us call Blanchard's (1981) model the Tobin–Blanchard model. The Tobin–Blanchard model emphasizes the interaction among asset values, interest rates, and output.

Among the macroeconomic variables, we select the variables used in the Tobin–Blanchard model since there is no other satisfactory theory that describes explicitly relationships between the stock market and the macroeconomy. Considering the existing state of the Japanese macroeconomy and sample periods, we add price index and exchange rates to above three variables. Therefore our VAR system consists of five variables: the stock market variable TOPIX, the income variable, interest rates, price index, and exchange rates. It should be emphasized that the Tobin–Blanchard model is described as structural form equations, and thus time series techniques can not be applied to testing the validity of the model.

The remainder of this chapter is organized as follows. Section 1.2 discusses the econometric methodology used in this chapter. Section 1.3 presents testing results of the unit roots and co-integrations. The VAR results, including the various estimation and hypothesis testing results, are given in Section 1.4. Section 1.5 concludes this chapter.

1.2 Econometric Methodology

This section outlines the econometric methodology used in this chapter. We use the VAR, vector autoregressions, specification to investigate various different causal relationships cited in the previous section.

Prior to conducting on VAR analysis, we need to investigate whether the data to be used in the VAR analysis are stationary or not. If the dth difference of a time series $\{X_t\}$, $\Delta^d X_t, d > 0$ is stationary, then it is said to be integrated of order d, $\{X_t\} \sim I(d)$. If a linear combination of integrated time series is stationary, then they are said to be co-integrated. The vector of weights in the co-integrated process is called the co-integrating vector. If the variables that we are interested in are integrated, then whether a co-integrating vector exists or not ought to be tested. Textbooks such as Banerjee *et al.* (1993) and Hamilton (1994), among others, provide detailed discussion and surveys about tests for unit roots and co-integrations, and estimation of co-integrating vectors.

It should be emphasized that these tests suffer from low power and/or size distortions. Moreover, since usually only small sample sizes are available, a slight change in sample period could result in different test conclusions. In other words, we are aware that unit root tests are easily implemented and that there are established ways to estimate cointegrating vectors, but at the same time we need to note that the tests and estimation methods are known to be not robust to small sample sizes, in particular. But these considerations themselves are preliminary to a further question.

If time series are found to be $I(1)$ say, then there may be two possible

routes that we could take:

1. If no co-integrating relationship can be found, then the usual practice is to specify the VAR in terms of first differenced variable, ΔX_t.

2. If a co-integration test does not reject the null hypothesis of co-integration, we may specify a vector error correction VAR model (VECM).

It should be noted that the VECM is not the only representation when the variables are co-integrated; see Watson (1994, p.2871).

The VAR in ΔX_t form, however, has certain drawbacks, when used in economic analysis. While it is not wrong from a statistics viewpoint to estimate the parameters of such a system, the most that we can find with such a system, is less than the first order relationship in the variables involved. Note that if $\{X_t\} \sim I(1)$ then $\Delta X_t \sim$ stationary, and this "stationary" component is composed of non–$\{X_t\}$ or an error term process.

One way to test causal relationships is to employ the Johansen-type VECM, which was proposed by Toda and Phillips (1993). This may be a natural way to proceed after we conducted Johansen's (1998, 1991) co-integration tests, since the test statistic is dependent on the outcomes of co-integration test. For reasons already stated, Toda and Phillips' (1993) procedure is unsatisfactory. Alternative procedures are the fully modified VAR approach proposed by Phillips (1995) and the lag-augmented VAR approach proposed by Toda and Yamamoto (1995) (T–Y here after). These procedures are designed to be robust to integration/co-integration properties of the time series, and therefore can be applied without *a priori* knowledge

of the presence/absence of unit roots. Yamada and Toda (1997, 1998) investigated finite sample properties of these test statistics for Granger causality, and found that test statistics based on Phillips (1995) suffer from size distortion and that tests by T–Y produce low power[4].

We use the results of T–Y to model the VAR for three reasons: (1) a test with uncontrolled size is not useful, even if it has reasonably high power. (2) As it is described later in this section, T–Y's procedure is very easy to implement. (3) since T–Y's VAR system is modeled in levels, *i.e.,* without differencing to achieve stationarity in each variable, interpretation of the test results is easily made.

The VAR system proposed by T–Y is modeled possibly in levels, but augmenting the lag length by *a priori* set maximum order of integration. They showed, albeit loss in efficiency due to the artificial lag length augmentation, standard asymptotic theory is applicable to such a system[5]. If T–Y's specification is used then the standard χ^2 based block exogeneity tests, *i.e.* Granger causality tests, can be performed, and after the Cholesky decomposition of the error term variance matrix, we may compute the forecast error variance decomposition (FEVD) and the impulse response function (IRF)[6].

In the following we shall introduce T–Y's results in more detail. Their most salient conclusion is: if the order of integration of the process does not exceed the true lag length of the model, we can apply the usual lag selection procedure to a possibly integrated or co-integrated VAR since the standard asymptotic theory is valid. Let us consider the following VAR system which

is T–Y's equation (24), p.242.

$$y_t = \gamma_0 + \gamma_1 t + J_1 y_{t-1} + \ldots + J_k y_{t-k} + \ldots + J_p y_{t-p} + \varepsilon_t \qquad (1.1)$$

where $J_{k+1} = \ldots = J_p = 0$ $(p \geq k+1)$. Suppose that we wish to test the hypothesis

$$H_0 : J_{m+1} = \ldots = J_p = 0 \qquad (1.2)$$

where $k \leq m \leq p-1$, in the estimated system

$$Y' = \hat{\Gamma}\mathcal{T}' + \hat{\Psi}Z' + \hat{\Phi}X' + \hat{\mathcal{E}}'$$

where

$$y_t = \begin{pmatrix} y_{1t} \\ \vdots \\ y_{nt} \end{pmatrix}, \; z_t = \begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-m} \end{pmatrix}, \; x_t = \begin{pmatrix} y_{t-m-1} \\ \vdots \\ y_{t-p} \end{pmatrix}, \; \mathcal{T}' = \begin{pmatrix} 1 & \ldots & 1 \\ 1 & \ldots & T \end{pmatrix},$$

$$Y' = (y_1, ..., y_T), \; Z' = (z_1, ..., z_T), \; X' = (x_1, ..., x_T), \; \hat{\mathcal{E}}' = (\hat{\varepsilon}_1, ..., \hat{\varepsilon}_T),$$

$$\hat{\Gamma} = (\hat{\gamma}_0, \hat{\gamma}_1), \; \hat{\Psi} = (\hat{J}_1, ..., \hat{J}_m), \text{ and } \hat{\Phi} = (\hat{J}_{m+1}, ..., \hat{J}_p).$$

According to T–Y, the Wald statistic, W, to test the hypothesis (1.2)

$$W = \hat{\phi}'[(X'QX)^{-1} \otimes \hat{\Sigma}_\varepsilon]^{-1}\hat{\phi} \qquad (1.3)$$

where

$$\hat{\phi} = vec(\hat{\Phi}), \quad \hat{\Sigma}_\varepsilon = T^{-1}\hat{\mathcal{E}}'\hat{\mathcal{E}},$$

$$Q = Q_{\mathcal{T}} - Q_{\mathcal{T}} Z (Z' Q_{\mathcal{T}} Z)^{-1} Z' Q_{\mathcal{T}},$$

and

$$Q_{\mathcal{T}} = I_T - \mathcal{T}(\mathcal{T}'\mathcal{T})^{-1}\mathcal{T},$$

has an asymptotic χ^2 distribution with $n^2(p-m)$ degrees of freedom under the null hypothesis (1.2) if $m \geq d_{\max}$ where d_{max} is the maximal order of integration that we suspect might occur in the process. For notational convenience, we use $vec(M)$ to stack the columns of a matrix M into a column vector while T–Y use it to stack the rows.

Having determined the lag length k, we estimate a $(k + d_{max})$ th order VAR. The coefficient matrices of the last d_{max} lagged vectors in the model may be ignored since these are regarded as zeros, and we can test linear or nonlinear restrictions on the first k coefficient matrices using the standard asymptotic theory. We restrict our attention to the case of $d_{max} = 2$ because we expect most economic time series encountered in empirical studies to be at most $I(2)$. Suppose we wish to test the linear restriction

$$H_0 : R\phi = 0. \tag{1.4}$$

We may construct the Wald statistic $W^\dagger$

$$W^\dagger = [R\hat{\phi}]'[R\{(X'QX)^{-1} \otimes \hat{\Sigma}_\varepsilon\}R]^{-1}[R\hat{\phi}] \tag{1.5}$$

where

$$x_t = \begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-m} \end{pmatrix}, \; z_t = \begin{pmatrix} y_{t-m-1} \\ \vdots \\ y_{t-m-2} \end{pmatrix}, \; \hat{\Phi} = (\hat{J}_1, ..., \hat{J}_m), \text{ and } \hat{\Psi} = (\hat{J}_{m+1}, \hat{J}_{m+2}),$$

has an asymptotic χ^2 distribution with m degrees of freedom under the null hypothesis (1.4) if $p \geq k + 2$.

1.3 The Data

As noted in Section 1.1, we used 19 years worth of monthly data from January 1976 to December 1994 on the industrial production index (IPI), the dollar/yen exchange rate (EX), the call rate (CR), the whole sale price index (WPI), and the TOPIX[7]. IPI, WPI and TOPIX have been seasonally adjusted by *SABL*, a seasonal adjustment method available on a software *S-PLUS*. *SABL* has been successfully used in stock price prediction, among others; see Shiba and Takeji (1994). In some models differenced series, $\Delta X_t = \ln X_t - \ln X_{t-1}$, have been used on IPI, WPI and the TOPIX.

1.3.1 Tests for Unit Roots

We first investigate the integration order of our individual time series using some well-known unit root test including augmented Dickey–Fuller (ADF) tests and Phillips and Perron's (1988) non-parametrically modified Dickey–Fuller test. The ADF test is based on the following regression model

$$\Delta y_t = \alpha + \mu t + \rho y_{t-1} + \sum_{i=1}^{p} \rho_i \Delta y_{t-i} + \varepsilon_t, \tag{1.6}$$

where y_t is a scalar in this section. α and μt are the drift and trend terms, respectively. If the data generating process (DGP) is $\alpha = \mu = 0, \rho = 1$ and equation (1.6) is estimated, then the estimate of ρ does not have the normal distribution asymptotically, but it follows the distribution which Dickey and Fuller derived. In practice we must choose the appropriate statistics for testing $\rho = 0$ depending on whether constant or trend terms are included in the regression. If $\mu \neq 0$ in the DGP, the t-ratio of ρ has the normal distribution. Similarly, if $\mu = 0, \alpha \neq 0$ in the DGP and we estimate the

regression with a constant term but without a trend term, then the t-ratio of ρ has the normal distribution.

Dolado, Jenkinson, and Sosvilla-Rivero (1990) suggest a procedure to test for a unit root when the form of the DGP is unknown. Their method is as follows: start with the least restrictive model that will generally include a trend term and a drift, *i.e.* equation (1.6), and use the ADF test statistic. If the null hypothesis of a unit root is rejected, there is no need to proceed. Conclude that $\{y_t\}$ sequence does not contain a unit root. If the null hypothesis is not rejected, test for the significance of the trend term under the null of a unit root. If the trend is significant, retest for the presence of a unit root using the normal distribution. If the trend is not significant, estimate equation (1.6) without the trend and use the ADF test statistic to test for the presence of a unit root. If the null hypothesis of a unit root is not rejected, test for the significance of the constant. If the drift is significant, retest for the presence of a unit root using the normal distribution. If the drift is not significant, estimate equation (1.6) without the trend or drift, then test for the presence of a unit root.

Instead of including lagged dependent variable in the right hand side to circumvent the possible serial correlation in the error terms, Phillips and Perron (1988) made a non-parametric correction for a serial correlation. In this approach, we first calculate the Dickey–Fuller unit root tests from regression equation (1.6) assuming $p = 0$. The statistics are then transformed to remove the effects of serial correlation on the asymptotic distribution of the test statistic. We used the Newey and West (1987) method to construct an estimate of the error variance from the estimated residuals. For both

ADF–type and Phillips–Perron type statistics, we start with the maximal lag length 15, and gradually decreased the lag length until the coefficient on the last included lag is significant by the t test; see Campbell and Perron (1991).

We use the following notation to present the results of ADF tests and Phillips–Perron tests. τ represents the τ test, *i.e.* t ratio test for $\rho = 0$, based on equation (1.6) without trend term and τ_μ is the τ test based on equation (1.6) with trend term. Φ is the one sided F type statistic testing the null hypothesis of $(\alpha, \rho) = (0, 0)$ in equation (1.6) without the trend and Φ_τ is the one sided F type statistic testing the null of $(\mu, \rho) = (0, 0)$ in equation (1.6) with the trend. The Z versions reported in Table 1.2 have the same interpretation but the Phillips–Perron non–parametric correction has been used instead of the lag augmentation.

As shown in Table 1.1, for each five series (in levels) the null of a unit root is not rejected at ten percent significant level, but after taking the first differences, the null is rejected.

In Table 1.2, $Z(\Phi)$ of IPI and five differenced variables reject the null at ten percent significant level. Applying Dolado, Jenkinson and Sosvilla-Rivero's procedure for IPI, the trend term is not significant under the null of a unit root, but the drift is significant. The value of $Z(\tau)$ of IPI is -2.2663 and we may reject the null using the normal distribution.

It must be noted, however, that Phillips–Perron type test are more likely to reject the null of a unit root, whether it is false or not; in the presence of negative moving average (MA) terms, the Phillips–Perron test tends to reject

Table 1.1: ADF Unit Root Test

	p	Φ_μ	τ_μ	Φ	τ
TOPIX	9	1.129	-1.050	1.867	-1.426
IPI	15	3.657	-2.461	2.892	-1.688
CR	3	3.505	-2.620	2.487	-2.174
EX	3	2.812	-2.345	2.308	-1.369
WPI	12	2.891	-2.050	1.961	-1.975
ΔTOPIX	7	7.248*	-3.802*	6.671*	-3.651*
ΔIPI	15	6.689*	-3.655*	5.515*	-3.306*
ΔCR	2	18.62*	-6.103*	18.62*	-6.103*
ΔEX	12	8.056*	-4.013*	8.080*	-4.019*
ΔWPI	2	7.991*	-3.993*	7.454*	-3.858*

Note: '*' indicate that it is significant at the ten percent level.

Table 1.2: Phillips–Perron Unit Root Test

	$Z(\Phi_\mu)$	$Z(\tau_\mu)$	$Z(\Phi)$	$Z(\tau)$
TOPIX	0.940	−0.650	2.928	−1.369
IPI	3.148	−1.898	7.109*	−2.266
CR	1.965	−1.900	1.251	−1.453
EX	2.159	−2.049	2.596	−1.217
WPI	3.074	−1.754	1.704	−1.817
ΔTOPIX	64.083*	−11.319*	63.694*	−11.285*
ΔIPI	233.65*	−21.595*	224.25*	−21.160*
ΔCR	64.987*	−11.401*	65.028*	−11.404*
ΔEX	96.791*	−13.901*	97.065*	−13.925*
ΔWPI	21.429*	−6.553*	19.518*	−6.253*

Note: '*' indicate that it is significant at the ten percent level.

the null whether or not the actual DGP contains negative MA components. Phillips and Perron (1988) present simulation evidence regarding the power of the Phillips–Perron type test. The DGP is taken to be

$$\begin{aligned} y_t &= \rho y_{t-1} + u_t, \\ u_t &= \varepsilon_t + \theta\varepsilon_{t-1},\ \varepsilon_t \sim \text{iid}(0, \sigma_\varepsilon^2);\ y_0 = 0. \end{aligned}$$

The Phillips–Perron type test generally has high power, but suffers substantial size distortions for $\theta < 0$, in sample sizes typically found in economics. See Banerjee *et al.* (1993) for more details.

The sample autocorrelation and the sample partial autocorrelation of the residuals for IPI decrease with an oscillation starting from the negative region. The residuals probably have negative MA components and this shows it is preferable to use the ADF test in this case. We conclude IPI has a unit root, supported with the ADF test's result. we hence treat our five variables as $I(1)$ hereafter.

1.3.2 Tests for Co–Integration

Once we have been unable to reject the null hypothesis of a unit root, we can build multivariate models that will enable us to investigate the absence or presence of co-integration relationships in our data set. We use the Engle–Granger (1987) test and the Johansen (1988, 1991) test.

Table 1.3 shows results of the Engle–Granger tests between two variables. For instance, using the residual from regressing "TOPIX" on the constant and "IPI", we calculated that the value of τ statistic under the null of no co-

Table 1.3: Engle–Granger Co-Integration Test

	TOPIX	IPI	CR	EX	WPI
TOPIX		−2.289	−1.677	−1.947	−1.355
IPI	−2.197		−1.804	−2.413	−1.028
CR	−2.730	−3.048*		−2.794	−2.582
EX	−2.115	−2.278	−2.009		−2.096
WPI	−1.886	−1.833	−2.048	−2.212	

Note: '*' indicate that it is significant at the ten percent level.

integration relationship is -2.2892. Except for the case of CR on IPI, Table 1.3 shows that there is no bivariate co-integration relationship.

Using the Johansen test, we investigate the co-integration in systems of equations, while we dealt with the bivariate co-integration regression using the Engle–Granger test. The trace statistic tests the null hypothesis that the number of distinct co-integrating vector is less than or equal to r against a general alternative. The maximal eigenvalue statistic tests the null that the number of co-integrating vector is r against the alternative of $r+1$. Note that the presence of the constant terms in the VAR system will effect the distribution of the test statistics, which is analogous to the effect of the presence of the constant and trend in testing a unit root. We use critical values provided by Osterwald-Lenum (1992).

Table 1.4 shows the result of the Johansen test using the estimated twelfth order VAR including five $I(1)$ variables, TOPIX, IPI, CR, EX, WPI, and the constant. We reject the null of $r = 0$ from the results of the trace test

and the maximal eigenvalue test. We therefore concluded that the number of co-integrating vector is one in this case. Table 1.5 shows the estimated normalized co-integrating vector and the estimated adjustment coefficient in each equation.

1.4 VAR Results

As mentioned in Section 1.2, we use the T–Y result to model the VAR system instead of the VECM. Let us present the lag length selection results. We estimated three models using equation (1.1) specification of the VAR:

1. model A: (TOPIX, ΔIPI, CR, EX, WPI)
2. model B: (TOPIX, ΔIPI, CR, EX, ΔWPI)
3. model C: (ΔTOPIX, ΔIPI, CR, EX, ΔWPI)

For models A and B, we set the maximum lag length l_{max} to be a year and a half, *i.e.*, $l_{max} = 18$, while for model C we set $l_{max} = 12$, one year. We would think such l_{max}'s are reasonable in view of the way variables are transformed in each model. The null hypothesis we wish to test is given by $H_0 : l = p - 1$, while the alternative hypothesis by $H_1 : l = p$, where l is the lag length. The Wald χ^2 test statistic as given in T–Y is distributed under H_0 as χ^2 with its degrees of freedom 25, $\chi^2(25)$, since there are five variables in the system. Table 1.6 gives the three models' computed Wald statistic along with their probability values. Results in Table 1.6 indicate that the chosen lag length, p^*, is 16 for models A and B, while 9 for model C. We have computed such

Table 1.4: Johansen Co–Integration Test

null hypothesis	trace	maximal-eigenvalue
$r \leq 4$	7.698	7.698
$r \leq 3$	17.303	9.605
$r \leq 2$	32.632	15.329
$r \leq 1$	55.048*	22.416
$r = 0$	92.890*	37.841*
eigenvalues: 0.037, 0.046, 0.073, 0.105, 0.170		

Notes: r denotes the rank of co-integration. '*' indicate that it is significant at the five percent level.

Table 1.5: Normalized Co–Integrating Vector and Adjustment Coefficients

	TOPIX	IPI	CR	EX	WPI
co-integrating vector	1.000	–0.597	3.236	6.668	13.49
adjustment coefficient	0.000551	–0.0000555	–0.000138	–0.000747	–0.0000364

Table 1.6: Lag Length Selection Wald Test (WT)

H_0 : pth lag coefficients=0				
p	18	17	16	
Model A	34.956	22.493	62.956	
prob–value	(8.9)	(60.7)	(0)	
p	18	17	16	
Model B	22.584	30.344	61.492	
prob–value	(60.2)	(21.2)	(0)	
p	12	11	10	9
Model C	32.532	32.702	33.144	40.817
prob–value	(14.4)	(13.9)	(12.8)	(2.4)

Notes: WT denotes the Wald test statistic as given in T–Y's paper. In all models, the figure in the last entry or column is significant at the five percent level. The critical value is given by $\chi^2_{0.05}(25) = 37.7$. 'prob–value' denotes probability values $\times 100$.

lag length selection criteria as the AIC and SBIC, but they tend to select very low order lag lengths such as 1 or 2.

We assumed d_{max}, the maximum order of integration, to be at most 2 for all three models, hence our chosen VAR is given by VAR(p^*+2). VAR(p^*+2) for the three models was put to block exogeneity test. Two test statistics have been computed: the Wald statistic (WT) as given in T–Y, and the usual degrees of freedom adjusted likelihood ratio (LRT) statistic. Since the LRT results are very close to that of the WT, we only present the WT results.

They are in three matrices with respective prob-values $\times 100$ attached inside the parenthesis. See Tables 1.7 to 1.9.

We note that in Tables 1.7 to 1.9, results are qualitatively very similar. To see the qualitative similarity in the tables, let us denote the stock market by **SM** and the macroeconomy block by **ME**. Then we notice that in all three models, (1) at least one of the variables in **ME** causes **SM**, *i.e.* **ME**$\longrightarrow$**SM**, but (2) **SM** does not cause any variable in the **ME** block. From these observations, we decided to concentrate on model A in the rest of this chapter to save space and to make this research more clearly focused. We have also computed all the results for models B and C, and they are available from the author upon request.

In the usual VAR analysis, forecast error variance decomposition (FEVD) and impulse response function (IRF) tend to give markedly different results depending on the way variables are arranged in the Cholesky decomposition. The order that variables are arranged, in a sense, implies a particular restriction that a researcher imposes on the VAR[8]. We, however, followed the traditional approach of the Cholesky decomposition since we observed that the results we obtained are robust to the different orderings. Table 1.10 gives FEVD of TOPIX and Table 1.11 presents that of ΔIPI, both due to the five variables in the system. The numbers in the tables indicate: for instance, in Table 1.10, "73.308" percent of "h=36" months ahead forecast error variance is due to TOPIX itself. The ordering of the variables are indicated in the top row, *i.e.,* EX, CR, WPI, TOPIX and then in the last position, ΔIPI. A lower triangular Cholesky decomposition is used.

Table 1.7: Model A Block Exogeneity WT

	TOPIX	ΔIPI	CR	EX	WPI
TOPIX	6070*	14.114	36.925*	20.939	30.220*
	(0)	(59.0)	(0.1)	(18.1)	(1.7)
ΔIPI	19.710	181*	21.891	28.405*	13.317
	(23.4)	(0)	(14.7)	(2.8)	(64.9)
CR	17.156	17.140	1152*	40.823*	83.398*
	(37.6)	(37.7)	(0)	(0.1)	(0)
EX	20.491	39.237*	38.283*	2122*	26.978*
	(19.9)	(0.1)	(0.1)	(0)	(4.2)
WPI	18.489	54.078*	59.881*	97.726*	3350*
	(29.6)	(0)	(0)	(0)	(0)

Notes: Figures in the parentheses are prob–values $\times 100$. '*' indicate that it is significant at the five percent level. The critical value is given by $\chi^2_{0.05}(16) = 26.3$. Lag length is $p = 16$.

Table 1.8: Model B Block Exogeneity WT

	TOPIX	ΔIPI	CR	EX	ΔWPI
TOPIX	6380*	12.727	26.544*	17.768	21.276
	(0)	(69.3)	(4.7)	(33.8)	(16.9)
ΔIPI	19.832	175*	26.175	26.140	11.956
	(32.8)	(0)	(5.2)	(5.2)	(74.3)
CR	17.077	16.658	2594*	50.101*	66.493*
	(38.1)	(40.8)	(0)	(0)	(0)
EX	20.563	38.547*	39.147*	2911*	31.671*
	(19.6)	(0.1)	(0.1)	(0)	(1.1)
ΔWPI	18.123	37.012*	30.865*	86.999*	48.9*
	(31.7)	(0.2)	(1.4)	(0)	(0)

Notes: Figures in the parentheses are prob–values $\times 100$. '*' indicate that it is significant at the five percent level. The critical value is given by $\chi^2_{0.05}(16) = 26.3$. Lag length is $p = 16$.

Table 1.9: Model C Block Exogeneity WT

	ΔTOPIX	ΔIPI	CR	EX	ΔWPI
ΔTOPIX	35.79*	7.794	17.003*	7.510	10.755
	(0)	(55.5)	(4.9)	(58.4)	(29.3)
ΔIPI	8.391	106*	18.474*	3.536	4.509
	(49.5)	(0)	(3.0)	(93.9)	(87.5)
CR	4.044	5.741	2149*	14.956	49.601*
	(90.8)	(76.6)	(0)	(9.2)	(0)
EX	8.687	6.887	9.265	1591*	9.802
	(46.7)	(64.9)	(41.3)	(0)	(36.7)
ΔWPI	10.272	16.057	12.729	88.135*	41.6*
	(32.9)	(6.6)	(17.5)	(0)	(0)

Notes: Figures in the parentheses are prob–values $\times 100$. '*' indicate that it is significant at the five percent level. The critical value is given by $\chi^2_{0.05}(9) = 16.9$. Lag length is $p = 9$.

Table 1.10: Forecast Error Variance Decomposition of TOPIX

h	EX	CR	WPI	TOPIX	ΔIPI
3	0.453	1.031	0.067	98.426	0.023
6	0.497	4.623	0.860	93.914	0.106
12	0.436	6.356	3.852	88.859	0.498
24	2.186	8.661	7.479	75.450	6.225
36	1.466	10.891	6.751	73.308	7.584

Notes: This table gives the FEVD result of TOPIX based on model A. The numbers in the tables indicate: for instance "73.308" percent of "h=36" months ahead forecast error variance is due to TOPIX itself. The ordering of the variables are indicated in the top row, *i.e.*, EX, CR, WPI, TOPIX and then in the last position, ΔIPI. A lower triangular Cholesky decomposition is used.

Table 1.11: Forecast Error Variance Decomposition of ΔIPI

h	EX	CR	WPI	TOPIX	ΔIPI
3	0.844	0.554	2.258	1.348	94.997
6	1.625	2.392	3.693	4.424	87.865
12	4.680	2.641	4.981	4.394	83.304
24	5.687	3.385	5.457	6.942	78.530
36	5.508	3.575	5.970	7.472	77.475

Table 1.12: Forecast Error Variance Decomposition of EX

h	EX	CR	WPI	TOPIX	ΔIPI
3	94.666	0.485	1.224	3.589	0.037
6	88.409	0.850	5.456	5.198	0.088
12	71.528	2.717	15.044	3.439	7.271
24	43.018	1.802	19.666	3.110	32.403
36	40.035	3.669	18.767	5.141	32.387

Table 1.13: Forecast Error Variance Decomposition of CR

h	EX	CR	WPI	TOPIX	ΔIPI
3	5.572	83.541	10.760	0.033	0.095
6	11.773	66.274	19.675	0.083	2.195
12	25.395	50.399	14.551	0.232	9.422
24	23.118	36.392	15.733	4.191	20.566
36	29.770	28.514	14.616	4.820	22.280

Table 1.14: Forecast Error Variance Decomposition of WPI

h	EX	CR	WPI	TOPIX	ΔIPI
3	39.009	0.942	57.221	1.028	1.801
6	50.417	1.256	40.446	4.038	3.844
12	56.813	1.318	23.701	2.609	15.559
24	28.009	9.576	9.117	8.132	45.166
36	25.909	12.621	10.654	19.947	30.869

We observe that FEVD's of TOPIX and ΔIPI indicate the two variables have strong exogenous element in them. For instance even after three years, TOPIX has 25 percent of its FEV accounted by itself, *i.e.*, FEV(TOPIX)$_{TOPIX,\ h=36}$=.25. But on the other hand, FEV(TOPIX)$_{ME,\ h=24}$=.23 increasing from FEV(TOPIX)$_{ME,\ h=1}$=.01, where ME denotes non stock market factors combined. In Tables 1.12 to 1.14, we present FEVD's of the rest of the variables in the system. After h=36 months, these three variables exhibit far less exogenous element in them; FEV(EX)$_{EX,\ h=36}$=.40 to FEV(WPI)$_{WPI,\ h=36}$=.11. Then the question is: what accounts for these variables' FEV after $h = 36$? It is interesting to see that the bulk of FEV after $h = 36$ is accounted by ΔIPI, and not by TOPIX. For instance FEV(CR)$_{TOPIX,\ h=36}$=.048 while FEV(CR)$_{\Delta IPI,\ h=36}$=.22.

1.5 Concluding Remarks

In this chapter we began with the question: what ought to be the right choice of measures that the Japanese government can take so that the slumping stock market, **SM**, can reverse its direction? We use Toda and Yamamoto's (1995) VAR specification to investigate the above question.

Using the block exogeneity test, we found **ME**$\longrightarrow$**SM**, where **ME** macroeconomic factors combined, but not **SM**$\longrightarrow$**ME**. This may indicate that if the government selects an appropriate set of macroeconomic policies, then that would stimulate **SM** also. At the same time, **SM** is not a good indicator of the future macroeconomic performance, as it is generally believed.

According to the FEVD and the IRF results, **SM** is pretty much exogenous. This implies that if the government can identify and use a set of feasible policies to intervene the stock market, then it should be worthwhile to do so. However, such policies should be "surprising" to the market participants in order to be effective, since what the FEVD does, is to decompose "forecast error" that is an unanticipated variable. The magnitude of **SM**'s exogeneity, however, is seen to diminish in the long–run after it reaches a plateau. This fact, in addition to the above observation on the difficulty of identifying correct "surprising" policies, may indicate that such policies as the PKO are not recommended. Such direct intervention into the stock market could be effective in the short–run, but it will likely to become ineffective in the long–run. What needs to be implemented, is not the PKO type policy, but a set of policies that would fundamentally strengthen the macroeconomy.

Footnotes

*1 The term "fundamentals" is usually meant to represent the importance of such variables as dividend yield and price–earning ratio, in explaining the variation in stock returns (see, *e.g.*, Cutler *et al.*, 1991), and not of the macroeconomic variables. Note, however, that such financial statement variables in turn are related to the macroeconomic variables.

*2 The word "P.K.O." is obviously coined from the word "P.K.O." and "P.K.F." of the United Nations. Just around the same time as the Japanese government was contemplating to send men to Cambodia, the government was in a position to decide whether to actively put moneys into the slamming stock market or not. Ordinarily, the term P.K.O. implies a set of government action to strengthen the stock market by (1) changing rule of the game in the market, *i.e.*, modifying institutional aspects of the stock market so that stock price may be prevented from slamming, and (2) buying up stocks using governmental agencies' such as postal savings, funds. Sometimes, the government even announces deliberately optimistic predictions that they hope to be self–fulfilling.

*3 There have been numerous studies that conclude variations in bonds and stocks can be explained by such financial statement variables as the dividend yield, term spread and so on. These terms are defined in Fama and French (1989). They also survey and empirically investigate the importance of the financial variables to explain movements in bonds and stocks. Young *et al.* (1991) find some financial statement variables' predictable power to be significant.

*4 Stock (1995) also pointed out the possibility that T–Y's approach produces poor power in some situations. Recently, Kurozumi and Yamamoto (1997) suggest a bias correction technique to improve size of T–Y's statistic.

*5 The super consistency property of a pure random walk process case has been known for sometime, *e.g.*, Fuller (1976, pp.356–369). Stock (1987) is responsible for extending this property.

*6 Morimune and Zhao (1995) discuss and compute Granger causality tests under non-stationarity.

*7 The main VAR results in section 1.4 were unchanged, when we employed different

set of sample periods. We, therefore, decided that we might safely ignore any probability of structural change in the whole sample period. Tsukuda and Miyakoshi (1995) dealt with the VAR based Granger causality tests allowing for structural breaks.

*8 The structural VAR representation [see Hamilton (1994, p.330)] is one way out of this.

Chapter 2
Comparing Predictive Performances of Several Models of Interest Futures

2.1 Introduction

This chapter compares the forecasting performance of six models, using the data of daily three month Euroyen interest futures. The six models include the Heath, Jarrow, and Morton (1992) (hereafter HJM) model, a short rate model, an ARIMA model, two VAR models, and a prediction method based on seasonal decomposition. We compare the in- and out-of-sample realization of the forward rate with the predicted value. The measure of performance that we focus on is root mean squared forecast error (RMSFE).

As is well known, in their celebrated paper, HJM proposed a continuous time instantaneous forward rate model that gives a general and complete description for variations of term structures of interest rates, and derived a necessary and sufficient condition for the existence of an equivalent martingale measure under which all discount bonds and associated contingent claims are priced or valued by the no-arbitrage concept. Although the HJM model has attracted a great deal of theoretical interests, only a few attempts have been made so far on the practical implementation of the model, since the no-arbitrage condition imposes an integral restriction on the drift and volatility functions with the market price processes of risk. Kamizono and Kariya (1996) specified the volatility term using a step function, and thus made the HJM model tractable.

Chan *et al.* (1992) introduced a spot rate model nesting many term structure models. They compared explanatory powers of the nested models to that of the nesting model. Brenner, Harjes, and Kroner (1996) investigated forecasting performances of several spot rate models focusing on their volatility. Amin and Morton (1993) examined six implied volatility functions for the HJM model. Although the HJM model is not meant to be used for forecasting, investors use it for pricing in practice. Our aim of this chapter is to compare the forecasting performances of six models including the HJM model. Section 2.2 describes our models and Section 2.3 data and estimation results. Section 2.4 reports forecasting performances and Section 2.5 concludes this chapter.

2.2 Models and Implementation Issues

2.2.1 The HJM Model and its Specification

We briefly introduce the HJM model and its specification by Kamizono and Kariya (1996).

HJM specify the movement of the instantaneous forward rate at time t for maturity T, $f(t,T)$, as the following Ito process;

$$df(t,T) = \alpha(t,T)dt + \sum_{i=0}^{n} \sigma_i(t,T)dW_i(t), \tag{2.1}$$

where $\{W_i(t)\}$'s are $(n+1)$ independent standard Brownian motions defined on the probability space $(\Omega, \mathcal{F}, P)$, and α and σ_i's represent respectively the drift and the volatility functions. α and σ_i's generally depend on the entire instantaneous forward rate for maturity s in $t \leq s \leq T$. HJM showed that

imposing non-arbitrage condition on the model under risk-neutral measure is equivalent to the following constraint on the drift function;

$$\alpha(t,T) = -\sigma_i(t,T)\left[\phi_i(t) - \int_t^T \sigma_i(t,u)du\right], \tag{2.2}$$

where $\phi_i(t)$'s are stochastic processes which do not depend on T. HJM also showed that $\phi_i(t)$'s can be regarded as the market prices of risk.

To make the no-arbitrage condition in (2.2) tractable, Kamizono and Kariya (1996) made assumptions as follows. First, for functions $\phi_i(t)$'s they simply assumed

$$\phi_i(t) = \phi_i \quad \text{(constant unknown parameter)}, \quad i = 0, \ldots, n. \tag{2.3}$$

We may note, however, that the market price of risk $\phi_i(t)$ is meant to express the attitude of investors, hence it could fluctuate. Second, they assumed the forward rates z_{jt}, $j = 1, \ldots, n$ for period $[S_j, S_{j+1}]$ and the spot rate z_{0t} for maturity S_0 are available as data. Third, instead of a piecewise flat function like Amin and Morton (1994), they specified the volatility using the following step functions $I(T)$ of maturity T;

$$\begin{aligned} \sigma_i(t,T) &= \xi_{i0}(t) I_{[t,S_1)}(T) + \sum_{j=2}^{n+1} \xi_{ij-1}(t) I_{[S_{j-1},S_j)}(T), \\ & \quad i = 0, \ldots, n. \end{aligned} \tag{2.4}$$

Integrating $\sigma_i(t,T)$ for $[S_j, S_{j+1}]$ gives

$$\xi_{ij}(t) = \frac{1}{S_{j+1} - S_j} \int_{S_j}^{S_{j+1}} \sigma_i(t,u)du.$$

At this stage, they regard the volatility $\xi_{ij}(t)$ of ith factor for period $[S_j, S_{j+1}]$ as an arbitrary stochastic process. Therefore, $\xi_{ij}(t)$ is the cumulative of

ith factor's volatility variation over the interval $[S_j, S_{j+1}]$. Applying these specifications to the HJM model, they made the discrete time approximation using observable forward rates z_{jt}'s,

$$z_{t+h} - z_t = \mu_t h + \sqrt{h}\Sigma_t' \varepsilon_{t+h}, \quad t = 0, h, \ldots, (N-1)h, \tag{2.5}$$

where

$$\begin{aligned}
z_t &= \begin{bmatrix} z_{0t} \\ \vdots \\ z_{nt} \end{bmatrix}, \mu_t = \begin{bmatrix} \mu_{0t} \\ \vdots \\ \mu_{nt} \end{bmatrix}, \Sigma_t = \begin{bmatrix} \xi_{00}(t) & \ldots & \xi_{0n}(t) \\ \vdots & \ddots & \vdots \\ \xi_{n0}(t) & \ldots & \xi_{nn}(t) \end{bmatrix}, \\
\varepsilon_t &= \begin{bmatrix} \varepsilon_{0t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix} \sim N_{n+1}(0, I_{n+1}), \\
\mu_{jt} &= \sigma_t'(\Sigma_t a_{jt} - \Phi) \quad j = 0, \ldots, n, \\
\sigma_t &= \begin{bmatrix} \xi_{0j}(t) \\ \vdots \\ \xi_{nj}(t) \end{bmatrix}, \Phi = \begin{bmatrix} \phi_0 \\ \vdots \\ \phi_n \end{bmatrix}, \\
a_{0t} &= \begin{bmatrix} \frac{S_0 - t}{2} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad a_{jt} = \begin{bmatrix} S_1 - t \\ S_2 - S_1 \\ \vdots \\ S_j - S_{j-1} \\ \frac{S_{j+1} - S_j}{2} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad j = 1, \ldots, n, \\
z_{jt} &= \frac{1}{S_{j+1} - S_j} \int_{S_j}^{S_{j+1}} f(t,T)dT \quad j = 0, \ldots, n.
\end{aligned}$$

Since the forward rates, z_{jt}'s, are expressed as the integration of the instantaneous forward rate, $f(t,T)$, they transform the HJM model with unobservable variables to the operational one with observable variables. For the parameter vector to be estimated, θ, the log-likelihood function given data z_t is

$$\begin{aligned}\ln L(\theta|z_t) &= -\frac{N}{2}\ln 2\pi - \frac{1}{2}\ln|h\Sigma_t'\Sigma_t| \\ &\quad -\frac{1}{2}(z_{t+h}-z_t-h\mu_t)'(h\Sigma_t'\Sigma_t)^{-1}(z_{t+h}-z_t-h\mu_t),\end{aligned}$$

where N is the numbers of the samples. Maximizing the log-likelihood $\ln L$ for the parameter θ gives the maximum likelihood estimate $\hat{\theta}$.

Kamizono and Kariya (1996) applied this model to three month Euroyen interest futures traded at the Tokyo International Financial Futures Exchange (TIFFE). For this empirical analysis, they specify Σ_t as follows;

$$\begin{aligned}\Sigma_t &= PJ_t, \qquad (2.6)\\ P &= \begin{bmatrix} \frac{1}{\sqrt{1-\rho^2}} & \frac{\rho}{\sqrt{1-\rho^2}} & \frac{\rho^2}{\sqrt{1-\rho^2}} & \cdots & \frac{\rho^n}{\sqrt{1-\rho^2}} \\ & 1 & \rho & \cdots & \rho^{n-1} \\ & & 1 & \cdots & \rho^{n-2} \\ & O & & \ddots & \vdots \\ & & & & 1 \end{bmatrix}, \\ J_t &= \mathrm{diag}\{\lambda_0 z_{0t}, \ldots, \lambda_n z_{nt}\},\end{aligned}$$

where $\mathrm{diag}\{\cdot\}$ represents the diagonal matrix, ρ is the correlation coefficient between forward rates of which maturities are next to each other. λ_j's may be regarded as volatility scale parameters.

In this chapter, since the fluctuations of spot rates heavily affected the

estimation results in Kamizono and Kariya (1996), we remove them. That is, we exclude from the model the parts where i or j becomes zero.

2.2.2 The Short Rate Model and its Specification

To analyze the variety of models for short rate, generalized continuous short rate model (GSR),

$$dr(t) = (\alpha + \beta r(t))dt + \psi[r(t)]^{\gamma} dW(t), \tag{2.7}$$

where $r(t)$ is the interest rate level, $W(t)$ is the one-dimensional Brownian motion, and $\alpha, \beta, \psi, \gamma$ are parameters. The GSR model nests many popular interest rate models (Table 2.1), and is used frequently; see, *e.g.*, Brenner, Harjes, and Kroner (1996, p.86).

We make the discrete time approximation of equation (2.7) using the same method as Kamizono and Kariya (1996),

$$z_{jt+h} - z_{jt} = (\alpha_j + \beta_j z_{jt})h + \psi_j z_{jt}^{\gamma_j} \sqrt{h}\varepsilon_{jt+h}, \quad j = 1, 2, 3, \tag{2.8}$$

where $\varepsilon_{jt} \sim \mathrm{NID}(0, 1)$. In this chapter we estimate the parameter of the GSR by maximum likelihood method, giving the value of γ of short-term interest models (the models (i) to (vii)) in Table 2.1; that is $\gamma = 0, \frac{1}{2}, 1, \frac{3}{2}$.

2.2.3 Time Series Models

We also apply the ARIMA models and VAR (vector autoregressive) models as the approach of time series analysis. We select the ARIMA model dimensions using the SBIC (Schwarz-Bayes information criterion). We estimate two VAR

Table 2.1: Relationships between the GSR and popular interest rate models

$$dr = (\alpha + \beta r)dt + \psi r^{\gamma} dW$$

	Model	α	β	ψ	γ
i.	Merton (1973)		0		0
ii.	Vasicek (1977)				0
iii.	CIR (1985) Square Root				1/2
iv.	Dothan (1978)	0	0		1
v.	Geometric Brownian Motion	0			1
vi.	Brennan and Schwartz (1980)				1
vii.	CIR (1980) Variable-Rate	0	0		3/2
viii.	Constant Elasticity of Variance process: Cox (1975) and Cox and Ross (1976)	0			

systems; a) using each of three forward rates as an endogenous variable, b) taking account of exogenous factors (hereafter VARf, VARx respectively). In this chapter we regard the information from the bond and foreign exchange markets as the factors which are thought to be influence interest futures markets. Hence we construct three-dimensional VAR models which include the forward rate, the Nikkei bonds index and the Nikkei currency index as endogenous variables for VARx. Taking account of effects of unit roots and co-integrations, we estimate VARf and VARx in levels proposed by Toda and Yamamoto (1995).

Shiba and Takeji (1994) have proposed a forecasting procedure for asset prices, using seasonal decomposition method, *e.g.*, *SABL* and X-11. Let X_t be the original series to be predicted, then *SABL* assumes a class of power transformed additive process for transformed X_t, $X_t^{(p)}$:

$$X_t^{(p)} = T_t + S_t + I_t,$$

where the power transformation $X_t^{(p)}$ is defined by

$$X^{(p)} = \begin{cases} X^p & \text{if } p > 0, \\ \ln X & \text{if } p = 0, \\ -X^p & \text{if } p < 0, \end{cases}$$

T_t is the trend component, S_t is the seasonal component, and I_t is called the irregular component. Shiba and Takeji (1994) have proposed that first we forecast each component and then we get predicted value from inverse transformation. *SABL* gives predicted values of the seasonal component. They specify ARIMA model for the irregular components, and make scenarios

for the trend components. We apply their method, using current values of the trend components as one-step-ahead predicted values instead of scenarios.

2.3 Data and Estimation Results

2.3.1 Data

The data we use in this chapter is the set of daily three month Euroyen interest futures traded at TIFFE. The maturities of the TIFFE interest futures are March, June, September and December. The underlying asset of the three month Euroyen interest futures is the three month Tokyo Interbank Offered Rate (TIBOR).

We assume futures rates and forward rates are equivalent following Kamizono and Kariya (1996), and regard (100 – TIBOR)/100 as the three month forward rate. In our choice of maturities, we observe the three futures rate $z_{S_1 t}, z_{S_2 t}, z_{S_3 t}$ for each day t where $S_1 < S_2 < S_3$ correspond to the nearest three maturities, and these $z_{S_j t}$'s correspond to z_{jt}'s in the Subsection 2.2.1. Our sample period is available for the period March 1, 1994 to August 31, 1995. We divide this period into the following 6 sample periods;

Period 1: Mar. 1, 1994 - May 31, 1994	Period 2: Jun. 1, 1994 - Aug. 31, 1994
Period 3: Sep. 1, 1994 - Nov. 30, 1994	Period 4: Dec. 1, 1994 - Feb. 28, 1995
Period 5: Mar. 1, 1995 - May 31, 1995	Period 6: Jun. 1, 1995 - Aug. 31, 1995

In each period, no shift occurs among different contracts and there are about 60 daily data. For example, in period 1, the contracts from June 94 to De-

cember 94 are used. For our analysis h is set to be 1/365 following Kamizono and Kariya (1996).

In the estimation of VAR models, we use the Nikkei bonds index (short-term) and the value of Japanese yen of the Nikkei currency index (hereafter *bond*, *cur*, respectively).

2.3.2 Unit Root Tests

We investigate the integration order of our individual forward rates, *bond* and *cur* using augmented Dickey–Fuller (ADF) tests. In selecting the lag length, we start with the maximal lag length 5, and gradually decrease the lag length until the coefficient of the last included lag is significant by the t test; see Campbell and Perron (1991). To compensate the low power of ADF tests, we apply the method proposed by Dolado, Jenkinson, and Sosvilla-Rivero (1990).

Dolado, Jenkinson, and Sosvilla–Rivero (1990) suggest a procedure to test for a unit root when the form of the DGP is unknown. Their method is as follows: start with the least restrictive of the plausible models (which will generally include a trend and drift) and use the ADF test statistic. If the null hypothesis of a unit root is rejected, there is no need to proceed. Conclude that the sequence does not contain a unit root. If the null hypothesis is not rejected, test for the significance of the trend term under the null of a unit root. If the trend is significant, retest for the presence of a unit root using the normal distribution. If the trend is not significant, estimate the model without the trend and use the ADF test statistic to test for the presence of a unit root. If the null hypothesis of a unit root is not rejected, test for the

Table 2.2: ADF unit root test

		p	τ_{ct}	trend	τ_c	const.	τ
Period 1	z_{1t}	0	-1.236	-1.052	-0.946	0.916	-0.816
	z_{2t}	4	-1.750	-0.741	-1.883	1.886	-0.003
	z_{3t}	0	-1.693	0.095	-1.706	1.680	-0.454
	bond	0	-4.215*	-4.355*	-1.046	1.035	-0.480
	cur	0	-2.230	0.116	-2.277	2.275*	-0.177
	Δz_{1t}	0	-7.809*	-1.066	-7.729*	-0.585	-7.756*
	Δz_{2t}	3	-2.947	-0.980	-2.783	0.095	-2.823*
	Δz_{3t}	0	-7.982*	-0.856	-7.958*	-0.034	-8.045*
	Δbond	0	-8.597*	-1.209	-8.488*	-0.739	-8.492*
	Δcur	0	-10.05*	-0.397	-10.12*	-0.245	-10.20*
Period 2	z_{1t}	0	-1.701	1.248	-1.626	1.622	-0.134
	z_{2t}	0	-1.716	0.871	-1.848	1.839	-0.226
	z_{3t}	0	-1.790	0.553	-1.888	1.886	-0.079
	bond	0	-2.028	1.542	-1.307	1.337	1.494
	cur	0	-1.615	-0.364	-2.092	2.108*	0.885
	Δz_{1t}	0	-8.166*	1.021	-8.101*	0.030	-8.167*
	Δz_{2t}	0	-9.323*	1.115	-9.241*	-0.038	-9.315*
	Δz_{3t}	0	-8.578*	0.638	-8.598*	0.144	-8.665*
	Δbond	0	-7.6058	0.351	-7.665*	1.309	-7.510*
	Δcur	0	-7.593*	-1.209	-7.472*	0.805	-7.449*

Notes: τ_{ct} is the statistic based on the model with constant and trend terms. τ_c is the statistic based on the model with constant term. τ is the statistic based on the model without trend or constant term. 'Trend' is the t value of trend term. 'Const.' is the t value of constant term. p is the lag length of ADF test. Δ is the first-difference operator. '*' indicates that it is significant at the five percent level.

Table 2.2: ADF unit root test: cont.

		p	τ_{ct}	trend	τ_c	const.	τ
Period 3	z_{1t}	0	-1.617	-1.529	-0.677	0.635	-1.535
	z_{2t}	0	-1.693	-1.545	-0.839	0.795	-1.287
	z_{3t}	0	-1.354	-1.398	-0.656	0.616	-1.281
	bond	0	-1.803	-0.171	-1.828	1.822	-0.594
	cur	4	-2.858	1.287	-2.541	2.540*	-0.0832
	Δz_{1t}	2	-5.261*	-0.507	-5.277*	-1.696	-4.912*
	Δz_{2t}	0	-7.596*	-0.263	-7.661*	-1.372	-7.480*
	Δz_{3t}	0	-8.116*	-0.653	-8.131*	-1.403	-7.943*
	Δbond	0	-8.618*	-0.417	-8.673*	-0.579	-8.705
	Δcur	1	-1.777	0.173	-1.923	1.924	0.261
Period 4	z_{1t}	0	-1.846	-1.513	-1.189	1.107	-1.945*
	z_{2t}	0	-2.269	-2.055*	-0.972	0.837	-2.017*
	z_{3t}	0	-2.595	-2.328*	-1.157	1.009	-1.996*
	bond	0	-2.617	-2.560*	-0.475	0.441	-2.359*
	cur	0	-3.233*	3.451*	-1.121	1.124	0.348
	Δz_{1t}	0	-5.666*	0.865	-7.871*	-1.731	-7.552*
	Δz_{2t}	0	-5.549*	0.592	-8.155*	-1.974*	-7.719*
	Δz_{3t}	0	-5.280*	0.567	-7.545*	-1.783	-7.198*
	Δbond	0	-8.076*	-0.650	-8.148*	-2.379*	-7.495
	Δcur	5	-4.767*	1.717	-4.371*	1.347	-4.126*

Notes: τ_{ct} is the statistic based on the model with constant and trend terms. τ_c is the statistic based on the model with constant term. τ is the statistic based on the model without trend or constant term. 'Trend' is the t value of trend term. 'Const.' is the t value of constant term. p is the lag length of ADF test. Δ is the first-difference operator. '*' indicates that it is significant at the five percent level.

Table 2.2: ADF unit root test: cont.

		p	τ_{ct}	trend	τ_c	const.	τ
Period 5	z_{1t}	0	-1.884	-1.513	-1.311	0.546	-4.002*
	z_{2t}	0	-1.853	-1.492	-1.180	0.424	-3.830*
	z_{3t}	0	-1.838	-1.478	-1.131	0.462	-3.331*
	bond	0	-1.429	-1.498	0.111	-0.648	-4.667*
	cur	0	-2.008	0.482	-2.364	2.426*	1.526
	Δz_{1t}	0	-7.350*	1.352	-7.176*	-4.114*	-5.206*
	Δz_{2t}	0	-6.482*	0.500	-6.510*	-2.943*	-5.471*
	Δz_{3t}	0	-6.446*	0.500	-6.469*	-2.633*	-5.641*
	Δbond	4	-3.477*	-0.291	-3.584*	-2.978*	-1.879*
	Δcur	0	-7.395*	-1.296	-7.241*	1.492	-7.013*
Period 6	z_{1t}	0	-2.215	-1.759	-1.336	1.215	-0.922
	z_{2t}	0	-2.199	-1.696	-1.416	1.273	-0.969
	z_{3t}	0	-2.193	-1.311	-1.773	1.664	-0.794
	bond	0	-1.585	0.627	-1.677	1.670	-0.184
	cur	0	-2.342	-2.707*	0.575	-0.679	-2.204*
	Δz_{1t}	0	-9.063*	0.332	-9.127*	-1.117	-9.051*
	Δz_{2t}	0	-8.563*	0.268	-8.632*	-1.143	-8.548*
	Δz_{3t}	0	-8.527*	0.301	-8.596*	-0.901	-8.574*
	Δbond	0	-7.961*	1.050	-7.888*	-0.223	-7.946*
	Δcur	0	-9.675*	-1.738	-9.368*	-2.504*	-8.673*

Notes: τ_{ct} is the statistic based on the model with constant and trend terms. τ_c is the statistic based on the model with constant term. τ is the statistic based on the model without trend or constant term. 'Trend' is the t value of trend term. 'Const.' is the t value of constant term. p is the lag length of ADF test. Δ is the first-difference operator. '*' indicates that it is significant at the five percent level.

significance of the constant. If the drift is significant, retest for the presence of a unit root using the normal distribution. If the drift is not significant, estimate the model without the trend or drift, then test for the presence of a unit root. We set the significance level of tests to be five percent.

As shown in Table 2.2, for most of our series (in levels) we are not able to reject the null of a unit root at five percent significant level. (z_{1t}, z_{2t}, z_{3t}) in period 4 reject the existence of unit roots, and seem to be trend stationary processes. (z_{1t}, z_{2t}, z_{3t}) in period 5, *bond* and *cur* also reject the null of unit roots. All the first-differenced sequences except *cur* in period 3 reject the existence of unit roots. Since we probably made *cur* in period 3 over-differenced, it can not reject the null. We conclude, therefore, that for all forward rates in each period the maximum order of possible integration is one.

2.3.3 Estimation Results

We estimate following six models,

1. ARIMA model, 2. VARf model, 3. VARx model,
4. GSR model, 5. HJM model, 6. *SABL*.

To select orders of ARIMA model, we first examine the sample autocorrelation and partial autocorrelation functions of each of the forward rates. In each series, the autocorrelation function decays geometrically, and the partial autocorrelation function at lag 1 has a spike. This suggests, at first sight, low-order AR models. We should, however, notice that the selection of the order of integration using correlograms is based on the large sample theory,

Table 2.3: ARIMA estimates

$$z_t = c + \phi z_{t-1} + \varepsilon_t$$

	Period 1			Period 2			Period 3		
	z_{1t}	z_{2t}	z_{3t}	z_{1t}	z_{2t}	z_{3t}	z_{1t}	z_{2t}	z_{3t}
ϕ	0.960	0.931	0.912	0.929	0.914	0.898	0.976	0.964	0.970
	(20.78)	(19.33)	(16.75)	(19.74)	(17.93)	(15.75)	(27.23)	(22.61)	(19.93)
c	0.090	0.169	0.232	0.163	0.219	0.291	0.057	0.094	0.088
	(0.84)	(1.43)	(1.60)	(1.51)	(1.69)	(1.80)	(0.63)	(0.79)	(0.58)
	Period 4			Period 5			Period 6		
	z_{1t}	z_{2t}	z_{3t}	z_{1t}	z_{2t}	z_{3t}	z_{1t}	z_{2t}	z_{3t}
ϕ	0.982	0.983	0.977	0.987	0.987	0.982	0.941	0.931	0.904
	(40.67)	(39.41)	(35.74)	(62.59)	(54.96)	(47.56)	(23.47)	(21.31)	(17.71)
c	0.040	0.036	0.053	0.002	-0.0008	0.006	0.046	0.052	0.081
	0.68)	(0.57)	(0.69)	(0.08)	(-0.02)	(0.16)	(1.23)	(1.35)	(1.70)

Note: The figures in parenthesis are t values.

Figure 2.1: HJM drift term estimates

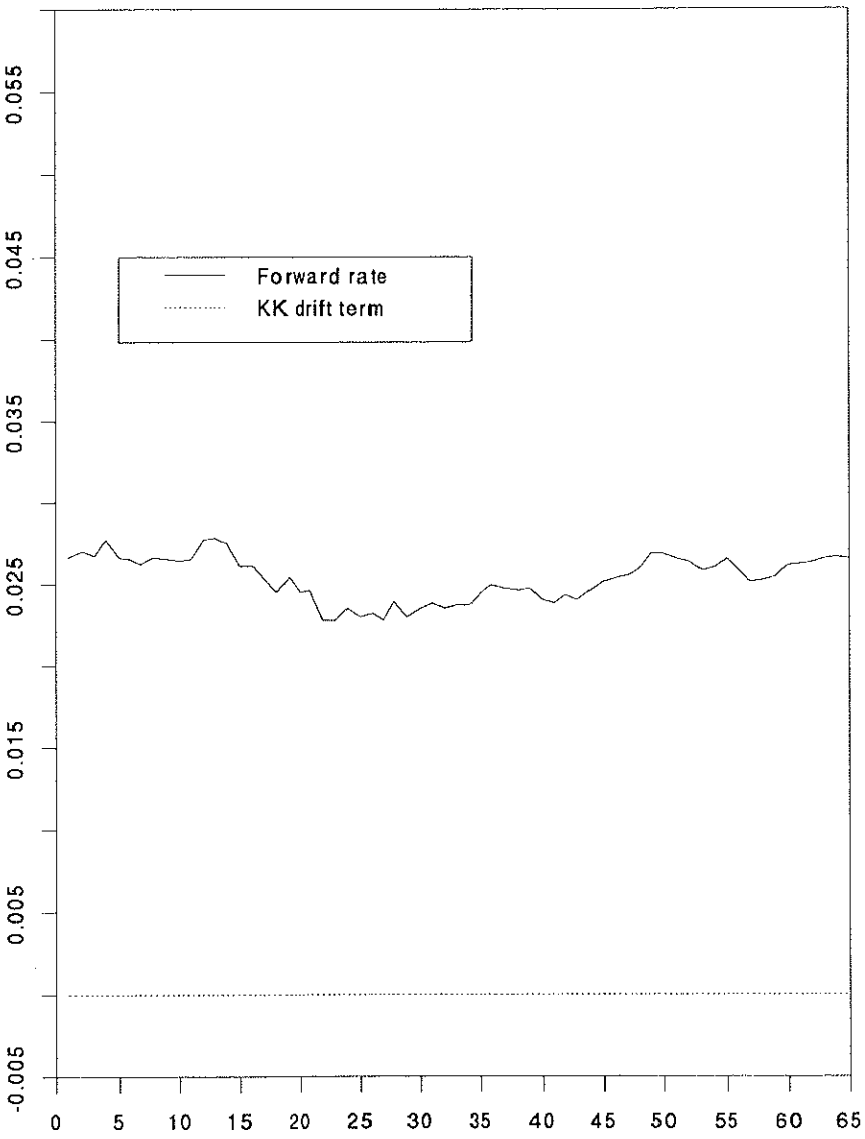

Table 2.4: Lag length selection Wald test: VARf

H_0 : pth lag coefficients= 0			
p	3	2	1
Period 1	4.1185	12.342	68.691*
	(0.903)	(0.195)	(0)
p	3	2	1
Period 2	7.0026	9.4023	56.963*
	(0.637)	(0.401)	(0)
p	3	2	1
Period 3	2.9702	5.4308	68.422*
	(0.965)	(0.795)	(0)

H_0 : pth lag coefficients= 0			
p	3	2	1
Period 4	2.5976	5.7696	55.455*
	(0.978)	(0.763)	(0)
p	5	4	3
Period 5	2.9215	1.6827	22.123*
	(0.967)	(0.996)	(.008)
p	3	2	1
Period 6	6.1128	2.8286	38.943*
	(0.729)	(0.971)	(0)

Notes: The figures in parenthesis are P-values. '*' indicates that it is significant at the five percent level. The critical value is given by $\chi^2_{0.05}(9) = 19.0$.

and that in small samples, correlograms of non-stationary data also converge quickly to zero. On the other hand, although we could not reject the null of unit roots for most of the series, we should not rely too much on this result because of small size of samples. Therefore, we estimate some models to use SBIC, and select AR(1) in each case. Table 2.3 shows estimates of AR(1). For example, coefficient of autoregressive term of z_{1t} in period 1, the forward rate of which maturity is June, 1994, is 0.960 and constant, 0.090.

To estimate VARf and VARx we take account of effects of unit roots and co-integrations, hence we use the VAR system in levels proposed by Toda and Yamamoto (1995). Except period 5 of VARf, the Wald statistic (see Toda and Yamamoto, 1996, p.244) select the lag length 1, hence we choose the lag length 2 according to their method[1]. In period 5 of VARf, the Wald

Table 2.5: Lag length selection Wald test: VARx

H_0 : pth lag coefficients= 0									
	Period 1			Period 2			Period 3		
p	z_{1t}	z_{2t}	z_{3t}	z_{1t}	z_{2t}	z_{3t}	z_{1t}	z_{2t}	z_{3t}
3	3.9562	6.7320	8.7477	5.8562	2.1064	2.2496	9.2284	5.0217	3.3567
	(0.914)	(0.665)	(0.461)	(0.754)	(0.990)	(0.987)	(0.416)	(0.832)	(0.948)
2	6.2017	6.5253	4.3968	4.7531	3.3050	5.8726	7.9819	6.2335	3.0843
	(0.720)	(0.686)	(0.883)	(0.855)	(0.951)	(0.753)	(0.536)	(0.716)	(0.961)
1	86.335*	78.016*	67.224*	77.207*	72.305*	78.988*	80.188*	76.290*	32.949*
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
	Period 4			Period 5			Period 6		
p	z_{1t}	z_{2t}	z_{3t}	z_{1t}	z_{2t}	z_{3t}	z_{1t}	z_{2t}	z_{3t}
3	7.3469	4.7412	3.5344	6.5154	4.2151	4.8294	2.3262	4.9302	2.2867
	(0.601)	(0.856)	(0.939)	(0.687)	(0.897)	(0.849)	(0.985)	(0.840)	(0.986)
2	2.5851	4.0083	4.5840	16.294	6.6346	3.0144	3.7503	7.0900	8.9360
	(0.979)	(0.911)	(0.869)	(0.061)	(0.675)	(0.964)	(0.927)	(0.628)	(0.443)
1	55.094*	30.565*	44.505*	214.21*	148.76*	91.404*	49.597*	57.392*	50.772*
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)

Notes: The figures in parenthesis are P-values. '*' indicates that it is significant at the five percent level. The critical value is given by $\chi^2_{0.05}(9) = 19.0$.

Table 2.6: GSR estimates

	Period 1			Period 2			Period 3		
	z_{1t}	z_{2t}	z_{3t}	z_{1t}	z_{2t}	z_{3t}	z_{1t}	z_{2t}	z_{3t}
α	0.322	0.682	0.915	0.612	0.874	1.108	0.197	0.324	0.337
	(0.70)	(1.42)	(1.58)	(1.41)	(1.83)	(1.73)	(0.54)	(0.78)	(0.72)
β	-14.396	-27.998	-34.850	-26.695	-34.617	-39.158	-8.387	-12.337	-11.554
	(-0.73)	(-1.44)	(-1.59)	(-1.41)	(-1.81)	(-1.72)	(-0.58)	(-0.83)	(-0.76)
ψ	1.651*	2.695*	3.105*	2.070*	2.679*	2.695*	0.961*	1.375*	1.338*
	(8.97)	(9.38)	(8.68)	(15.24)	(13.09)	(14.24)	(9.34)	(9.25)	(11.68)
γ	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
	Period 4			Period 5			Period 6		
	z_{1t}	z_{2t}	z_{3t}	z_{1t}	z_{2t}	z_{3t}	z_{1t}	z_{2t}	z_{3t}
α	0.243	0.204	0.304	-0.044	-0.186	-0.216	0.161	0.192	0.281
	(0.95)	(0.73)	(0.89)	(-0.44)	(-1.53)	(-1.46)	(1.15)	(1.25)	(1.38)
β	-10.817	-9.247	-12.494	-1.717	6.354	7.598	-20.516	-24.848	-32.557
	(-1.04)	(-0.86)	(-1.01)	(-0.25)	(0.73)	(0.75)	(-1.21)	(-1.27)	(-1.36)
ψ	0.958*	1.540*	1.868*	3.996*	5.120*	6.025*	10.384*	13.984*	15.114*
	(12.99)	(14.46)	(13.54)	(11.10)	(10.69)	(9.94)	(16.33)	(18.42)	(17.57)
γ	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5

Notes: The figures in parenthesis are t values. '*' indicates that it is significant at the five percent level.

Table 2.7: HJM estimates

	Period 1	Period 2	Period 3	Period 4	Period 5	Period 6
λ_1	0.112*	0.109*	0.069*	0.071*	0.148*	0.313*
	(17.04)	(16.20)	(15.27)	(18.44)	(18.32)	(30.79)
λ_2	0.187*	0.146*	0.102*	0.115*	0.187*	0.384*
	(19.61)	(19.22)	(15.09)	(18.93)	(20.26)	(31.32)
λ_3	0.220*	0.153*	0.102*	0.144*	0.221*	0.422*
	(12.40)	(24.10)	(21.21)	(18.72)	(16.24)	(16.49)
ϕ_1	1.838	0.302	3.862	5.069	9.455*	2.266
	(0.67)	(0.10)	(1.47)	(1.94)	(3.43)	(0.71)
ϕ_2	-2.331	0.672	-0.280	1.985	-0.457	1.466
	(-1.04)	(0.30)	(-0.12)	(0.67)	(-0.18)	(0.63)
ϕ_3	0.108	-0.839	0.878	1.002	-1.737	-1.061
	(0.04)	(-0.28)	(0.22)	(0.31)	(-0.46)	(-0.34)
ρ	0.902	0.944	0.897	0.877	0.959	0.953

Notes: The figures in parenthesis are t values. '*' indicates that it is significant at the five percent level.

statistic selects the lag length 3, hence we choose the lag length 4. Table 2.4 indicates the values of Wald statistic for VARf, and Table 2.5 for VARx.

For GSR estimation, we select Variable-Rate model(CIR, 1980) of Table 2.1 for all series. Table 2.6 reports GSR estimates.

Table 2.7 presents HJM estimates by Kamizono and Kariya (1996) specification. Predictability of drift term of HJM models will effect the results. Figure 2.1 shows estimates of drift term of z_{2t} in period 1, the forward rate of which maturity is September, 1994, and itself. Since the estimates are almost zero, they will hardly affect the prediction result.

2.4 Forecasting Performance

We define the root mean squared forecast error as

$$RMSFE = \sqrt{\frac{1}{N^*}\sum_{t=1}^{N^*}(z_{jt} - \hat{z}_{jt})^2},$$

where N^* is the length of forecasting period. We use this notation because it is more convenient to see levels than mean squared forecast error. We compare three cases of forecasting performance; (1) in-sample forecasting, (2) out-of-sample forecasting for the last five days before maturity date, (3) out-of-sample forecasting for the middle five days of each period, *i.e.*, about five weeks before maturity date. We must notice that we can not use *SABL* for in-sample prediction since it is by seasonal decomposition method, and that we can not use HJM for out-of-sample prediction since HJM model needs all samples of each period to estimate the drift term.

Table 2.8: Root mean squared forecast error ($\times 100$): in-sample

	Model	ARIMA	VARf	VARx	GSR	HJM	SABL
Period 1	z_{1t}	0.235	0.213	0.236	0.240	0.242	
	z_{2t}	0.410	0.388	0.427	0.432	0.440	
	z_{3t}	0.547	0.490	0.552	0.558	0.572	
Period 2	z_{1t}	0.304	0.299	0.298	0.304	0.311	
	z_{2t}	0.452	0.432	0.445	0.453	0.465	
	z_{3t}	0.539	0.516	0.541	0.542	0.557	
Period 3	z_{1t}	0.155	0.151	0.155	0.156	0.157	
	z_{2t}	0.256	0.249	0.258	0.259	0.260	
	z_{3t}	0.289	0.277	0.289	0.290	0.292	
Period 4	z_{1t}	0.149	0.147	0.148	0.150	0.152	
	z_{2t}	0.263	0.253	0.255	0.262	0.265	
	z_{3t}	0.351	0.329	0.341	0.349	0.354	
Period 5	z_{1t}	0.334	0.245	0.307	0.332	0.334	
	z_{2t}	0.412	0.327	0.393	0.423	0.415	
	z_{3t}	0.509	0.412	0.490	0.528	0.520	
Period 6	z_{1t}	0.343	0.355	0.341	0.359	0.366	
	z_{2t}	0.431	0.447	0.422	0.459	0.469	
	z_{3t}	0.529	0.548	0.531	0.564	0.580	

Note: Since *SABL* is seasonal decomposition method, we can't use it for in-sample prediction.

Table 2.9: Root mean squared forecast error ($\times 100$): last 5 days

	Model	ARIMA	VARf	VARx	GSR	HJM	SABL	
Period 1	z_{1t}	0.0474	0.0523	0.0445	0.0466	0.0416	0.0249	(1,0,0)
	z_{2t}	0.0958	0.0988	0.0968	0.0958	0.0978	0.0448	(1,0,0)
	z_{3t}	0.144	0.148	0.143	0.144	0.150	0.0618	(1,0,2)
Period 2	z_{1t}	0.0377	0.0330	0.0298	0.0395	0.0331	0.0724	(1,0,0)
	z_{2t}	0.0442	0.0518	0.0455	0.0487	0.0314	0.0685	(1,0,0)
	z_{3t}	0.0582	0.0645	0.0506	0.0621	0.0435	0.0748	(1,0,2)
Period 3	z_{1t}	0.0403	0.0409	0.0402	0.0407	0.0386	0.0265	(1,0,2)
	z_{2t}	0.0740	0.0730	0.0785	0.0756	0.0705	0.0644	(1,0,2)
	z_{3t}	0.144	0.136	0.163	0.147	0.131	0.0914	(1,0,2)
Period 4	z_{1t}	0.0169	0.0142	0.0193	0.0169	0.0169	0.00725	(1,0,0)
	z_{2t}	0.0432	0.0563	0.0589	0.0439	0.0440	0.0233	(1,0,0)
	z_{3t}	0.0812	0.0940	0.0896	0.0830	0.0805	0.0544	(1,0,2)
Period 5	z_{1t}	0.118	0.0655	0.107	0.124	0.130	0.0922	(1,0,2)
	z_{2t}	0.177	0.124	0.193	0.181	0.189	0.102	(1,0,2)
	z_{3t}	0.193	0.138	0.218	0.197	0.199	0.112	(1,0,2)
Period 6	z_{1t}	0.0616	0.0871	0.0623	0.0619	0.0584	0.0471	(1,0,0)
	z_{2t}	0.0852	0.143	0.0806	0.0873	0.0748	0.112	(1,0,0)
	z_{3t}	0.115	0.189	0.114	0.116	0.105	0.100	(1,0,0)

Note: The figures in parenthesis are selected order of ARIMA(p,d,q) model used to forecast irregular components of the *SABL* method.

Table 2.10: Root mean squared forecast error ($\times 100$): middle 5 days

	Model	ARIMA	VARf	VARx	GSR	HJM	SABL	
Period 1	z_{1t}	0.0397	0.0582	0.0680	0.0381	0.0423	0.0226	(1,0,0)
	z_{2t}	0.139	0.148	0.152	0.136	0.144	0.0601	(1,0,0)
	z_{3t}	0.174	0.199	0.202	0.171	0.185	0.0857	(1,0,0)
Period 2	z_{1t}	0.0816	0.0755	0.116	0.0814	0.0847	0.108	(1,0,0)
	z_{2t}	0.108	0.0919	0.163	0.107	0.111	0.125	(1,0,0)
	z_{3t}	0.125	0.107	0.198	0.124	0.128	0.125	(1,0,0)
Period 3	z_{1t}	0.0463	0.0515	0.0636	0.0466	0.0439	0.0204	(1,0,0)
	z_{2t}	0.0725	0.0659	0.0899	0.0725	0.0701	2.721	(1,0,0)
	z_{3t}	0.0793	0.0705	0.0894	0.0785	0.0826	2.140	(1,0,0)
Period 4	z_{1t}	0.0340	0.0354	0.0362	0.0333	0.0336	0.0136	(1,0,0)
	z_{2t}	0.0793	0.0887	0.0871	0.0784	0.0751	0.0320	(1,0,2)
	z_{3t}	0.127	0.148	0.127	0.125	0.124	0.0398	(1,0,2)
Period 5	z_{1t}	0.0805	0.0621	0.0776	0.0776	0.0754	0.0145	(2,0,2)
	z_{2t}	0.110	0.0986	0.106	0.107	0.106	0.0327	(2,0,2)
	z_{3t}	0.154	0.140	0.158	0.149	0.151	0.0465	(1,0,2)
Period 6	z_{1t}	0.0719	0.0671	0.0842	0.0741	0.0685	0.0350	(1,0,0)
	z_{2t}	0.105	0.0970	0.117	0.109	0.102	0.0580	(1,0,0)
	z_{3t}	0.115	0.118	0.128	0.118	0.114	0.0709	(1,0,0)

Note: The figures in parenthesis are selected order of ARIMA(p,d,q) model used to forecast irregular components of the *SABL* method.

Firstly, we compute the in-sample RMSFE (Table 2.8). For example, for z_{1t} in period 1, the minimum RMSFE is 0.213 by VARf and the maximum is 0.242 by HJM. We conclude from Table 2.8 that in-sample forecasting performance of the five models is hardly different from each other. Secondly, Table 2.9 shows out-of-sample performances for last 5 days. We insert in-sample performances of HJM model for comparison. For z_{1t} in period 1, the minimum RMSFE is 0.0249 by *SABL* and the maximum is 0.0523 by VARf. In period 1, RMSFEs by *SABL* is almost half of other methods. In period 2, however, RMSFEs by *SABL* is almost twice of others. Since trend components of *SABL* will affect largely forecasting results, if the non-linear trend of a forward rate changes drastically in the forecasting period, then the *SABL*'s performance will be poor, and if the non-linear trend does not change, it will be excellent. We conclude from Table 2.9 that *SABL* produce excellent or poor results, which depend on changes of trend components. Results of out-of-sample performances for middle 5 days (Table 2.10) give us the same conclusion.

2.5 Concluding Remarks

We compare the in- and out-of-sample forecasting performance of Heath, Jarrow, and Morton (1992) model, a short rate model, ARIMA, VAR, and a prediction method based on *SABL*, using the data of daily three month Euroyen interest futures. RMSE of the forecasts from the HJM model resemble to those of the other models, except for the *SABL*. *SABL*'s performance is sometimes excellent, but occasionally very poor. It seems that such differences in the *SABL* performances have been caused by the changes in non-liner

trend components.

Footnotes

*1 Toda and Yamamoto (1995) have proposed that VAR system be modeled possibly in levels, *i.e.*, without differencing to achieve stationarity in each variable, but augmenting the lag length by *a priori* setting the maximum order of integration. From the result of unit root tests (maximum order of integration is 1) and Wald test (the lag length of VAR is 1), we estimate the VAR system in levels of order $2(=1+1)$.

Chapter 3
A New Method to Estimate Stochastic Volatility Models: A Log-GARCH Approach

3.1 Introduction

Changes in asset return variance or volatility over time may be modeled using the GARCH models, developed by Engle (1982) and Bollerslev (1986). In GARCH models, such effects are captured by letting the conditional volatility be a function of squares of previous observations and past volatilities. Since the models are formulated in terms of the conditional distribution, the maximum likelihood estimation may be implemented straightforwardly. A wide range of GARCH models has now appeared in the econometric literature; see, *e.g.*, surveys by Bollerslev, Engle, and Nelson (1995) and Palm (1996).

An alternative to the GARCH type modeling is to use an unobserved volatility component model. In this approach, the logarithm of an unobserved volatility is modeled as a linear stochastic process, such as an autoregression. Models of this kind are called the stochastic volatility (SV) models. Just like the GARCH models, SV models are used both in mathematical finance and financial econometrics. In mathematical finance, Scott (1987), Hull and White (1987), Wiggins (1987), Chesney and Scott (1989), and Melino and Turnbull (1990) were interested in pricing options assuming continuous time SV models for the underlying asset. In financial econometrics, continuous

time SV models are usually discretized. Clark (1973), Tauchen and Pitts (1983), and Andersen (1996) showed that asset returns follow a SV model under the assumption that asset price is a function of a random process of information arrival. Surveys of SV models are given by Taylor (1994) and Ghysels, Harvey, and Renault (1996), among others.

Statistical properties of SV models are easily obtained from the properties of the process generating the volatility component. Their main disadvantage, however, is that they are difficult to estimate by the maximum likelihood estimation method. Taylor (1986), Melino and Turnbull (1990) and Andersen and Sørensen (1996) used the method of moments (MM) to avoid the integration problems associated with evaluating the likelihood directly. Nelson (1988), Harvey, Ruiz, and Shephard (1994) and Ruiz (1994) employed approximate Kalman filtering methods in their quasi-maximum likelihood (QML) estimation. The Monte Carlo evidence of Jacquier, Polson, and Rossi (1994), however, implies that MM and Kalman filtering procedures suffer from poor finite sample performance because they do not depend on the exact likelihood.

When researchers can neglect certain computational costs, there are better alternatives based on the exact likelihood: Danielsson and Richard (1993) and Danielsson (1994a) proposed simulation-based maximum likelihood (SML) procedures; Watanabe (1997a) developed nonlinear filtering maximum likelihood (NFML) procedures; Jacquier, Polson, and Rossi (1994) suggested a Bayesian inference and used the Markov chain Monte Carlo (MCMC). Shephard (1995) used the simulated EM algorithm. Although these methods are computationally intensive, experimental results of Jacquier, Polson, and

Rossi (1994), Danielsson (1994b) and Watanabe (1997a) show that these estimators outperform MM and Kalman filter approaches.

The log-GARCH models, proposed by Geweke (1986) and Pantula (1986), are the logarithmic extension of the GARCH models. Their models may be interpreted as a special case of the Exponential GARCH models, originally developed by Nelson (1991). We derive a log-GARCH representation of a class of SV models, including linear regression models with ARMA(p,q)-SV errors. To estimate these SV models, we propose a new QML method via the log-GARCH approach based on either a Gaussian or a standardized t distribution.

We conduct Monte Carlo experiments to analyze the finite sample property of our method to estimate simple SV models. In parameter estimation, the performance of QML estimator via the log-GARCH approach is better than that of GMM estimators and QML estimators via the Kalman filter; and close to that of NFML estimator, which is a computationally intensive method. In volatility estimation using approximate Kalman filtering methods, smoothed estimates with QML estimates via the log-GARCH approach based on standardized t distribution outperforms those with GMM estimators and QML estimators via the Kalman filter.

Above simulation results show that some computationally intensive methods are preferable to the log-GARCH approach when researchers can neglect estimation time. In this sense, it is worth developing a computationally intensive method for ARMA-SV models; see appendix 3.G.

Organization of this chapter is as follows. Section 3.2 shows that a simple

SV process in a sequence, say y_t, can be interpreted as a log-GARCH process in y_t. Section 3.3 proposes a QML method via the log-GARCH approach and presents Monte Carlo results. Section 3.4 reports a brief empirical findings for the yen/dollar daily exchange rate and Section 3.5 concludes the chapter.

3.2 Stochastic Volatility Processes

A simple stationary SV model is given by

$$\begin{aligned} y_t &= \sqrt{h_t}\eta_t, \quad \eta_t \sim \text{NID}(0,1), \\ \ln h_t &= \gamma + \phi \ln h_{t-1} + \sigma_\nu \nu_t, \quad \nu_t \sim \text{NID}(0,1), \end{aligned} \tag{3.1}$$

where η_t is generated independently of ν_t and $|\phi| < 1$. Working with logarithms ensures that h_t is always positive. As noted in the previous section, MM and Kalman filter approaches are easier to conduct than computer intensive methods, SML, NFML and Bayesian MCMC. Some experimental results, however, imply that computer intensive estimators outperform MM and Kalman filter approaches.

As pointed out by Harvey, Ruiz, and Shephard (1994) and Shephard (1996), the simple SV process in y_t is equivalent to an ARMA(1,1) process in $\ln y_t^2$ with a skewed non-Gaussian noise. In this section, we first derive the parameters of the ARMA(1,1) representation using the results of Hamilton (1994, chapter 4). We next show that the simple SV process in y_t is equivalent to a log-GARCH(1,1) process in y_t with a symmetric non-Gaussian noise.

In the simple SV model (3.1), transforming y_t by taking logarithms of the squares and concentrating out h_t, we obtain

$$\ln\, y_t^2 = \phi \ln\, y_{t-1}^2 + \gamma + \sigma_\nu \nu_t + \ln\, \eta_t^2 - \phi \ln\, \eta_{t-1}^2. \tag{3.2}$$

The mean and variance of $\ln\eta_t^2$ are known to be

$$c_\eta \equiv E[\ln\eta_t^2] = \left[\psi\left(\frac{1}{2}\right) - \ln\left(\frac{1}{2}\right)\right] \simeq -1.27036, \quad Var[\ln\eta_t^2] = \psi'\left(\frac{1}{2}\right) = \frac{\pi^2}{2},$$

where $\psi(\cdot)$ is the Digamma function and $\psi'(\cdot)$ is the Trigamma function, defined as

$$\psi(z) \equiv d[\ln\Gamma(z)]/dz = \Gamma'(z)/\Gamma(z), \quad \psi'(z) \equiv d\psi(z)/dz = d^2[\ln\Gamma(z)]/d^2z;$$

see Abramovitz and Stegun (1970, p.943).

Equation (3.2) can be rewritten as

$$\ln y_t^2 = (1-\phi)c_\eta + \gamma + \phi \ln\, y_{t-1}^2 + \sigma_\nu\nu_t + \left(\ln\eta_t^2 - c_\eta\right) - \phi\left(\ln\eta_{t-1}^2 - c_\eta\right),$$
$$\left(\ln\eta_t^2 - c_\eta\right) \sim \text{iid}(0, \pi^2/2), \tag{3.3}$$

where $\text{iid}(0,\pi^2/2)$ denotes an independently identically distributed process with mean zero and variance $\pi^2/2$. Let us note that a white noise process variable combined with an MA(1) variable reduces to an MA(1) process; see Hamilton (1994, chapter 4, pp.102–106). Using this fact, the last three terms of equation (3.3) becomes an MA(1) process,

$$\sigma_\nu\nu_t + \left(\ln\eta_t^2 - c_\eta\right) - \phi\left(\ln\eta_{t-1}^2 - c_\eta\right) = \varepsilon_t - \theta\varepsilon_{t-1}, \quad \varepsilon_t \sim \text{WN}(0,\sigma^2) \tag{3.4}$$

where $\text{WN}(0,\sigma^2)$ denotes a white noise process with mean zero and variance σ^2, and

$$\sigma^2 = \frac{\pi^2\phi}{2\theta}, \quad \theta = \frac{1}{2\phi}\left[1 + \phi^2 + \frac{2\sigma_\nu^2}{\pi^2} - \sqrt{\left(1 - \phi^2 + \frac{2\sigma_\nu^2}{\pi^2}\right)^2 + \frac{8\phi^2\sigma_\nu^2}{\pi^2}}\right]. \tag{3.5}$$

The other solution of the quadratic equation for θ does not satisfy the invertibility condition. Note that equation (3.5) implies $\phi\theta > 0$ and $0 < |\theta| < |\phi| < 1$.

As a result, the model reduces to an ARMA(1,1) process in $\ln y_t^2$:

$$\ln y_t^2 = (1-\phi)c_\eta + \gamma + \phi \ln y_{t-1}^2 + \varepsilon_t - \theta\varepsilon_{t-1}, \quad \varepsilon_t \sim \mathrm{WN}(0, \sigma^2). \tag{3.6}$$

where ε_t is a skewed and leptokurtic white noise; see appendix 3.A.

We next show a simple SV process in y_t is equivalent to a log-GARCH(1,1) process in y_t with a symmetric non-Gaussian noise. Define a standardized process z_t and a positive predetermined variable σ_t by

$$z_t \equiv \frac{1}{c_z}\left(\eta_t \prod_{i=1}^{\infty} |\eta_{t-i}|^{(\theta-\phi)\theta^{i-1}}\right) \exp\left[-\frac{c_\eta(1-\phi)}{2(1-\theta)} + \frac{\sigma_\nu}{2}\nu_t + \frac{\sigma_\nu}{2}\sum_{i=1}^{\infty}\theta^i\nu_{t-i}\right], \tag{3.7}$$

$$\ln\sigma_t^2 \equiv \ln c_z^2 + \frac{1}{1-\theta}(\gamma + (1-\phi)c_\eta) + (\phi-\theta)(1-\theta L)^{-1}\ln y_{t-1}^2, \tag{3.8}$$

where

$$\begin{aligned} c_z \;=\; \exp\Bigg[&\frac{\sigma_\nu^2}{4(1-\theta^2)} - \frac{c_\eta}{2} - \frac{(\theta-\phi)\theta}{2(1-\theta)}\psi\left(\frac{1}{2}\right) \\ &+\frac{1}{2}\sum_{i=1}^{\infty}\left[\ln\Gamma\left((\theta-\phi)\theta^i + \frac{1}{2}\right) - \ln\Gamma\left(\frac{1}{2}\right)\right]\Bigg], \end{aligned}$$

and L denotes the lag operator. z_t is a weak stationary process that has, at least, up to fourth order moments if $|\phi| < 1$ and $\sigma_\nu > 0$; see appendix 3.B. σ_t is measurable with respect to the time $t-1$ information set.

Rewriting ε_t in terms of z_t or σ_t as

$$\varepsilon_t = \ln z_t^2 - \ln c_z^2 = \ln y_t^2 - \ln\sigma_t^2 - \ln c_z^2 \tag{3.9}$$

and substituting equation (3.9) to (3.6), we obtain an equivalent representation of the log-AR(1) SV process,

$$\begin{aligned} y_t &= \sigma_t z_t, \quad z_t \sim \mathrm{WN}(0,1), \\ \ln\sigma_t^2 &= \left[\gamma + (1-\phi)c_\eta + (1-\theta)\ln c_z^2\right] + (\phi-\theta)\ln y_{t-1}^2 + \theta\ln\sigma_{t-1}^2 \end{aligned} \tag{3.10}$$

Note that z_t is serially uncorrelated with mean zero and variance one. Therefore the log-AR(1) SV process in y_t can be interpreted as a log-GARCH(1,1) process in y_t which has a heavy-tailed and symmetric conditional distribution.

Reparameterizing $(\gamma, \phi, \sigma_\nu)$ as

$$\begin{aligned} \alpha_0 &\equiv \gamma + (1-\phi)c_\eta + (1-\theta)\ln c_z^2, \\ \alpha_1 &\equiv \phi - \theta, \\ \beta_1 &\equiv \theta, \end{aligned}$$

we have

$$\begin{aligned} y_t &= \sigma_t z_t, \\ \ln\sigma_t^2 &= \alpha_0 + \alpha_1 \ln y_{t-1}^2 + \beta_1 \ln\sigma_{t-1}^2, \end{aligned} \tag{3.11}$$

where $|\alpha_1 + \beta_1| < 1$ and $(\alpha_1 + \beta_1)\beta_1 > 0$. The inverse transformations for this mapping are

$$\begin{aligned} \gamma &= \alpha_0 + \alpha_1 c_\eta + \frac{\alpha_1(1-(\alpha_1+\beta_1)\beta_1)\pi^2}{4\beta_1(1+\beta_1)} + \alpha_1\beta_1\psi\left(\frac{1}{2}\right) \\ &\qquad + (1-\beta_1)\sum_{i=1}^{\infty}\left[\ln\Gamma\left(\frac{1}{2} - \alpha_1\beta_1^i\right) - \ln\Gamma\left(\frac{1}{2}\right)\right], \\ \phi &= \alpha_1 + \beta_1, \\ \sigma_\nu &= \sqrt{\frac{(\phi-\theta)(1-\theta\phi)\pi^2}{2\theta}} = \sqrt{\frac{\alpha_1(1-(\alpha_1+\beta_1)\beta_1)\pi^2}{2\beta_1}}. \end{aligned}$$

Note that

$$\begin{aligned}\theta &= \beta_1, \\ c_z^2 &= \exp\left[\frac{\alpha_1(1-(\alpha_1+\beta_1)\beta_1)\pi^2}{4\beta_1(1-\beta_1^2)} - c_\eta + \frac{\alpha_1\beta_1}{1-\beta_1}\psi\left(\frac{1}{2}\right)\right. \\ &\qquad \left. + \sum_{i=1}^{\infty}\left[\ln\Gamma\left(\frac{1}{2}-\alpha_1\beta_1^i\right) - \ln\Gamma\left(\frac{1}{2}\right)\right]\right].\end{aligned}$$

We now consider a linear regression model with ARMA(p,q)-SV errors, or simply ARMA-SV model:

$$y_t = X_t\delta + u_t, \tag{3.12}$$

$$A(L)u_t = B(L)e_t, \tag{3.13}$$

$$e_t = \sqrt{h_t}\eta_t, \quad \eta_t \sim \mathrm{NID}(0,1), \tag{3.14}$$

$$\ln h_t = \gamma + \phi\ln h_{t-1} + \sigma_\nu\nu_t, \quad \nu_t \sim \mathrm{NID}(0,1), \tag{3.15}$$

where η_t is generated independently of ν_t, X_t is a $1\times k$ vector, δ is a $k\times 1$ parameter vector,

$$A(L) = 1 - a_1L - \cdots - a_pL^p,$$
$$B(L) = 1 - b_1L - \cdots - b_qL^q,$$

and L is the lag operator. This ARMA-SV model is a straightforward extension if we take account of the empirical results that many asset return series may be expressed as ARMA processes.

The ARMA-SV model of (3.12)-(3.15) can be interpreted as an ARMA-log-GARCH(1,1) model in a similar fashion to the simple SV model (3.1):

$$y_t = X_t\delta + u_t,$$

$$
\begin{aligned}
A(L)u_t &= B(L)e_t, \\
e_t &= \sigma_t z_t, \qquad z_t \sim \mathrm{WN}(0,1), \qquad (3.16) \\
\ln\sigma_t^2 &= \alpha_0 + \alpha_1 \ln e_{t-1}^2 + \beta_1 \ln\sigma_{t-1}^2, \qquad (3.17)
\end{aligned}
$$

where the definitions of the transformed parameters $(\alpha_0, \alpha_1, \beta_1)$ are the same as in the simple SV case. z_t has the heavy-tailed and symmetric conditional distribution

3.3 Econometric Methodology

3.3.1 A New QML Estimation Method

Under a set of mild regularity conditions, the quasi-maximum likelihood estimator proposed by White (1982) is consistent and asymptotically normal. Lee and Hansen (1994) established the consistency and asymptotic normality properties of the quasi-maximum likelihood estimator of the GARCH(1,1) and IGARCH(1,1) models assuming that the standardized variable z_t is stationary and ergodic with a bounded fourth conditional moment. Unfortunately, as with other ARCH models including EGARCH models, a satisfactory asymptotic theory for the log-GARCH is as yet unavailable. In the remainder of this chapter, we assume that the quasi-maximum likelihood estimator is consistent and asymptotically normal[1]. The asymptotic distribution for the QML estimator of the parameter vector $\omega \equiv (a_0, a_1, b_1)'$ takes the form

$$
\sqrt{T}(\hat{\omega}_{QML} - \omega_0) \overset{a}{\sim} N(0, A_0^{-1} B_0 A_0^{-1}) \qquad (3.18)
$$

where A_0 is the information matrix evaluated at the true parameter vector, and B_0 is the expected value of the outer product of the gradients evaluated

at the true parameters. Hence, the asymptotic distribution for the QML estimator of $f(\omega) = (\gamma, \phi, \sigma_\nu)'$ takes the form

$$\sqrt{T}(f(\hat{\omega}_{QML}) - f(\omega_0)) \overset{a}{\sim} N\left(0, \left(\frac{\partial f(\omega_0)}{\partial \omega'}\right) A_0^{-1} B_0 A_0^{-1} \left(\frac{\partial f(\omega_0)}{\partial \omega'}\right)'\right). \quad (3.19)$$

As in Melino and Turnbull (1990), Harvey, Ruiz, and Shephard (1994) and Ruiz (1994), once we have the parameter estimates, then we can obtain smoothed estimates of $\ln h_t$ by applying the approximate Kalman filter, described in Anderson and Moore (1979, chapter 8). The smoothed estimates of h_t may be converted from the smoothed estimates of log-volatility $\ln h_t$ using standard properties of lognormal distribution[2]. See appendix 3.C for a complete discussion. In case of ARMA-SV models, in addition to parameter estimates, we need the fitted values of e_t for smoothings.

3.3.2 Monte Carlo Experiments

In this subsection, we investigate the finite sample property of the QML estimator via log-GARCH approach for the simple SV model. Jacquier, Polson, and Rossi (1994) have surveyed the literature and considered parameters designs that seem to be adequate in empirical studies. Following Andersen and Sørensen (1996) and Watanabe (1997a), we focus on the following three parameter settings of Jacquier, Polson, and Rossi (1994):

$$\begin{aligned}(\gamma, \phi, \sigma_\nu) &= (-0.736, 0.90, 0.363) \\ &= (-0.368, 0.95, 0.260) \\ &= (-0.147, 0.98, 0.166).\end{aligned}$$

While Jacquier, Polson, and Rossi (1994) consider sample size of $T = 500$ in the majority of their experiments, we use $T = 2000$, which is not uncommon in studies using daily data. The number of replications is 1000.

To construct the log-GARCH quasi-log-likelihood, we may assume z_t in equation (3.10) to be either one of the following two: (a) a standard normal variable, or (b) a standardized t variable favored by Bollerslev (1987), among others. As noted in previous section, z_t has a heavy-tailed and symmetric distribution. If the true distribution is symmetric, then the Gaussian QML estimator obtains estimates that are close to those of the exact ML estimator[3]. We thus use the normal distribution. Taking account of the thick tails, we also use the standardized t distribution.

When the distribution of z_t conditioned on the information up to time $t-1$ assumed to be standardized t with degrees of freedom λ, the conditional quasi-distribution of y_t is standardized t with mean zero and variance σ_t^2, *i.e.*,

$$f(y_t) = \Gamma\left(\frac{\lambda+1}{2}\right)\Gamma\left(\frac{\lambda}{2}\right)^{-1}\left[(\lambda-2)\sigma_t^2\right]^{-1/2}\left[1+\frac{y_t^2}{(\lambda-2)\sigma_t^2}\right]^{-(\lambda+1)/2} \qquad (\lambda > 2),$$

where $f(y_t)$ denotes the conditional quasi-density function of y_t. Analytic derivatives of the log-GARCH quasi-log-likelihood are used instead of numerical differentiation[4].

Table 3.1 shows the mean and the root mean squared error (RMSE) of various estimators. Results of GMM, SML and Bayesian MCMC estimators are respectively obtained from Table 5 and Table 9 of Andersen and Sørensen (1996), Table 2 of Danielsson (1994b) and Table 9 of Jacquier, Polson, and Rossi (1994). Results of QML estimators via the Kalman filter approach and NFML estimators are obtained from Table 1 of Watanabe (1997a). Note that

the results in Table 3.1 have been compiled from different sources, and each author has used different random numbers. Appendices 3.C–3.F provide a discussion and review of the Kalman filter approach, the GMM estimation, SML and NFML methods, respectively. The Bayesian MCMC method is discussed in appendix 3.G.

The biases of QML estimates via the log-GARCH approach based on the normal or the standardized t distribution are smaller than those of GMM estimates, QML estimates via the Kalman filter approach and NFML estimates when $N = 25$, where N is the number of nodes used in NFML. These biases resemble NFML estimates when $N = 50$, SML and Bayesian MCMC estimates. The RMSE of QML estimates via the log-GARCH approach based on the normal or the standardized t distribution are smaller than those of GMM estimates or QML estimates via the Kalman filter approach, and are larger than those of NFML estimates when $N = 50$, SML and Bayesian MCMC estimates. These results imply that 1) QML estimators via the log-GARCH approach outperform GMM estimators and QML estimators via the Kalman filter approach, and 2) the performance of estimators based on computationally intensive methods except NFML estimators when $N = 25$ are mainly superior to that of QML estimators via the log-GARCH approach.

3.3.3 Smoothing Performance

We compare the smoothing performance of approximate Kalman filtering conditional on the true parameters, QML estimates via the log-GARCH approach based on the normal and those based on the standardized t distribution. Following Jacquier, Polson, and Rossi (1994) and Watanabe (1997a),

Table 3.1: Comparison of QML estimator via log-GARCH approach and other estimators, $T = 2000$ (1000 replications)

	γ	ϕ	σ_ν	γ	ϕ	σ_ν	γ	ϕ	σ_ν
Method	-.736	.900	.363	-.368	.950	.260	-.147	.980	.166
LGARCH-n	-.766	.895	.351	-.393	.946	.253	-.175	.976	.167
	(.269)	(.036)	(.074)	(.147)	(.020)	(.048)	(.080)	(.011)	(.032)
LGARCH-t	-.773	.894	.363	-.396	.954	.261	-.175	.976	.170
	(.234)	(.031)	(.066)	(.134)	(.018)	(.044)	(.074)	(.010)	(.030)
GMM	-.592	.920	.279	-.286	.961	.190	-.140	.981	.125
	(.311)	(.042)	(.115)	(.197)	(.027)	(.099)	(.112)	(.015)	(.068)
Kalman	-.837	.887	.383	-.419	.943	.271	-.182	.975	.174
filter	(.369)	(.050)	(.095)	(.187)	(.025)	(.061)	(.087)	(.012)	(.037)
NFML	-.812	.890	.406	-.426	.942	.294	-.194	.974	.197
$N = 25$	(.199)	(.027)	(.068)	(.124)	(.017)	(.052)	(.083)	(.011)	(.043)
NFML	-.776	.895	.368	-.406	.945	.264	-.178	.976	.169
$N = 50$	(.168)	(.023)	(.041)	(.106)	(.014)	(.032)	(.067)	(.009)	(.024)
SML	-.721	.902	.359	Not Available					
	(.15)	(.02)	(.039)						
Bayesian	-.762	.896	.359						
MCMC	(.15)	(.02)	(.034)						

Notes: 'LGARCH-n' denotes the log-GARCH approach based on the normal distribution. 'LGARCH-t' denotes the log-GARCH approach based on the standardized t distribution. Results of GMM, SML and Bayesian MCMC estimators are respectively obtained from Table 5 and Table 9 of Andersen and Sørensen (1996), Table 2 of Danielsson (1994b) and Table 9 of Jacquier, Polson, and Rossi (1994). Results of QML estimators via the Kalman filter approach and NFML estimators are obtained from Table 1 of Watanabe (1997a). N is the number of nodes used in NFML. The table shows the mean and RMSE (in parentheses).

Table 3.2: RMSE smoothing performance of approximate Kalman filtering, $T = 2000$ (1000 replications)

	ϕ		
Method	.90	.95	.98
True parameters	6.856	6.099	5.127
LGARCH-n	6.923	6.127	5.133
LGARCH-t	6.884	6.114	5.125
True parameters	6.85	6.08	5.24
KF	6.92	6.13	5.30

Notes: The second half of Table 3.2 are obtained from Table 2 of Watanabe (1997a). 'LGARCH-n' denotes the log-GARCH approach based on the normal distribution. 'LGARCH-t' denotes the log-GARCH approach based on the standardized t distribution. 'KF' denotes the Kalman filter approach. RMSE $\times 10000$ is displayed.

we compute the grand RMSE,

$$\text{RMSE} = \sqrt{\frac{1}{1000(T-199)} \sum_{i=1}^{1000} \sum_{t=100}^{T-100} (h_{i,t} - \hat{h}_{i,t})^2},$$

where $T = 2000$, $h_{i,t}$ is the volatility simulated at period t in the ith simulation and $\hat{h}_{i,t}$ is its smoothed estimate. The smoothed estimates of $h_{i,t}$ may be converted from the smoothed estimates of log-volatility $\ln h_{i,t}$ using standard properties of lognormal distribution.

Table 3.2 presents the RMSE of smoothed estimates of volatility. The second half of Table 3.2 are obtained from Table 2 of Watanabe (1997a). Smoothing solutions with QML estimates via the log-GARCH approach based on the normal distribution resemble those of the Kalman filtering method. Smoothing solutions with QML estimates via the log-GARCH approach based on the standardized t distribution dominates these two results. Since Jacquier, Polson, and Rossi (1994) reported that RMSE of smoothing

solutions conditional on QML estimates via the Kalman filter are smaller than those with GMM estimates, our results show that smoothing solutions via the log-GARCH approach based on the standardized t distribution dominates those with GMM estimates or QML estimates via the Kalman filter.

3.4 Test for Integration in Log-Volatility

The appropriate procedure for testing for integration in variance is not yet clear. One possible way is to apply augmented Dickey-Fuller test and/or Philips-Perron test to $\ln y_t^2$. However, the reliability of such unit root tests in this situation is questionable. According to Section 3.2 and Harvey, Ruiz, and Shephard (1994), simple SV process (3.1) in y_t can be interpreted as ARMA(1,1) process in $\ln y_t^2$:

$$\ln y_t^2 = \gamma^* + \phi \ln y_{t-1}^2 + \varepsilon_t - \theta \varepsilon_{t-1},$$

where $\gamma^* = \gamma + c_\eta(1-\phi)$. Since the variance of ε_t typically dominates the variance of $\sigma_\nu \nu_t$, the parameter θ will be close to unity for values of ϕ close to one. For example, when $\phi = .98$ and $\sigma_\nu = .166$, which was used in the previous Monte Carlo experiments, θ is .92. As shown in Pantula (1991) and Schwert (1989), when moving-average parameter is very close to one, unit root tests reject the null hypothesis of a unit root too often since the model is difficult to distinguish from white noise.

The other possible way depends on the results of Section 3.2, which showed the simple SV process in y_t of equation (3.1) can be interpreted as a log-GARCH(1,1) model in y_t of (3.11). Since $\phi = \alpha_1 + \beta_1$, the null

hypothesis for $\phi = 1$ can be tested by t statistics for the null hypothesis that $\alpha_1 + \beta_1 = 1$. As is the usual test for IGARCH(1,1) model, this test can be easily conducted. In the following section, we present the example of t test.

3.5 Empirical Example: Daily Exchange Rates

This section illustrates our log-GARCH approach by fitting SV models to the yen/dollar exchange rates. The data consist of daily yen/dollar close exchange rates, S_t, from the Tokyo interbank market. There are total of 1484 daily observations, from January 4, 1991 to December 30, 1996. We calculate the rate of change, r_t, by taking logarithmic difference between the close of two successive trading days.

As we noted in the introduction to this dissertation, tests in a unit root must be carried out before estimation of a model. If we want to test hypotheses about exchange rates, and if we take account of time-varying volatilities, then we must investigate the stationarity for level and volatility of exchange rates. A broad consensus has emerged that nominal exchange rates over the free float period are best described as non-stationary, or $I(1)$, type processes; see, *e.g.*, Baillie and Bollerslev (1989). We shall therefore concentrate on modeling the nominal returns as an univariate SV model and testing the efficient market hypothesis and the persistent volatility hypothesis. Note that a fuller study of exchange rates lies outside the scope of this dissertation.

Table 3.3 provides some summary statistics of the data. The data are unimodal and approximately symmetric, with higher peaks and fatter tail than the Gaussian distribution.

Table 3.3: Summary Statistics of log price change
$r_t \times 100 = \ln(S_t/S_{t-1}) \times 100$

Mean	Std.dev.	Minimum	Maximum
-0.0104	0.6434	-3.6258	3.3937
Skewness	**Kurtosis**	**Studentized range**	**NOBS**
-0.1873	6.5473	10.9094	1484
(0.0636)	(0.0162)		

Sample period: 4 January 1991 to 30 December 1996.
Standard errors in parentheses.
Note: The standard errors are computed as follows:
$\sqrt{6/\text{NOBS}}$ for the coefficient of skewness,
$\sqrt{24/\text{NOBS}}$ for the coefficient of kurtosis.

We consider the following MA(1)-SV model:

$$
\begin{aligned}
r_t &= \delta_0 + \delta_1 W_t + e_t - b_1 e_{t-1} \\
e_t &= \sqrt{h_t}\eta_t \quad \eta_t \sim \text{NID}(0,1), \\
\ln h_t &= \gamma_0 + \gamma_1 W_t - \gamma_1 \phi W_{t-1} + \phi \ln h_{t-1} + \sigma_\nu \nu_t, \quad \nu_t \sim \text{NID}(0,1),
\end{aligned}
$$

where W_t denotes a weekend dummy equal to one following a closure of the market. The MA(1) term is included to take account of the weak serial dependence in the mean. Following Baillie and Bollerslev (1989), the weekend dummy is entered in variance to allow for an impulse effect.

The above MA(1)-SV model in r_t may be interpreted as an MA(1)-log-GARCH(1,1) model in r_t with the similar discussion to Section 3.2,

$$
\begin{aligned}
r_t &= \delta_0 + \delta_1 W_t + e_t - b_1 e_{t-1}, \\
e_t &= \sigma_t z_t, \\
\ln \sigma_t^2 &= \alpha_0 + \gamma_1 W_t - \gamma_1(\alpha_1 + \beta_1) W_{t-1} + \alpha_1 \ln e_{t-1}^2 + \beta_1 \ln \sigma_{t-1}^2,
\end{aligned}
$$

where

$$\begin{aligned}
\alpha_0 &\equiv \gamma_0 + (1-\phi)c_\eta + (1-\theta)\ \ln c_z^2, \\
\alpha_1 &\equiv \phi - \theta, \\
\beta_1 &\equiv \theta.
\end{aligned}$$

Table 3.4 reports the QML estimates based on the normal distribution and the standardized t distribution. The robust QML covariance estimators of Bollerslev and Wooldridge (1992) are used to compute the standard errors. Note that they cannot be used to test whether σ_ν is significantly different from zero.

The estimate of λ is 4.527 and indicates the standardized t distribution is preferred to the normal distribution. In both cases, the coefficients of weekend dummy in level, δ_1, are not significant at the five percent level, and the coefficients in volatility, γ_1, are significant. These results support the empirical findings of Hsieh (1988,1989) and Baillie and Bollerslev (1989). Estimates of ϕ are both significant and indicate strong degree of persistence in volatility. For the standardized t distribution case, the t value for $H_0 : \phi = 1$, which is calculated under the log-GARCH(1,1) representation, is -2.2668 and is rejected the null of unit root in log-volatility. Ruiz (1994), however, found much higher value for ϕ and fitted a random walk stochastic volatility model. It should be pointed out that the Ruiz (1994) study was conducted on October 1, 1981 to June 28, 1885, while our period is from January 4 1991, to December 30, 1996.

Table 3.4: MA(1)-SV Models

	LGARCH-n	LGARCH-t
δ_0	-0.000128	-0.000020
	(0.000004)	(0.000005)
δ_1	-0.000035	0.000026
	(0.000688)	(0.000778)
b_1	-0.005212	0.04527
	(0.001133)	(0.00113)
γ_0	-0.5939	-0.4243
	(0.0066)	(0.0029)
γ_1	0.2703	0.3245
	(0.0040)	(0.0198)
ϕ	0.9418	0.9577
	(0.04481)	(0.0186)
σ_ν	0.2077	0.1613
	(0.1412)	(0.0590)
λ	—	4.527
		(0.00056)
quasi-log-likelihood	5432.76	6347.28

Notes: 'LGARCH-n' denotes the log-GARCH approach based on the normal distribution. 'LGARCH-t' denotes the log-GARCH approach based on the standardized t distribution. Standard error is in parentheses.

3.6 Concluding Remarks

This chapter derived a log-GARCH representation of a class of SV models, including linear regression models with ARMA(p,q)-SV errors. To estimate these SV models, we proposed a new QML method via the log-GARCH approach based on either a Gaussian or a standardized t distribution. We conducted Monte Carlo experiments to analyze the finite sample property of our method to estimate simple SV models. In parameter estimation, the performance of QML estimator via the log-GARCH approach is better than that of GMM estimators and QML estimators via the Kalman filter; and close to that of NFML estimators, which is a computationally intensive method. It should be emphasized that both Bayesian MCMC and SML estimators outperform QML estimator via the log-GARCH approach when the results are available, and that since the performance of NFML estimators depends on the number of nodes, increasing the number may provide better performance. In volatility estimation using approximate Kalman filtering methods, smoothed estimates with QML estimates via the log-GARCH approach based on standardized t distribution outperforms those with GMM estimators and QML estimators via the Kalman filter. We also provided the testing procedure of unit root in log-volatility. The empirical example of daily observations on the yen/dollar exchange rate returns indicates that, for the period which we analyzed, 1) the log-volatility process is stationary and 2) volatility is higher on the day that follows weekend or holiday compared to that of the other days, and that 3) the market is inefficient.

Monte Carlo experiments show that there are better alternatives to estimate simple SV models than the log-GARCH approach when researchers

can neglect such computational costs as below. Jacquier, Polson, and Rossi (1994) reported that typical estimation time of a simple SV model via the Bayesian MCMC is approximately 14 minutes of CPU time by the newest workstation at the time. The SML method takes much more CPU time than the Bayesian MCMC. The NFML method is faster than the Bayesian MCMC, but it requires about half an hour on a PC. Rapid advance in computers will eliminate these costs in the near future. Appendix 3.G developed methods of analyzing ARMA(p,q)-SV regression error models in a Bayesian framework via the Markov chain Monte Carlo.

Using the daily yen/dollar exchange rate, we showed that our Bayesian MCMC technique performs well. This empirical results supports the above findings. While estimates by the Bayesian MCMC technique are slightly different from those by the log-GARCH approach, standard errors by the former are smaller than those by the latter. In this estimation the Bayesian MCMC method on Pentium II PC 330 MHz takes about four hours. Thus, in practice, one should use the Bayesian MCMC technique rather than the log-GARCH approach whenever a fast speed computer server is available. The log-GARCH approach is still useful if one is only interested in parameter estimates.

Both methods have certain advantages, but they still leave room for extensions. First, using a generalized t distribution as the conditional distribution in the log-GARCH approach may improve the finite sample performances of QML estimators, since a generalized t distribution has two shape parameters while a standardized t distribution contains only one shape parameter. Secondly, some authors are interested in capturing comovements in volatility; see

Harvey, Ruiz, and Shephard (1994) and Jacquier, Polson, and Rossi (1995), among others. Thus, both of the log-GARCH approach and the Bayesian MCMC technique should be extended to multivariate SV models. These tasks await further research.

Appendix

The derivation of the moments of ε_t and z_t in this chapter is available upon request.

3.A ε_t and its Moments

In terms of ν_t and η_t, ε_t can be expressed as

$$\varepsilon_t = \sigma_\nu \nu_t + \sigma_\nu \sum_{i=1}^{\infty} \theta^i \nu_{t-i} + \left(\ln\eta_t^2 - c_\eta\right) + \sum_{i=1}^{\infty} (\theta - \phi)\theta^{i-1} \left(\ln\eta_{t-i}^2 - c_\eta\right).$$

Since $|\theta| < |\phi| < 1$, $\sum_{i=0}^{\infty} |\theta|^i < \infty$ and $\sum_{i=1}^{\infty} |\theta - \phi||\theta|^{i-1} < \infty$. Therefore ε_t is a weak stationary process. Its moments up to fourth order are given by

$$\begin{aligned}
&E(\varepsilon_t) = 0, \qquad E(\varepsilon_t^2) = \frac{\pi^2\phi}{2\theta}, \\
&\frac{E(\varepsilon_t^3)}{\left[E(\varepsilon_t^2)\right]^{3/2}} = \frac{1}{\pi^3}\left[1 + \frac{(\theta-\phi)^3}{1-\theta^3}\right]\left(\frac{2\theta}{\phi}\right)^{3/2} \psi^{(2)}\left(\frac{1}{2}\right), \\
&\frac{E(\varepsilon_t^4)}{\left[E(\varepsilon_t^2)\right]^2} = \frac{4\theta^2}{\phi^2}\left[1 + \frac{(\theta-\phi)^4}{1-\theta^4}\right] + 3, \\
&E(\varepsilon_t\varepsilon_{t-j}) = 0, \qquad \text{for } j > 1.
\end{aligned}$$

3.B z_t and its Moments

First we show that $c_z < \infty$, that is

$$\Gamma\left((\theta-\phi)\theta^i + \frac{1}{2}\right) > 0 \quad \text{and} \quad \sum_{i=1}^{\infty}\left[\ln\Gamma\left((\theta-\phi)\theta^i + \frac{1}{2}\right) - \ln\Gamma\left(\frac{1}{2}\right)\right] < \infty.$$

Since ϕ and θ always have the same sign and $0 < |\theta| < |\phi| < 1$,

$$0 < (\phi - \theta)\theta = \frac{\phi^2}{4} - \frac{1}{4\phi^2}\left[1 + \frac{2\sigma_\nu^2}{\pi^2} - \sqrt{\left(1 - \phi^2 + \frac{2\sigma_\nu^2}{\pi^2}\right)^2 + \frac{8\phi^2\sigma_\nu^2}{\pi^2}}\right]^2 \leq \frac{\phi^2}{4} < \frac{1}{4}.$$

(i) If $\theta > 0$, then $0 < (\phi - \theta)\theta^i < (\phi - \theta)\theta < 1/4$.

(ii) If $\theta < 0$ and i is odd, $i = 2j + 1$, then $0 < (\phi - \theta)\theta \cdot \theta^{2j} < 1/4$ since $|\theta| < 1$.

(iii) If $\theta < 0$ and i is even, $i = 2j$, then $(\theta - \phi)\theta^{2j} > 0$.

Therefore $(\theta - \phi)\theta^i + 1/2 > 0$, and thus $\Gamma((\theta - \phi)\theta^i + 1/2) > 0$.

Define a sequence,

$$s_i = \ln\Gamma\left((\theta - \phi)\theta^i + \frac{1}{2}\right) - \ln\Gamma\left(\frac{1}{2}\right).$$

By the Taylor expansion, s_i can be expressed as

$$s_i = (\theta - \phi)\theta^i\psi\left(\frac{1}{2}\right) + o(\theta^i).$$

Note that $b_i = o(a_i)$ implies $\lim_{i\to\infty} b_i/a_i = 0$, and that the Digamma function, $\psi(\cdot)$, is the first derivative of $\ln\Gamma(\cdot)$. Thus

$$\begin{aligned}\frac{s_{i+1}}{s_i} &= \frac{(\theta - \phi)\theta^{i+1}\psi(1/2) + o(\theta^i)}{(\theta - \phi)\theta^i\psi(1/2) + o(\theta^i)} \\ &= \frac{\theta\psi(1/2) + o(1)}{\psi(1/2) + o(1)},\end{aligned}$$

and $\lim_{i\to\infty} s_{i+1}/s_i = \theta$. Since there exists a positive constant K such that $|s_{i+1}/s_i| < K < 1$ for a sufficiently large i, $\sum_{i=1}^{\infty} s_i$ converges absolutely. We have shown that $c_z < \infty$.

Second we show the weak stationarity of z_t. z_t can be rewritten by

$$z_t = \frac{\eta_t}{c_z}\exp\left[\sum_{i=1}^{\infty}(\theta-\phi)\theta^{i-1}|\eta_{t-i}| - \frac{c_\eta(1-\phi)}{2(1-\theta)} + \frac{\sigma_\nu}{2}\nu_t + \frac{\sigma_\nu}{2}\sum_{i=1}^{\infty}\theta^i\nu_{t-i}\right].$$

Therefore z_t is a weak stationary process since $\sum_{i=1}^{\infty}|\theta-\phi||\theta|^{i-1} < \infty$ and $\sum_{i=0}^{\infty}|\theta|^i < \infty$.

The moments of z_t are given by

$$\begin{aligned}
&E(z_t) = 0, \quad E(z_t^2) = 1, \quad E(z_t^3) = 0, \\
&E(z_t^4) = \frac{3}{c_z^4}\exp\left[\frac{2\sigma_\nu^2}{1-\theta^2} - 2c_\eta - \frac{2(\theta-\phi)\theta}{1-\theta}\psi\left(\frac{1}{2}\right)\right. \\
&\qquad\qquad \left. + \sum_{i=1}^{\infty}\left[\ln\Gamma\left(2(\theta-\phi)\theta^i + \frac{1}{2}\right) - \ln\Gamma\left(\frac{1}{2}\right)\right]\right], \\
&E(z_t z_{t-j}) = 0, \qquad \text{for } j > 1.
\end{aligned}$$

Similar way to show $c_z < \infty$ reveals that $E(z_t^4) < \infty$ if $|\phi| < 1$ and $\sigma_\nu > 0$. All odd-order moments are equal to zero. For even-order moments, rth moment $(r = 6, 8, \ldots)$ exists if $(\phi - \theta)\theta < 1/r$.

3.C Approximate Kalman Filtering and QML Estimation

The simple SV model,

$$\begin{aligned}
y_t &= \sqrt{h_t}\eta_t, \quad \eta_t \sim \text{NID}(0,1), \\
\ln h_t &= \gamma + \phi \ln h_{t-1} + \sigma_\nu \nu_t, \quad \nu_t \sim \text{NID}(0,1),
\end{aligned}$$

can be rewritten as

$$\begin{aligned}
\ln y_t^2 &= -1.27 + \ln h_t + \zeta_t, \quad E(\zeta_t) = 0,\ Var(\zeta_t) = \pi^2/2, \\
\ln h_t &= \gamma + \phi \ln h_{t-1} + \sigma_\nu \nu_t, \quad \nu_t \sim \text{NID}(0,1).
\end{aligned}$$

If the distribution of ζ_t is approximated by a normal distribution, the preceding system becomes a standard dynamic linear model, to which the Kalman filter can be applied. The Kalman filter requires three sets of equations; a prediction and updating set which are run forward through the data, and smoothing equations, which are run backward through the data. We follow the standard notation of Anderson and Moore (1979). Let $\overline{\ln h_{t|t-1}}$ be the prediction of $\ln h_t$ based on the information available at time $t-1$. $\Omega_{t|t-1}$ is the variance of the prediction. Let $\overline{\ln h_{t|t}}$ be the update that uses the information at time t and $\Omega_{t|t}$ the variance of the update. The equations that recursively compute the predictions and updatings are given by

$$\begin{aligned}\overline{\ln h_{t|t-1}} &= \gamma + \phi\overline{\ln h_{t-1|t-1}},\\ \Omega_{t|t-1} &= \phi^2\Omega_{t-1|t-1} + \sigma_\nu^2,\end{aligned}$$

and

$$\begin{aligned}\overline{\ln h_{t|t}} &= \overline{\ln h_{t|t-1}} + \frac{\Omega_{t|t-1}}{f_t}\left[\ln y_t^2 + 1.27 - \overline{\ln h_{t|t-1}}\right],\\ \Omega_{t|t} &= \Omega_{t|t-1}(1 - \Omega_{t|t-1}/f_t),\end{aligned}$$

where $f_t = \Omega_{t|t-1} + \pi^2/2$. Once the predictions and updates are computed for $t = 1,\ldots,T$, we can obtain the smoothed estimates, $\overline{\ln h_{t|T}}$, which is the estimate of $\ln h_t$ given all information in the sample. $\Omega_{t|T}$ denotes the variance of $\overline{\ln h_{t|T}}$. The smoothing equations are

$$\begin{aligned}\overline{\ln h_{t|T}} &= \overline{\ln h_{t|t}} + P_t[\overline{\ln h_{t+1|T}} - \overline{\ln h_{t+1|t}}],\\ \Omega_{t|T} &= \Omega_{t|t} + P_t^2(\Omega_{t+1|T} - \Omega_{t+1|t}),\end{aligned}$$

where $P_t = \phi\Omega_{t|t}/\Omega_{t+1|t}$. The system is initialized at the unconditional values, $\Omega_0 = \sigma_\nu^2/(1-\phi^2)$ and $\overline{\ln h_0} = \gamma/(1-\phi)$.

The prediction, updating, and smoothing estimates of h_t are computed using standard properties of the lognormal distribution.

The quasi-likelihood is defined and computed using the prediction error decomposition (see Harvey, 1989)

$$L(\gamma, \phi, \sigma_\nu) \propto -\frac{1}{2}\sum_{t=1}^{T} f_t - \frac{1}{2}\sum_{t=1}^{T} \frac{v_t^2}{f_t},$$

where f_t is the prediction error variance just defined and v_t is the one-step-ahead prediction error, $v_t = \ln y_t^2 + 1.27 - \overline{\ln h_{t|t-1}}$.

3.D GMM Estimation

We briefly survey the GMM estimation used in Andersen and Sørensen (1996). We concentrate on 14 moments case, since their Monte Carlo results suggest that the inclusion of their full 24 moments provides very little information regarding the parameters relative to what is contained in the initial 14 moments.

The SV model is

$$\begin{aligned} y_t &= \sqrt{h_t}\eta_t, \quad \eta_t \sim \text{NID}(0,1), \\ \ln h_t &= \gamma + \phi \ln h_{t-1} + \sigma_\nu \nu_t, \quad \nu_t \sim \text{NID}(0,1). \end{aligned}$$

Then,

$$\ln h_t \sim N\left(\frac{\gamma}{1-\phi}, \frac{\sigma_\nu^2}{1-\phi^2}\right) \sim N(\mu_h, \sigma_h^2).$$

Since the correlation between ν_t and η_t is assumed to be 0, this makes the calculation of moments easy, *e.g.*, $E[|y_t|^r]$ becomes $E[|\eta_t|^r]E[h_t^{r/2}]$. The 14 moments used in Andersen and Sørensen (1996) are as follows: $E[|y_t|] =$

$\sqrt{2/\pi}\exp(\mu_h/2+\sigma_h^2/8)$, $E[y_t^2] = \exp(\mu_h+\sigma_h^2/2)$, $E[|y_t|^3] = 2\sqrt{2/\pi}\exp(3\mu_h/2+9\sigma_h^2/8)$, $E[y_t^4] = 3\exp(2\mu_h + 2\sigma_h^2)$, $E[|y_t y_{t-i}|] = (2/\pi)\exp(\mu_h + \sigma_h^2(1+\phi^i)/4)$ $(i = 2, 4, 6, 8, 10)$, $E[y_t^2 y_{t-1}^2] = \exp(2\mu_h + \sigma_h^2(1+\phi^i))$ $(i = 1, 3, 5, 7, 9)$.

GMM estimation exploits the convergence of selected sample moments to their unconditionally expected values. We denote, hereafter, $\omega = (\gamma, \phi, \sigma_\nu)'$. Let the vector of sample realizations of the moments at time t be $m_t(\omega) = (m_{1t}(\omega), \ldots, m_{14t}(\omega))'$, and let the sample moments be $M_T(\omega) = (M_{1T}, \ldots, M_{14T})$, where $M_{iT}(\omega) = \sum_{t=j+1}^{T} m_{it}(\omega)/(T-j)$, for $i = 1, \ldots, 14$, and j is the maximum lag between the variable defining the sample moments. The corresponding vector of analytical moments is denoted $A(\omega)$. The GMM estimator, $\hat{\omega}_T$, minimizes the distance between $A(\omega)$ and $M_T(\omega)$ in the following quadratic form: $\hat{\omega}_T = \arg\min_\omega (M_T(\omega) - A(\omega))' \Lambda_T^{-1} (M_T(\omega) - A(\omega))$, where Λ_T is a positive definite and possibly random weighting matrix. Hansen (1982) showed that, under suitable regularity conditions, $\hat{\omega}_T$ is consistent and asymptotically normal: $T^{1/2}(\hat{\omega}_T - \omega_0) \sim N(0, \Omega)$. The optimal choice of weighting matrix, Λ^{-1}, in the sense of minimizing the asymptotic covariance matrix, Ω, is given by the inverse of the covariance matrix of the appropriately standardized moment conditions:

$$\Lambda = \lim_{T\to\infty} E\left[\sum_{t,\tau=1}^{T} (m_t - A(\omega_0))(m_t - A(\omega_0))'/T\right].$$

This matrix may be estimated by a kernel estimator for the spectral density of the vector of sample moments at frequency 0. The use of an appropriate weighting matrix is important. Andersen and Sørensen (1996) found that: (1) estimation using a fixed number of lags in the weighting matrix generally is inferior to using the plug-in estimator of lag length suggested by

Andrews (1991), and that (2) the prewhitening method for the weighting matrix suggested by Andrews and Monahan (1992) can be helpful in several settings, and that (3) the automatic bandwidth choice proposed by Newey and West (1994) is appropriate for large samples in which the GMM Bartlett-kernel procedure combined with this automatic bandwidth choice provides inference of a quality that other practical methods will be hard pressed to improve on.

3.E Simulated Maximum Likelihood Methods

The simple SV model is

$$\begin{aligned} y_t &= \sqrt{h_t}\eta_t, \quad \eta_t \sim \text{NID}(0,1), \\ \ln h_t &= \gamma + \phi \ln h_{t-1} + \sigma_\nu \nu_t, \quad \nu_t \sim \text{NID}(0,1), \end{aligned}$$

where η_t is generated independently of ν_t. h_t is a latent variable and the first task is estimation of the parameter vector $\omega = (\gamma, \phi, \sigma_\nu)'$. As Jacquier, Polson, and Rossi (1994) pointed out, this model cannot be easily estimated because the vector of latent observations, $\{h_t\}_{t=1}^T$, has to be integrated out of the joint density of volatility and returns in one T-dimensional integration,

$$l(\omega) = \int_{R^T} f(y, h|\omega) dh,$$

to obtain the likelihood function $l(\omega)$. This integral may not have an analytical solution, and numerical methods must be employed for evaluation.

In Simulated maximum likelihood (SML) estimation of the SV model, the integration in $l(\omega)$ is evaluated by simulation and subsequently maximized

with respect to parameters. Define two functions, a remainder function (RF) $h(\Lambda, Y)$ and an importance function (IF) $\mu(\Lambda|Y)$, where $\lambda_t = \ln h_t$, $\Lambda = \{\lambda_t\}_{t=1}^T$, and $Y = \{y_t\}_{t=1}^T$, such that their product equals the joint density of the latent and observable variables,

$$f(Y, \Lambda|\omega) \equiv h(\Lambda, Y)\mu(\Lambda|Y), \tag{3.A1}$$

and the conditional expectation of $h(\Lambda, Y)$ with respect to $\mu(\Lambda|Y)$ is

$$\begin{aligned} f(Y|\omega) &= E_\mu[h(\Lambda, Y)] \\ &= \int_{R^T} h(\Lambda, Y)\mu(\Lambda|Y)d\Lambda \\ &= \int_{R^T} f(Y, \Lambda|\omega)d\Lambda \end{aligned} \tag{3.A2}$$

An Monte Calro sample for $l(\omega)$ can be constructed by sampling N vectors of the simulated latent variable, Λ_n, from $\mu(\Lambda|Y)$ and inserting each into $h(\Lambda, Y)$. Under appropriate conditions

$$\frac{1}{N}\sum_{n=1}^{N} h(\Lambda_n, Y) = \bar{l}_N(\omega) \xrightarrow{as} l(\omega),$$

where $\xrightarrow{as}$ denotes almost sure convergence. Subsequently $\bar{l}_N(\omega)$ is maximized to obtain parameter estimates. The error in the Monte Calro estimate of the likelihood, $\bar{l}_N(\omega)$, and the parameter, $\hat{\omega}_N$, can be made arbitrarily small by increasing the number of simulations, N. The initial functions that fulfill conditions (3.A1) and (3.A2) can be obtained from a simple factorization of the joint density:

$$\begin{aligned} \mu_0(\Lambda|Y) &\equiv \prod_{t=1}^{T} f(\lambda_t|\omega, \lambda_{t-1}) \\ h_0(\Lambda, Y) &\equiv \prod_{t=1}^{T} f(y_t|\omega, y_{t-1}, \lambda_t). \end{aligned}$$

Although analytically tractable, these initial RF's and IF's are not efficient because millions of simulations are required to achieve reasonable accuracy in the Monte Calro estimate of $l(\omega)$, thus effectively rendering the computation time infinite. To overcome the inefficiency, Danielsson and Richard (1993) proposed an acceleration method whereby the initial RF's and IF's are rewritten in a way that preserves conditions (3.A1) and (3.A2) while minimizing the Monte Calro sampling variance $Var(\bar{l}_N(\omega))$. This requires solving

$$\arg\min_{\mu} \int_{R^T} [h(Y,\Lambda) - f(Y|\omega)]^2 \mu(\Lambda|Y) d\Lambda. \tag{3.A3}$$

An analytical solution to this optimization problem is not available, so numerical method must be employed. Define an auxiliary function, $\xi(\Lambda, Q)$, where Q is a matrix of parameters, and multiply and divide it into the initial IF and RF, respectively. The ratio of the remainder and auxiliary functions and the product of the auxiliary and importance functions form new remainder and importance functions:

$$\begin{aligned} f(Y|\omega) &= \int_{R^T} \left(\frac{h_0(\Lambda, Y)}{\xi(\Lambda, Q)} \right) \mu_0(\Lambda|Y) \xi(\Lambda, Q) d\Lambda \\ &= \int_{R^T} h(\Lambda, Y) \mu(\Lambda|Y) d\Lambda. \end{aligned}$$

The expectation in (3.A2) is not affected, but the variance in (3.A3) changes. The optimization problem in (3.A3) is, therefore, equivalent to finding the auxiliary function that minimizes the sampling variance $Var(\bar{l}_N(\omega))$. Analytical and computational considerations dictate that the choice of $\xi(\Lambda, Q)$ be tractable and thus imply the $\xi(\Lambda, Q)$ selected is not necessarily the best but only sufficiently good. The function chosen for $\xi(\Lambda, Q)$ is an exponential second-order polynomial of the latent variables. The number of elements

in Q is then $2T + 1$, all of which have to be estimated. These parameters could be estimated in a single global nonlinear optimization, but the computational cost could be prohibitive. Instead the parameters are estimated using a linear regression of the log remainder function in each time period on a constant, the simulated latent variable, and the simulated latent variable squared. This process requires iteration until the matrix Q has converged.

Danielsson (1994b) reported that typical estimation time of an SV model is 5 to 20 minutes of workstation time.

3.F Nonlinear Filtering Maximum Likelihood Estimation

A nonlinear filtering technique proposed by Kitagawa (1987) can directly be applied to estimation SV models without any transformation and yields the exact likelihood. Watanabe (1993, 1997) referred to it as the nonlinear filtering maximum likelihood (NFML) estimation.

The simple SV model is

$$\begin{aligned} y_t &= \sqrt{h_t}\eta_t, \qquad \eta_t \sim \text{NID}(0, 1), \\ \ln h_t &= \gamma + \phi \ln h_{t-1} + \sigma_\nu \nu_t, \qquad \nu_t \sim \text{NID}(0, 1), \end{aligned}$$

where η_t is generated independently of ν_t. Define $x_t \equiv \ln h_t$ and $Y_t \equiv (y_1, \ldots, y_t)$. Then we obtain two conditional probability density functions,

$$\begin{aligned} f(x_t|x_{t-1}, Y_{t-1}) &= \frac{1}{\sqrt{2\pi}\sigma_\nu} \exp\left(-\frac{(x_t - \gamma - \phi x_{t-1})^2}{2\sigma_\nu^2}\right), \\ f(y_t|x_t, Y_{t-1}) &= \frac{1}{\sqrt{2\pi \exp(x_t)}} \exp\left(-\frac{y_t^2}{2\exp(x_t)}\right). \end{aligned}$$

Although $f(x_t|x_{t-1}, Y_{t-1}) = f(x_t|x_{t-1})$ and $f(y_t|x_t, Y_{t-1}) = f(y_t|x_t)$ in the simple SV model, we will keep writing Y_{t-1} for more general models. The prediction and updating equations are given by

$$\begin{aligned} f(x_t|Y_{t-1}) &= \int_{-\infty}^{\infty} f(x_t, x_{t-1}|Y_{t-1})dx_{t-1} \\ &= \int_{-\infty}^{\infty} f(x_t|x_{t-1}, Y_{t-1})f(x_{t-1}|y_{t-1})dx_{t-1}, \end{aligned} \tag{3.A4}$$

and

$$\begin{aligned} f(x_t|Y_t) &= f(x_t|y_t, Y_{t-1}) \\ &= \frac{f(x_t, x_{t-1}|Y_{t-1})}{f(y_t|Y_{t-1})} \\ &= \frac{f(x_t|x_{t-1}, Y_{t-1})f(x_{t-1}|y_{t-1})}{f(y_t|Y_{t-1})}, \end{aligned} \tag{3.A5}$$

where the denominator (3.A5) is given by

$$\begin{aligned} f(y_t|Y_{t-1}) &= \int_{-\infty}^{\infty} f(y_t, x_t|Y_{t-1})dx_t \\ &= \int_{-\infty}^{\infty} f(y_t|x_t, Y_{t-1})f(x_t|Y_{t-1})dx_t. \end{aligned} \tag{3.A6}$$

The problem with this filter is that it is difficult to solve analytically. Kitagawa (1987) has suggested a liner spline technique for solving this filter. The basic idea is to approximate the relevant probability density functions for each period by piecewise linear functions. Specifically, $f(x_t|Y_{t-1})$ and $f(y_t|Y_{t-1})$ for $t = 1, \ldots, T$ are approximated by piecewise linear functions, specifying the corresponding number of segments, location of nodes, and values of the density function at each node. If, For each period, $N + 1$ number of nodes is selected, then they are sorted in order of size. We denote them as $x_t^{(0)}, x_t^{(1)}, \ldots, x_t^{(N)}$ for $t = 1, \ldots, T$.

Using the trapezoid rule, we approximate equations (3.A4)-(3.A6) as follows:

$$\begin{aligned} f(x_t^{(i)}|Y_{t-1}) &= \int_{-\infty}^{\infty} f(x_t^{(i)}|x_{t-1}, Y_{t-1}) f(x_{t-1}|Y_{t-1}) dx_{t-1} \\ &\approx \sum_{n=1}^{N} \int_{x_{t-1}^{(n-1)}}^{x_{t-1}^{(n)}} f(x_t^{(i)}|x_{t-1}, Y_{t-1}) f(x_{t-1}|Y_{t-1}) dx_{t-1} \\ &\approx \frac{1}{2} \sum_{n=1}^{N} (x_{t-1}^{(n)} - x_{t-1}^{(n-1)}) \left(f(x_t^{(i)}|x_{t-1}^{(n-1)}, Y_{t-1}) f(x_{t-1}^{(n-1)}|Y_{t-1}) \right. \\ &\quad \left. + f(x_t^{(i)}|x_{t-1}^{(n)}, Y_{t-1}) f(x_{t-1}^{(n)}|Y_{t-1}) \right), (i = 0, 1, \ldots, N), \quad \text{(3.A7)} \end{aligned}$$

and

$$f(x_{t-1}^{(i)}|Y_t) = \frac{f(x_t|x_{t-1}^{(i)}, Y_{t-1}) f(x_{t-1}^{(i)}|y_{t-1})}{f(y_t|Y_{t-1})}, \quad (i = 0, 1, \ldots, N), \qquad \text{(3.A8)}$$

where

$$\begin{aligned} f(y_t|Y_{t-1}) &= \int_{-\infty}^{\infty} f(y_t|x_t, Y_{t-1}) f(x_t|Y_{t-1}) dx_t \\ &\approx \sum_{n=1}^{N} \int_{x_{t-1}^{(n-1)}}^{x_{t-1}^{(n)}} f(y_t|x_t, Y_{t-1}) f(x_t|Y_{t-1}) dx_t \\ &\approx \frac{1}{2} \sum_{n=1}^{N} (x_t^{(n)} - x_t^{(n-1)}) \left(f(y_t|x_t^{(n-1)}, Y_{t-1}) f(x_t^{(n-1)}|Y_{t-1}) \right. \\ &\quad \left. + f(y_t|x_t^{(n)}, Y_{t-1}) f(x_t^{(n)}|Y_{t-1}) \right), (i = 0, 1, \ldots, N). \quad \text{(3.A9)} \end{aligned}$$

Once the distribution of the initial state variable, *i.e.*, $f(x_1)(\equiv f(x_1|Y_0))$ is set, equations (3.A7)-(3.A9) can be solved recursively. The unconditional distribution of x_t is the normal with mean $\gamma/(1-\phi)$ and variance $\sigma_\nu^2/(1-\phi^2)$, which is used for the distribution of the initial state variable.

The log-likelihood is defined as:

$$\ln L(\omega) = \sum_{t=1}^{T} \ln f(y_t|Y_{t-1}). \qquad \text{(3.A10)}$$

Given the parameter values, executing our filter yields the value of $f(y_t|Y_{t-1})$ for each period. Using these values, we can evaluate the log-likelihood defined by equation (3.A10). The SV parameters are estimated by maximizing this log-likelihood.

Watanabe (1997) reported that, for the simulation design used in Subsection 3.3.1, estimating the SV parameters by the NFML procedure using GAUSS version 3.1 on Pentium PC's takes approximately 10 minutes when $N = 25$ and a half hour when $N = 50$.

3.G Bayesian Markov chain Monte Carlo

Monte Carlo simulations in chapter 3 suggest that some computer intensive methods performs better compared to the log-GARCH approach when researchers can afford to neglect computational costs. In this sense, it is worth developing a computationally intensive method for ARMA-SV models.

In this appendix we develop a new method of analyzing ARMA(p,q)-SV regression error models using the MCMC technique in a Bayesian framework. To this effect, we modify a Markov chain sampling scheme developed by Jacquier, Polson, and Rossi (1994) and Nakatsuma (1996). Nakatsuma (1996) designed a Markov chain sampling scheme for a linear regression model with an ARMA(p,q)-GARCH(r,s) error. He used a particular state space representation of an ARMA model with a heteroskedastic variance, and modified the method of Chib and Greenberg (1994) for an ARMA error. While Nakatsuma's (1996) model is a deterministic GARCH type, we modify his sampling scheme to the case that allows a stochastic variance.

State Space Model and Prior Assumptions

We consider the ARMA-SV model,(3.12)-(3.15). Let $a \equiv (a_1, \cdots, a_p)'$: $p \times 1$, $b \equiv (b_1, \cdots, b_q)'$: $q \times 1$, $\omega_1 \equiv (\delta', a', b')'$: $(k+p+q) \times 1$, $\omega_2 \equiv (\gamma, \phi', \sigma_\nu^2)' : 3 \times 1$, $\omega \equiv (\omega_1', \omega_2')' : (k+p+q+3) \times 1$, $Y \equiv (y_1, \cdots, y_T)' : T \times 1$, $X \equiv (X_1', \cdots, X_T')' : T \times k$, and $h \equiv (h_1, \cdots, h_T)' : T \times 1$.

We consider a state space expression of the ARMA model (3.12)-(3.14) given the latent variable vector h:

$$y_t = X_t\delta + z\alpha_t + G_t\eta_t, \tag{3.A11}$$

$$\alpha_{t+1} = T\alpha_t + H_t\eta_t, \tag{3.A12}$$

where $z \equiv [1, 0, \cdots, 0] : 1 \times m$, $G_t \equiv h_t^{1/2}$,

$$T \equiv \begin{bmatrix} a_1 & & \\ a_2 & & I_{m-1} \\ \vdots & & \\ a_m & 0 \cdots & 0 \end{bmatrix} : m \times m, \quad H_t \equiv h_t^{1/2} \begin{bmatrix} a_1 - b_1 \\ a_2 - b_2 \\ \vdots \\ a_m - b_m \end{bmatrix} : m \times 1,$$

$m = \max\{p, q\}$, $a_j = 0$ for $j > p$, $b_j = 0$ for $j > q$, and $\eta_t \sim NID(0, 1)$. We assume $\alpha_0 = 0$. Obviously, $e_t = \sqrt{h_t}\eta_t$. This state space expression[5] is not the same as the typical expression of the ARMA model given in textbooks such as Harvey (1993) or Hamilton (1994). We can, however, easily verify that this expression reduces to equations (3.12)-(3.14).

We make the following assumptions.

Assumption 1 The latent variable vector h are generated by (3.15), and the data Y are generated by (3.12)-(3.14), with p,q known and non-stochastic X.

Assumption 2 All roots of $A(L) = 0$ lie outside of the unit circle.

Assumption 3 All roots of $B(L) = 0$ lie outside of the unit circle.

Assumption 4 $|\phi| < 1$.

Assumption 5 The prior distribution of the vector ω, $p(\omega)$, is given by

$$\begin{aligned}
p(\omega) &= p(\omega_1)p(\omega_2), \\
p(\omega_1) &= p(\delta)p(a)p(b), \\
&\quad \delta \sim N_k(\delta_0, D_0^{-1}), \quad a \sim N_p(a_0, A_0^{-1})I_{S_a}, \quad b \sim N_q(b_0, B_0^{-1})I_{S_b}, \\
p(\omega_2) &= p(\check{\phi}|\sigma_\nu^2)p(\sigma_\nu^2), \\
&\quad \check{\phi}|\sigma_\nu^2 \sim N_2(\phi_0, \sigma_\nu^2\Phi_0^{-1})I_{S_\phi}, \quad \sigma_\nu^2 \sim IG(\nu_0/2, s_0/2),
\end{aligned}$$

where $\check{\phi} = (\gamma, \phi)' : 2 \times 1$, N_i is the i-variate Gaussian distribution, IG is the inverted gamma distribution, I_A is the indicator function of the set A, S_a is the set of a that satisfies Assumption 2, S_b is the set of b that satisfies Assumption 3, S_ϕ is the set of ϕ that satisfies Assumption 4, the hyperparameters δ_0,D_0,a_0,A_0,b_0,B_0, ϕ_0,Φ_0,ν_0,and s_0 are known.

Assumption 2 and 3 are related to the stationarity and invertibility of the error term. Assumption 4 guarantees the stationarity of the latent variable $\ln h_t$.

Nakatsuma (1996) shows that y_t is given by

$$y_t = X_t\delta + \sum_{j=1}^{p} a_j(y_{t-j} - X_{t-j}\delta) + e_t - \sum_{j=1}^{q} b_j e_{t-j} + (a_t - b_t)e_0, \qquad (3.A13)$$

and y_t does not depend on e_0 for $t \geq m$ $(= \max\{p, q\})$. One of advantages of our state space expression is that we only need to obtain e_0, which is a

scalar, in order to start the recursion of the state space model. Chib and Greenberg (1994) derived a similar expression for a regression model with an ARMA(p,q) error. Their expression, however, is based on the initial state variable α_0, instead of the initial error term e_0. Another advantage of our state space model is that we can evaluate e_0 given the data, unobserved volatilities and the rest of the parameters without using any smoothing. Using equation (3.A13), we obtain the following equations:

$$\begin{aligned}
y_1 - X_1\delta &= (a_1 - b_1)e_0 + e_1, \\
y_2 - X_2\delta - a_1(y_1 - X_1\delta) &= (a_2 - b_2)e_0 + e_2 - b_1 e_1, \\
y_3 - X_3\delta - a_1(y_2 - X_2\delta) - a_2(y_1 - X_1\delta) &= (a_3 - b_3)e_0 + e_3 - b_1 e_2 - b_2 e_1, \\
&\vdots \qquad\qquad \vdots \\
y_T - X_T\delta - \sum_{j=1}^{p} a_j(y_{T-j} - X_{T-j}\delta) &= (a_T - b_T)e_0 + e_T - \sum_{j=1}^{q} b_j e_{T-j},
\end{aligned}$$

or

$$Pu = Ze_0 + Q\Sigma^{1/2}\eta,$$

where $u \equiv (y_1 - X_1\delta, y_2 - X_2\delta, \cdots, y_T - X_T\delta)' : T \times 1$, $Z \equiv (a_1 - b_1, a_2 - b_2, \cdots, a_T - b_T)' : T \times 1$, $\eta \equiv (\eta_1, \eta_2, \cdots, \eta_T) : T \times 1$, $\Sigma \equiv \mathrm{diag}\{h_1, h_2, \cdots, h_T\} : T \times T$,

$$P \equiv \begin{bmatrix}
1 & & & & & & \\
-a_1 & 1 & & & & & \\
-a_2 & -a_1 & 1 & & & & \\
\vdots & & & \ddots & & & \\
-a_p & \cdots & -a_2 & -a_1 & 1 & & \\
\vdots & & & & & \ddots & \\
0 & \cdots & -a_p & \cdots & -a_2 & -a_1 & 1
\end{bmatrix} : T \times T,$$

$$Q \equiv \begin{bmatrix} 1 & & & & & & \\ -b_1 & 1 & & & & & \\ -b_2 & -b_1 & 1 & & & & \\ \vdots & & & \ddots & & & \\ -b_q & \cdots & -b_2 & -b_1 & 1 & & \\ \vdots & & & & & \ddots & \\ 0 & \cdots & -b_q & \cdots & -b_2 & -b_1 & 1 \end{bmatrix} : T \times T,$$

and $\eta_t \sim NID(0,1)$. The estimate of e_0 by the GLS,

$$\hat{e}_0 = [Z'(Q\Sigma Q')^{-1}Z]^{-1}Z'(Q\Sigma Q')^{-1}Pu. \tag{3.A14}$$

As noted above, we can start the recursion of (3.A12) and (3.A13) once we obtain e_0 by (3.A14).

Full Conditional Distribution

The joint posterior of (h, ω) is given by the Bayes theorem,

$$\pi(h, \omega|Y) \propto p(Y|h, \omega_1)p(h|\omega_2)p(\omega)$$

where

$$\begin{aligned} p(h|\omega_2) &= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\sigma_\nu} \exp\left[-\frac{1}{2\sigma_\nu^2}(\ln h_t - \gamma - \phi \ln h_{t-1})^2\right], \\ p(Y|h, \omega_1) &= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi h_t}} \exp\left[-\frac{1}{2h_t}(y_t - y_{t|t-1})^2\right], \\ y_{t|t-1} &\equiv X_t\delta + (1 - A(L))(y_t - X_t\delta) + (B(L) - 1)e_t, \end{aligned}$$

and $y_{t|t-1}$ is the one-step-ahead prediction of y_t given X_t and information up to the period $t-1$. Breaking the joint posterior into various conditional distributions is the key to constructing the appropriate Markov chain

sampler. We simulate ω and h from the following conditional densities: $\pi(\delta|Y,\omega_{-\delta},h)$, $\pi(a|Y,\omega_{-a},h)$, $\pi(b|Y,\omega_{-b},h)$, $\pi(\check{\phi}|Y,\omega_{-\check{\phi}},h)$, $\pi(\sigma_\nu^2|Y,\omega_{-\sigma_\nu^2},h)$ and $\pi(h_t|y_t,\omega,h_{t-1},h_{t+1})$, where, e.g., $\omega_{-\delta}$ denotes all the parameters in ω other than δ.

We introduce two lemmas shown by Nakatsuma (1996). In the paper, he diagonalized the density $f(Y|\omega_1,h)$ by recursive transformations of the data, and produced a regression relationship for δ and a.

Definition 1 Let the scalars $y_0^* = u_0$, $y_s = y_s^* = 0$ for $s < 0$ and the vectors $X_s = X_s^* = 0$, $s \leq 0$. For t=1,...,T, define

$$\begin{aligned} y_t^* &= y_t - \sum_{i=1}^{p} a_i y_{t-i} - \sum_{j=1}^{q} b_j y_{t-j}^*, \\ X_t^* &= X_t - \sum_{i=1}^{p} a_i X_{t-i} - \sum_{j=1}^{q} b_j X_{t-j}^*. \end{aligned}$$

Using this definition, we obtain the following lemma:

Lemma 1 [Nakatsuma (1996, p.14)] Let Y^* be the $T \times 1$vector of the y_t^* and let the $T \times k$ matrix with X_t^* as its t th row. Then

$$p(Y^*|h,\omega_1) = (2\pi)^{-T/2}|\Sigma|^{-1/2}\exp\left[-\frac{1}{2}(Y^* - X^*\delta)'\Sigma^{-1}(Y^* - X^*\delta)\right].$$

Proof. Verify that $y_1^* - X_t^*\delta = e_1$ and proceed by induction, making use of

$$y_t^* - X_t^*\delta = y_t - X_t\delta - \sum_{i=1}^{p} a_i(y_{t-i} - X_{t-i}\delta) - \sum_{j=1}^{q} b_j(y_{y-j}^* - X_{y-j}^*\delta). \text{ ///}$$

If we note the regression relationship $Y^* = X^*\delta + e$, where $e \sim N_T(0,\Sigma)$, the full conditional distribution of δ immediately follows. By repeatedly applying transformations given in Definition 2 below, we arrive at the full conditional distribution of a.

Definition 2 Let $\tilde{y}_0 = e_0$, and let the scalars $y_s = \tilde{y}_s = 0$, $s < 0$. For $s < 0$, let the scalars $y_s = \tilde{y}_s = 0$. For $t = 1, \ldots, T$, define

$$\begin{aligned} \tilde{y}_t &= y_t - X_t\delta - \sum_{j=1}^{q} b_j \tilde{y}_{t-j}, \\ \tilde{X}_t &= (\tilde{y}_{t-1}, \cdots, \tilde{y}_{t-p}) : 1 \times p. \end{aligned}$$

With this definition we can prove the following lemma.

Lemma 2 [Nakatsuma(1996, p.15)] Let $\tilde{Y}$ be the $T \times 1$ column vector and let $\tilde{X} \equiv (\tilde{X}_1', \cdots, \tilde{X}_T')' : T \times p$, $\Sigma \equiv \text{diag}\{h_1, \cdots, h_T\} : T \times T$. Then

$$f(Y|h, \omega_1) = (2\pi)^{-T/2}|\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(\tilde{Y} - \tilde{X}a)'\Sigma^{-1}(\tilde{Y} - \tilde{X}a)\right].$$

Proof. Verify that $\tilde{y}_1 - \tilde{X}_1 a = e_1$ and proceed by induction, making use of that

$$\tilde{y}_t - \tilde{X}_t a = y_t - X_t\delta - \sum_{j=1}^{p} a_j(y_{t-j} - X_{t-j}\delta) - \sum_{j=1}^{q} b_j(\tilde{y}_{t-j} - \tilde{X}_{t-j}a). \text{ ///}$$

A corollary of this result is that $\tilde{Y} = \tilde{X}a + e$, where $e \sim N_T(0, \Sigma)$. Lemma 1 and 2 shown by Nakatsuma (1996) are expansions of the results of Chib and Greenberg (1994).

Let $LH = (\ln h_1, \cdots, \ln h_T)' : T \times 1$, $LX = \begin{bmatrix} 1 & \cdots & 1 \\ \ln h_0 & \cdots & \ln h_{T-1} \end{bmatrix}' : T \times 2$, $\mu_t = (\gamma(1-\phi) + \phi(\ln h_{t+1} + \ln h_{t-1}))/(1+\phi^2)$ and $\sigma_h^2 = \sigma_\nu^2/(1+\phi^2)$.

We now present the full conditional distributions that are used in the simulation of the regression model with ARMA(p, q)-SV errors.

Proposition 1 Under Assumptions 1-5, the full conditional distributions for $(\delta, a, b, \phi, \sigma_\nu)$ and h_t are given by

(i) $\delta|Y,\omega_{-\delta},h,\sim N_k\left((D_0+X^{*\prime}\Sigma^{-1}X^*)^{-1}(D_0\delta_0+X^{*\prime}\Sigma^{-1}Y^*),(D_0+X^{*\prime}\Sigma^{-1}X^*)^{-1}\right)$,

(ii) $a|Y,\omega_{-a},h\sim N_p\left((A_0+\tilde{X}'\Sigma^{-1}\tilde{X})^{-1}(A_0a_0+\tilde{X}\Sigma^{-1}\tilde{y}),(A_0+\tilde{X}'\Sigma^{-1}\tilde{X})^{-1}\right)I_{S_a}$,

(iii) $$\pi(b|Y,\omega_{-b},h)\propto\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi h_t}}\exp\left[-\frac{e_t(b)^2}{2h_t}\right]$$
$$\times(2\pi)^{-q/2}|B_0|^{1/2}\exp\left[-\frac{1}{2}(b-b_0)'B_0(b-b_0)\right]I_{S_b},$$

(iv) $\check{\phi}|Y,\omega_{-\check{\phi}},h\sim N_2\left((LX'LX+\Phi_0)^{-1}(LX'LH+\Phi_0\phi_0),\sigma_\nu^2(LX'LX+\Phi_0)^{-1}\right)I_{S_\phi}$,

(v) $$\sigma_\nu^2|Y,\omega_{-\sigma_\nu^2},h\sim IG\left((T+\nu_0+1)/2,((\check{\phi}-\phi_0)'\Phi_0(\check{\phi}-\phi_0)\right.$$
$$\left.+(LH-LX\check{\phi})'(LH-LX\check{\phi})+s_0)/2\right),$$

(vi) $$\pi(h_t|y_t,\omega,h_{t-1},h_{t+1})\propto h_t^{-1/2}\exp\left[-\frac{(y_t-y_{t|t-1})^2}{2h_t}\right]\exp\left[-\frac{(\ln h_t-\mu_t)^2}{2\sigma_h^2}\right].$$

Proof.

Given Y, $\omega_{-\delta}$, and h, the regression relationship $Y^*=X^*\delta+e$, where $e\sim N_T(0,\Sigma)$, is derived by Lemma 1. Thus the full conditional distribution of δ is given by (i). Similarly, (ii) follows from Lemma 2 and Assumption 1.

The full conditional distribution is proportional to the joint posterior of (h,ω), by its definition. Thus (iii), (v), and (vi) are given straight-forwardly. (iv) follows from the definition of the full conditional distribution and Assumption 3. ///

Finally, let us note that we may easily obtain the full conditional distributions if we use the vague priors instead of the informative priors of Assumption 5.

Assumption 6 The prior distribution $P(\omega)$ is given by

$$p(\omega) \;=\; I_{S_a}I_{S_b}I_{S_\phi}/\sigma_\nu^2.$$

Corollary 1 Under Assumptions 1-4 and 6, the full conditional distributions for $(\delta, a, b, \check{\phi}, \sigma_\nu)$ and h_t are given by

(i) $\delta|Y, \omega_{-\delta}, h, \sim N_k\left((X^{*\prime}\Sigma^{-1}X^*)^{-1}X^{*\prime}\Sigma^{-1}Y^*, (X^{*\prime}\Sigma^{-1}X^*)^{-1}\right)$,

(ii) $a|Y, \omega_{-a}, h \sim N_p\left((\tilde{X}'\Sigma^{-1}\tilde{X})^{-1}\tilde{X}\Sigma^{-1}\tilde{y}, (\tilde{X}'\Sigma^{-1}\tilde{X})^{-1}\right) I_{S_a}$,

(iii) $\pi(b|Y, \omega_{-b}, h) \propto \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi h_t}} \exp\left[-\frac{e_t(b)^2}{2h_t}\right] I_{S_b}$,

(iv) $\check{\phi}|Y, \omega_{-\check{\phi}}, h \sim N_2\left((LX'LX)^{-1}LX'LH, \sigma_\nu^2(LX'LX)^{-1}\right) I_{S_\phi}$,

(v) $\sigma_\nu^2|Y, \omega_{-\sigma_\nu^2}, h \sim IG\left((T+1)/2, (LH - LX\check{\phi})'(LH - LX\check{\phi})/2\right)$,

(vi) $\pi(h_t|y_t, \omega, h_{t-1}, h_{t+1}) \propto h_t^{-1/2}\exp\left[-\frac{(y_t - y_{t|t-1})^2}{2h_t}\right] \exp\left[-\frac{(\ln h_t - \mu_t)^2}{2\sigma_h^2}\right]$.

Proof. Similar to the Proof of Proposition 1. ///

Implementation Issue

Proposition 1 implies that while sampling from the full conditional distributions of $\delta, a, \check{\phi}, \sigma_\nu^2$ can be done by Gibbs sampling algorithm, sampling from b and h_t requires MH accept/reject algorithm.

Generation of b is a little more complicated since the error term, e_t, is a nonlinear function of b given h and ω_{-b}. Note that $e_t = 0$ for $t < 0$ by Definitions 1 and 2. To deal with this complexity, Chib and Greenberg (1994) proposed to linearize e_t by the first-order Taylor expansion:

$$e_t(b) \simeq e_t(b^*) + \psi_t(b - b^*), \tag{3.A15}$$

where

$$e_t(b^*) \equiv y_t^*(b^*) - X_t^*(b^*)\delta, \quad \psi_t = [\psi_{1t}, \cdots, \psi_{qt}] \equiv \left.\frac{\partial e_t}{\partial b}\right|_{b=b^*} : 1 \times q,$$

and ψ_t is the first order derivative of e_t evaluated at b^* given by the following recursion:

$$\psi_{it} = -e_{t-i}(b^*) - \sum_{j=1}^{q} b_j^* \psi_{i,t-j}, \quad (i = 1, \ldots, q)$$

where $\psi_{it} = 0$ for all $t \leq 0$. The remaining problem is at which point Taylor series expansion should be carried out. Nakatsuma (1996) used the following weighted nonlinear least squares estimator of b:

$$b^* \equiv \underset{b}{\operatorname{argmin}} \sum_{t=1}^{T} \frac{1}{h_t} \left(y_t^*(b) - X_t^*(b)\delta\right)^2, \tag{3.A16}$$

to approximate $p(Y|X, h, \omega_1)$. By assuming a natural conjugate prior for b, the approximation of the full conditional distribution of b is given as a truncated Gaussian distribution,

$$N\left((B_0 + \Psi'\Sigma^{-1}\Psi)^{-1}[B_0 b_0 + \Psi'\Sigma^{-1}(\Psi b^* - e(b^*))], (B_0 + \Psi'\Sigma^{-1}\Psi)^{-1}\right) I_{S_b} \tag{3.A17}$$

where $\Psi \equiv (\psi_1' \cdots, \psi_T')' : T \times q$ and $e = (e_1, \cdots, e_T)' : T \times 1$. We use it as a candidate generating density $g(b|Y, X, \omega_{-b}, h)$ in order to generate a candidate for b

Similarly, by assuming a flat prior for b, the approximation of the full conditional distribution of b is given as a truncated Gaussian distribution,

$$N\left([\Psi'\Sigma^{-1}\Psi]^{-1}\Psi'\Sigma^{-1}(\Psi b^* - e(b^*)), [\Psi'\Sigma^{-1}\Psi]^{-1}\right) I_{S_b}. \tag{3.A18}$$

Although Jacquier, Polson, and Rossi (1994) argued that it was difficult to find the bounding function of $\pi(h_t|y_t, X_t, \omega, h_{t-1}, h_{t+1})$, we can derive it

using the idea of Pitt and Shephard (1995). The logarithm of the density of $h_t|y_t, X_t, \omega, h_{t-1}, h_{t+1}$ is by (vi) of Proposition 1,

$$\ln(\pi(h_t|y_t, x_t, \omega, h_{t-1}, h_{t+1})) = \text{const} + \ln(p_t^*),$$

where

$$\ln(p_t^*) = -\frac{1}{2}\ln\, h_t - \frac{(y_t - y_{t|t-1})^2}{2h_t} - \frac{1}{2\sigma_h^2}(\ln\, h_t - \mu_t)^2. \tag{3.A19}$$

Let $h_t = \exp\, v_t$ and using the first-order Taylor series expansion of $\exp(-v_t)$ around $v_t^*(= \ln\, h_t^*)$, we obtain

$$\begin{aligned} \frac{1}{h_t} = \exp(-v_t) &\geq \exp(-v_t^*) - (v_t - v_t^*)\exp(-v_t^*) \\ &= \frac{1}{h_t^*} - (\ln\, h_t - \ln\, h_t^*)/h_t^*. \end{aligned}$$

Rewriting (24) using the above inequality, we have a bounding function,

$$\begin{aligned} \ln(p_t^*) &\leq -\frac{1}{2}\ln\, h_t - \frac{(y_t - y_{t|t-1})^2}{2}\left\{\frac{1}{h_t^*}(1 + \ln\, h_t^* - \ln\, h_t)\right\} - \frac{1}{2\sigma_h^2}(\ln\, h_t - \mu_t)^2 \\ &= \ln(g_t^*). \end{aligned}$$

The normalized version of g_t^* is a Gaussian density which has the mean and the variance

$$m_t^* = \mu_t + \frac{(y_t - y_{t|t-1})^2\sigma_h^2}{2h_t^*} - \frac{3\sigma_h^2}{2} \quad \text{and} \quad \sigma_h^2.$$

Hence, we can sample from $p(h_t|\cdot)$ by proposing $h_t \sim N(m_t^*, \sigma_h^2)$ and accepting with probability p_t^*/g_t^*. The remaining problem is at which point Taylor series expansion should be carried out. Watanabe (1996) choose h_t^* for efficient draw in the sense that it accepts the draw with higher probability. The closer to one is p_t^*/g_t^*, the more efficiently can we draw. Following Watanabe (1996), we first derive the value of h_t which corresponds to the peak of

p_t^* very roughly by iterating the Newton method for a few times. Then we choose h_t^* so that the peaks of p_t^* and g_t^* coincide with each other.

Empirical Example: Daily Exchange Rates

We developed a method of analyzing ARMA(p,q)-SV regression error models using the Markov chain Monte Carlo technique. In the following we use the same data in Section 3.4 and investigate performances of the Bayesian MCMC method. It should be noted that we do not intend to examine statistical difference between the Bayesian MCMC and the log-GARCH approach, since our MCMC method is developed under the Bayesian framework and since it is not an easy task to test the difference between the Bayesian and frequentist's results. Again, the data consist of daily yen/dollar close exchange rates, S_t, from the Tokyo interbank market. There are total of 1484 daily observations, from January 4, 1991 to December 30, 1996.

In spite of prefiltering the return series to take out AR terms and day-of-the-week effects in the mean returns, we consider the following ARMA(1,1)-SV model:

$$\begin{aligned}
\ln S_t &= \delta_0 + \delta_M D_{Mt} + \delta_T D_{Tt} + \delta_W D_{Wt} + \delta_R D_{Rt} + \delta_H Hol_t + a_1 \ln S_{t-1} + u_t, \\
u_t &= e_t - b_1 e_{t-1}, \\
e_t &= \sqrt{h_t}\eta_t, \quad \eta_t \sim \mathrm{NID}(0,1), \\
\ln h_t &= \gamma + \phi \ln h_{t-1} + \sigma_\nu \nu_t, \quad \nu_t \sim \mathrm{NID}(0,1),
\end{aligned}$$

where D_{Mt}, D_{Tt}, D_{Wt} and D_{Rt} are dummy variables for Monday, Tuesday, Wednesday and Thursday. Hol_t is the number of holidays (excluding weekends) between the $(t-1)$st and tth trading day. Taking account of the unit

root case for ϕ, we relax Assumption 4. Compared to the model in Section 3.4, this model is slightly extended with respect to the level specification, *i.e.*, we consider an ARMA(1,1)-SV model in $\ln S_t$ instead of the MA(1)-SV model in $r_t = \ln S_t - \ln S_{t-1}$. This specification enables us to check whether $\ln S_t$ is an $I(1)$ process.

For our implementation of the MCMC algorithm, we use QML estimates and standard errors via the log-GARCH approach as the prior information. The Markov chain sampler is run for 25000 draws such that the first m draws are discarded and then the next n $(= 25000 - m)$ are retained. All sample moments are computed as sample averages. It has sometimes been suggested that inferences should be based on every k th iteration of each sequence, with k set to some value high enough that successive draws of the parameter vector, ω, are approximately independent. This strategy can not be used in this situation since the set of simulated values is not so large that reducing the number of simulations by a factor of k gives important savings in storage and computation time.

Numerical standard errors are computed for the posterior mean by the batch means method described in Ripley (1987). In particular, the n simulated values were placed into v batches of n/v observations. The batch size is increased until the lag 1 correlation of the batch means is less than 0.05. The numerical standard errors are estimated as $s/\sqrt{v}$, where s is the standard deviation of the batch means. Alternatively, the spectral density approach of Geweke (1992) can be used to compute the numerical standard errors. Following Watanabe (1997b), We do not use Geweke's (1992) numerical standard error because of its sensitivity to the specification of the

spectral window.

Geweke (1992) has constructed a convergence diagnostic after $m+n$ iterations, where the first m draws discarded and next n draws retained. Though created for the Gibbs sampler, Geweke's (1992) method may be applied to the output of any MCMC algorithm; see, for example, Cowles and Carlin (1996). Markov chain sampler yields draws from the posterior only as the number of passes gets large, and hence comparison of early n_A passes with late n_B passes can reveal failures of convergence. Let $\omega_{\cdot}^{(i)}$ be the ith draw of a parameter and let

$$\begin{aligned} \bar{\omega}_{\cdot}^{A} &= \frac{1}{n_A}\sum_{j=1}^{n_A}\omega_{\cdot}^{(j)}, \\ \bar{\omega}_{\cdot}^{B} &= \frac{1}{n_B}\sum_{j=n_C}^{n}\omega_{\cdot}^{(j)}, \end{aligned}$$

where $n_C = n - n_B + 1$, and let nse_A and nse_B are the numerical standard errors for the two estimates, $\bar{\omega}_{\cdot}^{A}$ and $\bar{\omega}_{\cdot}^{B}$, calculated as in the previous paragraph. If the sequence of $\omega_{\cdot}^{(i)}$ is stationary, the ratios n_A/n and n_B/n are held fixed, and $n_A + n_B < n$, then by the central limit theorem, the distribution of the convergence diagnostic (CD),

$$(\bar{\omega}_{\cdot}^{A} - \bar{\omega}_{\cdot}^{B}) \Big/ \sqrt{\mathrm{nse}_A^2 + \mathrm{nse}_B^2}\,,$$

approaches a standard normal as n tends to infinity. Following the suggestion of Geweke (1992), we calculated this statistic by setting $n_A = 0.1n$ and $n_B = 0.5n$.

Table 3.5 presents the log-GARCH approach and the Bayes results for the ARMA-SV model. For the log-GARCH approach, the robust QML co-

Table 3.5: ARMA(1,1)-SV Models

Parameter	log-GARCH estimate	Posterior distribution Mean	Std. dev.	Lower 95% limit	Upper 95% limit	Corr.	CD
δ_0	0.01048	0.01048	0.00079	0.00916	0.00118	0.013	0.296
	(0.00081)	(0.000018)					
δ_M	-0.00027	-0.00017	0.00024	-0.00056	0.00021	0.009	1.314
	(0.00070)	(0.000006)					
δ_T	-0.00083	-0.00090	0.00025	-0.00131	-0.00049	0.014	0.921
	(0.00073)	(0.000007)					
δ_W	0.00027	0.00018	0.00026	-0.00025	0.00059	0.001	0.603
	(0.00063)	(0.000007)					
δ_R	0.00014	-0.00014	0.00026	-0.00057	0.00029	-0.026	0.969
	(0.00067)	(0.000013)					
δ_H	0.00162	0.00100	0.00032	0.00046	0.00154	0.070	0.271
	(0.00066)	(0.000009)					
a_1	0.9978	0.9978	0.00017	0.9955	0.9981	0.011	-0.577
	(0.0021)	(0.000004)					
b_1	-0.0184	-0.0006	0.00090	-0.0022	0.0008	0.343	-0.449
	(0.0011)	(0.000025)					
γ	-0.949	-1.008	0.00895	-1.023	-0.994	0.037	0.480
	(1.299)	(0.000265)					
ϕ	0.908	0.910	0.00075	0.908	0.911	0.018	-1.095
	(0.030)	(0.000036)					
σ_ν^2	0.0775	0.0293	0.00500	0.0226	0.0391	0.950	-0.243
	(0.0289)	(0.000674)					

Notes: Numerical standard error of posterior mean is in parentheses. Correlation denotes the first-order correlation of the Markov chain run. For the log-GARCH approach, standard error is in parentheses. 1500 simulations.

variance estimators of Bollerslev and Wooldridge (1992) are used to compute the standard errors. Note that they cannot be used to test whether σ_ν^2 is significantly different from zero. According to convergence diagnostics values, the null hypothesis that the sequence of 1500 samples is stationary cannot be rejected at five percent significance level for all parameters. Therefore $m = 23500$ and $n = 1500$ are chosen. With this statistics, we select $m = 23500$ and $n = 1500$. The marginal posterior means for δ_0, δ_M, δ_T, δ_W, δ_R, and δ_H are nearly zero. These results support the empirical findings of Hsieh (1988,1989) and Baillie and Bollerslev (1989). Marginal posterior mean of a_1 is nearly one, and this implies a unit root in $\ln S_t$. The marginal posterior means for SV parameters $(\gamma,\phi,\sigma_\nu^2)$ are $(-1.008,0.910,0.0275)$. Ruiz (1994) found much higher value for ϕ. It should be pointed out that the Ruiz (1994) study was conducted on October 1, 1981 to June 28, 1885, while our period is from January 4 1991, to December 30, 1996. The yen/dollar exchange rate data exhibit a high degree of persistence in volatility although posterior is massed well away from the unit root case.

In this appendix, We developed methods of analyzing ARMA(p,q)-SV regression error models in a Bayesian framework via the Markov chain Monte Carlo. Using the daily yen/dollar exchange rate, we showed that our Bayesian MCMC technique performs well.

Footnotes

∗1 This is the usual practice in papers that use ARCH models. See, *e.g.*, Bollerslev, Engle, and Nelson (1995).

*2 There are two more accurate computationally intensive smoothing procedures. Jacquier, Polson, and Rossi (1994) used a method based on Bayesian Markov chain Monte Carlo. Watanabe (1997a) proposed a smoothing procedure based on a nonlinear filtering method. Watanabe's (1997a) Monte Carlo evidence indicates that his nonlinear smoothing conditional on NFML estimates produces more efficient volatility estimates than the approximate Kalman filtering conditional on true parameters.

*3 See Bollerslev, Engle, and Nelson (1995, pp.2983-2984).

*4 Fiorentini, Calzolari, and Panattoni (1996) establish a computational advantage of analytic derivatives of the GARCH log-likelihood over the numerical differentiation. The advantage is large especially when second-order derivatives are involved in the computation.

*5 The data are offered by the Ueda Harlow co. for the academic purpose.

*6 The state space representations of this type are found in De Jong (1991), Koopman (1993), De Jong and Shephard (1995), Nakatsuma (1996), among others.

References

[1] Abramovitz, M. and N. Stegun, 1970, *Handbook of Mathematical Functions* (Dover Publications, NY).

[2] Amin, K.I. and A.J. Morton, 1994, "Implied volatility functions in arbitrage-free term structure models," *Journal of Financial Economics* **35**, 141–180.

[3] Andersen, T.G., 1996, "Return volatility and trading volume: an information flow interpretation of stochastic volatility," *Journal of Finance* **51**, 169–204.

[4] Andersen, T.G. and B.E. Sørensen, 1996, "GMM estimation of a stochastic volatility model: a Monte Carlo study," *Journal of Business and Economic Statistics* **14**, 328–353.

[5] Anderson, B.D.O. and J.B. Moore, 1979, *Optimal Filtering*, (Prentice-Hall, NJ).

[6] Andrews, D.W.K, 1991, "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica* **59**, 817–858.

[7] Andrews, D.W.K. and J.C. Monahan, 1992, "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator," *Econometrica* **60**, 953–967.

[8] Asai, M., 1998, "A new method to estimate stochastic volatility models: a log-GARCH approach," *Journal of the Japan Statistical Society* **28**, 101–114.

[9] Asai, M. and T. Shiba, 1996, "The Japanese stock market and the macroeconomy: an empirical investigation," *Financial Engineering and the Japanese Markets* **2**, 259–267.

[10] Asai, M., T. Takahashi, and T. Shiba, 1998, "The forecasting performance of models of interest futures: HJM models and others (in Japanese)," in K.

Morimune and T. Kariya, eds., *Risuku Kanri to Kinnyu-shyouken Toushi Senryaku, JAFEE Journal*, Vol.2 (Toyoh Keizai Shinpousha, Tokyo), 151–171.

[11] Baba, Y., 1995, "Estimation of the money demand function by error-correction models (in Japanese)," in Y. Honda, ed., *Nihon no Keiki* (Yuhikaku, Tokyo), 111–128.

[12] Baillie, R.T. and T. Bollerslev, 1989, "The message in daily exchange rates: a conditional-variance tail," *Journal of Business and Economic Statistics* **7**, 297–305.

[13] Balvers, R.J., T.F. Cosimano, and B. MacDonald, 1990, "Predicting stock returns in an efficient market," *Journal of Finance* **45**, 1109–1128.

[14] Banerjee, A., J. Dolado, J.W. Galbraith, and D.F. Hendry, 1993, *Co-integration, Error Correction, and the Econometric Analysis of Non-stationary Data* (Oxford University Press, Oxford).

[15] Blanchard, O.J., 1981, "Output, the stock market, and interest rates," *The American Economic Review* **71**, 132–143.

[16] Bollerslev, T., 1986, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics* **31**, 307–327.

[17] Bollerslev, T., 1987, "A conditional heteroskedastic time series model for speculative prices and rate of return," *Review of Economics and Statistics* **69**, 542–547.

[18] Bollerslev, T., R.F. Engle, and D.B. Nelson, 1995, "ARCH models," in R.F. Engle and D. McFadden, eds., *Handbook of Econometrics*, Vol.4 (North-Holland, Amsterdam), 2959–3038.

[19] Bollerslev, T. and J.M. Wooldridge, 1992, "Quasi maximum likelihood estimation and inference in dynamic models with time varying covariances," *Econometric Reviews* **11**, 143–172.

[20] Brennan, M.J. and E.S. Schwartz, 1980, "Analyzing convertible bonds," *Journal of Financial and Quantitative Analysis* **15**, 907–929.

[21] Brenner, R.J., R.H. Harjes, and K. Kroner, 1996, "Another look at models of the short-term interest rate," *Journal of Financial and Quantitative Analysis* **31**, 85–107.

[22] Campbell, J.Y. and R.J. Shiller, 1988, "Stock prices, earnings, and expected dividends," *Journal of Finance* **43**, 661–676.

[23] Campbell, J.Y. and P. Perron, 1991, "Pitfalls and opportunities: what macroeconomists should know about unit roots," in O.J. Blanchard and S. Fischer, eds., *NBER Macroeconomics Annual 1991* (MIT Press, Cambridge, MA).

[24] Chan, K.C., G.A. Karolyi, F.A. Longstaff, and A.B. Sanders, 1992, "An empirical comparison of alternative models of the short-term interest rate," *The Journal of Finance* **47**, 1209–1227.

[25] Chesney, M. and L.O. Scott, 1989, "Pricing European options: a comparison of the modified Black-Scholes model and a random variance model," *Journal of Financial and Quantitative Analysis* **24**, 267–284.

[26] Chib, S. and E. Greenberg, 1994, "Bayes inference in regression models with ARMA(p,q) errors," *Journal of Econometrics* **64**, 183–286.

[27] Clark, P.K., 1973, " A subordinated stochastic process model with finite variance for speculative prices," *Econometrica* **41**, 135–156.

[28] Cochrane, J.H., 1991, "Production–based asset pricing and the link between stock returns and economic fluctuations," *Journal of Finance* **46**, 209–237.

[29] Cowles, M.K. and B.P. Carlin, 1996, "Markov chain Monte Carlo convergence diagnostics: a comparative review," *Journal of the American Statistical Association* **91**, 863–904.

[30] Cox, J.C. and S.A. Ross, 1976, "The valuation of options for alternative stochastic processes," *Journal of Financial Economics* **3**, 145-166.

[31] Cox, J.C., J. Ingersoll, and S.A. Ross, 1980, "An analysis of variable rate loan contracts," *Journal of Finance* **35**, 389–403.

[32] Cox, J.C., J. Ingersoll, and S.A. Ross, 1985, "A theory of the term structure of interest rates," *Econometrica* **53**, 385–407.

[33] Cutler, D.M., J.M. Poterba, and L.H. Summers, 1991, "Speculative dynamics," *Review of Economic Studies* **48**, 529–546.

[34] Danielsson, J., 1994a, "Stochastic volatility in asset prices: estimation with simulated maximum likelihood," *Journal of Econometrics* **64**, 375–400.

[35] Danielsson, J., 1994b, "Bayesian analysis of stochastic volatilities: comment," *Journal of Business and Economic Statistics* **12**, 393–395.

[36] Danielsson, J. and J.F. Richard, 1993, "Quadratic acceleration for Monte Carlo likelihood evaluation," *Journal of Applied Econometrics* **8**, 153–174.

[37] Davidson, J.E.H., D.F. Hendry, F. Srba, and S. Yeo, 1978, "Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom," *Economic Journal* **88**, 661-692.

[38] De Jong, P., 1991, "The diffuse Kalman filter," *Annals of Statistics*, **19**, 1073–1083.

[39] De Jong, P. and N. Shephard, 1995, "The simulation smoother for time series models," *Biometrika* **82**, 339–350.

[40] Dickey, D.A. and W.A. Fuller, 1979, "Distributions of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association* **74** 427–431.

[41] Dolado, J.J., T. Jenkinson, S. Sosvilla-Rivero, 1990, "Cointegration and unit roots," *Journal of Economic Surveys* 4, 249–273.

[42] Dothan, U.L., 1978, "On the term structure of interest rates," *Journal of Financial Economics* **16**, 321–343.

[43] Engle, R.F., 1982, "Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation," *Econometrica* **50**, 987–1008.

[44] Engle, R.F. and T. Bollerslev, 1986, "Modeling the persistence of conditional variances," *Econometric Reviews* **5**, 1–50.

[45] Engle, R.F. and C.W.J. Granger, 1987, "Co-integration and error-correction: representation, estimation, and testing," *Econometrica* **55**, 251–276.

[46] Fama, E.F., 1990, "Stock returns, real returns, and economic activity," *Journal of Finance* **45**, 1089–1108.

[47] Fama, E.F. and K.R. French, 1988, "Dividend yields and expected stock returns," *Journal of Financial Economics* **22**, 3–25.

[48] Fama, E.F. and K.R. French, 1989, "Business conditions and expected returns on stocks and bonds," *Journal of Financial Economics* **25**, 23–49.

[49] Fiorentini, G., G. Calzolari, and L. Panattoni, 1996, "Analytic derivatives and the computation of GARCH estimates," *Journal of Applied Econometrics* **11**, 399–417.

[50] Fuller, W.A., 1976, *Introduction to Statistical Time Series* (John Wiley).

[51] Geweke, J., 1986, "Modeling the persistence of conditional variances: a comment," *Econometric Reviews* **5**, 57–61.

[52] Geweke, J., 1992, "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments," in J.M. Bernardo, J. Berger, A.P. David, and A.F.M. Smith, eds., *Bayesian Statistics*, Vol.4 (Oxford Univ. Press, Oxford), 169–193.

[53] Ghysels, E., A.C. Harvey, and E. Renault, 1996, "Stochastic volatility," in G.S. Maddala and C.R. Rao, eds., *Handbook of Statistics*, Vol.14 (Elsevier Science B.V.), 119–191.

[54] Granger, C.W.J., 1983, "Co-integrated variables and error-correcting models," Unpublished Discussion paper 83-13 (University of California, San Diego).

[55] Granger, C.W.J. and P. Newbold, 1973, "Spurious regression in econometrics," *Journal of Econometrics* **2**, 111-120.

[56] Hamilton, J.D., 1994, *Time Series Analysis*, (Princeton University Press, Princeton, NJ).

[57] Hansen, L.P., 1982, "Large sample properties of generalized method of moments estimators," *Econometrica* **50**, 1029–1054.

[58] Harvey, A.C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge University Press, Cambridge).

[59] Harvey, A.C., 1990, *The Econometric Analysis of Time Series*, 2nd ed. (Philip Allan).

[60] Harvey, A.C., 1993, *Time Series Models*, 2nd ed. (MIT Press, Cambridge, MA).

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

[62] Heath, D., R.A. Jarrow, and A.J. Morton, 1992, "Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation," *Econometrica* **60**, 77–105.

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

[64] Hsieh, D.A., 1988, "The statistical properties of daily foreign exchange rates: 1974–1983," *Journal of International Economics* **24**, 129–145.

[65] Hsieh, D.A., 1989, "Modeling heteroskedasticity in daily foreign-exchange rates," *Journal of Business and Economic Statistics* **7**, 307–317.

[66] Jacquier, E., N.G. Polson, and P.E. Rossi, 1994, "Bayesian analysis of stochastic volatilities," *Journal of Business and Economic Statistics* **12**, 371–389.

[67] Jacquier, E., N.G. Polson, and P.E. Rossi, 1995, "Multivariate and prior distributions for stochastic volatility models," Discussion paper CIRANO.

[68] Johansen, S., 1988, "Statistical analysis of cointegrating vectors," *Journal of Economic and Dynamic Control* **12**, 231–54.

[69] Johansen, S., 1991, "Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models," *Econometrica* **59**, 1551–1580.

[70] Kamizono, K. and T. Kariya, 1996, "An implementation of the HJM model with application to Japanese interest futures," *Financial Engineering and the Japanese Markets* **3**, 151–170.

[71] Keim, D.B. and R.F. Stambaugh, 1986, "Predicting returns in the stock and bond markets," *Journal of Financial Economics* **17**, 357–390.

[72] Kitagawa, G., 1987, "Non-Gaussian state space modeling of nonstationary time series," *Journal of the American Statistical Association* **82**, 1032–1063.

[73] Kitasaka, S., 1995, "On the worldwide stock price increase in the 1980's (in Japanese)," *Nihon Keizai Kenkyu* **28**, 1–30.

[74] Koopman, S.J., 1993, "Disturbance smoother for state space models," *Biometrika* **80**, 117–126.

[75] Kurozumi, E. and T. Yamamoto, 1997, "Bias correction in vector autoregressions with possibly integrated processes," Unpublished paper (Hitotsubashi University, Tokyo).

[76] Lee, S.W. and B.E. Hansen, 1994, "Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator," *Econometric Theory* **10**, 29-52.

[77] Melino, A. and S.M. Turnbull, 1990, "Pricing foreign currency options with stochastic volatility," *Journal of Econometrics* **45**, 239–265.

[78] Merton, R.C., 1973, "Theory of rational option pricing," *Bell Journal of Economics and Management Science* **4**, 141–183.

[79] Morimune, K. and G.Q. Zhao, 1995, "The unit root analyses of the causality between money and income," Institute of Economic Research Discussion Paper No.412 (Kyoto University, Kyoto).

[80] Nakatsuma, T., 1996, "Bayesian analysis of ARMA-GARCH models: a Markov chain sampling approach," Unpublished paper (Department of Economics, Rutgers University).

[81] Nelson, C.R. and M.J. Kim, 1993, "Predictable stock returns: the role of small sample bias," *Journal of Finance* **58**, 641–661.

[82] Nelson, D.B., 1988, "The time-series behaviour of stock market volatility and returns," Unpublished Ph.D. dissertation (MIT, Cambridge, MA).

[83] Nelson, D.B., 1990, "Stationarity and persistence in the GARCH(1,1) model," *Econometric Theory* **6**, 318–334.

[84] Nelson, D.B., 1991, "Conditional heteroskedasticity in asset returns: a new approach," *Econometrica* **59**, 347–370.

[85] Newey, W.K. and K.D. West, 1987, "A simple positive semi-definite heteroskedasticity and autocorrelation-consistent covariance matrix," *Econometrica* **55**, 703–708.

[86] Newey, W.K. and K.D. West, 1994, "Automatic lag selection in covariance matrix estimation," *Review of Economic Studies* **61**, 631–653.

[87] Osterwald-Lenum, M., 1992, "A note with fractiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics: Four cases," *Oxford Bulletin of Economics and Statistics* **54**, 461–72.

[88] Palm, F.C., 1996, "GARCH models of volatility," in G.S. Maddala and C.R. Rao, eds., *Handbook of Statistics*, Vol.14 (Elsevier Science B.V.), 209–240.

[89] Pantula, S.G., 1986, "Modeling the persistence of conditional variances: a comment," *Econometric Reviews* **5**, 79–97.

[90] Phillips, P.C.B., 1986, "Understanding spurious regressions in econometrics," *Journal of Econometrics* **33** 311–340.

[91] Phillips, P.C.B., 1995, "Fully modified leas squares and vector autoregression," *Econometrica* **63**, 1023–1078.

[92] Phillips, P.C.B. and P. Perron, 1988, "Testing for a unit root in time series regression," *Biometrika* **75**, 335–346.

[93] Pitt, M. and N. Shephard, 1995, "Parameter-driven exponential family models," Unpublished paper (Nuffield College, Oxford).

[94] Ripley, B.D., 1987, *Stochastic Simulation* (John Wiley & Sons).

[95] Ruiz, E., 1994, "Quasi-maximum likelihood estimation of stochastic volatility models," *Journal of Econometrics* **63**, 289–306.

[96] Schwert, G.W., 1989, "Stock returns and real activity: a century evidence," *Journal of Finance* **45**, 1237–1257.

[97] Scott, L.O., 1987, "Option pricing when the variance changes randomly: theory, estimation and an application," *Journal of Financial and Quantitative Analysis* **22**, 419–438.

[98] Shephard, N., 1995, "Fitting nonlinear time-series models with applications to stochastic variance models," in H.K. van Dijk, A. Monfort, and B. Brown, eds., *Econometrics Inference using Simulation Techniques* (John Wiley & Sons), 151–168.

[99] Shephard, N., 1996, "Statistical aspects of ARCH and stochastic volatility," in D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, eds., *Time Series Models in Econometrics, Finance and other Fields* (Chapman & Hall), 1–67.

[100] Shephard, N. and M.K. Pitt, 1997, "Likelihood analysis of non-Gaussian measurement time series," *Biometrika* **84**, 653–667.

[101] Shiba, T. and Y. Takeji, 1994, "Asset Price prediction using seasonal decomposition," *Financial Engineering and the Japanese Markets* **1**, 37–53.

[102] Siegel, J.J., 1991, "Does it pay stock investors to forecast the business cycle?," *Journal of Portfolio Management*, Fall issue, 27–34.

[103] Smith, A.F.M. and G.O. Roberts (1993), "Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods," *Journal of the Royal Statistical Society* B**55**, 3–24.

[104] Stock, J., 1987, "Asymptotic properties of least squares estimators of cointegrating vectors," *Econometrica* **55**, 1035–1056.

[105] Stock, J., 1995, "Cointegration, long-run comovements, and long-horizon forecasting," presented at the World Congress of the Econometric Society, Tokyo.

[106] Takeuchi, Y., 1991, "Trends and structural changes in macroeconomic time series," *Journal of the Japan Statistical Society* **21**, 13–25

[107] Tauchen, G. and M. Pitts, 1983, The price variability-volume relationship on speculative markets," *Econometrica* **51**, 485–505.

[108] Taylor, S.J., 1986, *Modeling Financial Time Series* (John Wiley, Chichester).

[109] Taylor, S.J., 1994, "Modeling stochastic volatility: a review and comparative study," *Mathematical Finance* **4**, 183–204.

[110] Tobin, J., 1969, "A general equilibrium approach to monetary theory," *Journal of Money, Credit and Banking* **1**, 15–29.

[111] Toda, H. and P.C.B Phillips, 1994, "Vector autoregression and causality," *Econometrica* **61**, 1367–1393.

[112] Toda, H. and T. Yamamoto, 1995, "Statistical inference in vector autoregressions with possibly integrated processes," *Journal of Econometrics* **66**, 225–250.

[113] Tsukuda Y. and T. Miyakoshi, 1995, "Granger causality between money and income for the Japanese economy in the presence of a structural change," Unpublished paper (Tohoku University, Miyagi).

[114] Vasicek, O.A., 1977, "An equilibrium characterization of the term structure," *Journal of Financial Economics* **5**, 177–188.

[115] Watanabe, T., 1993, "Time series properties of returns, volatility, and trading volume in financial markets," Unpublished Ph.D. dissertation (Department of Economics, Yale University).

[116] Watanabe, T., 1996, "Permanent and transitory component models of stock return volatility: quasi-maximum likelihood estimation and Bayes inference," Institute of Economic Research Discussion Paper No.434 (Kyoto University, Kyoto).

[117] Watanabe, T., 1997a, "A nonlinear filtering approach to stochastic volatility models with an application to daily stock returns," Unpublished paper (Tokyo Metropolitan University, Tokyo).

[118] Watanabe, T., 1997b, "Option pricing under stochastic volatility: a Bayesian Markov chain Monte Carlo simulation approach," presented at the 1997 Biwako Keiryoukeizaigaku Kaigi (Biwako Econometrics Conference), Shiga-ken, Japan.

[119] Watson, M.W., 1994, "Vector autoregressions and cointegration," in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics*, Vol.4 (North-Holland, Amsterdam), 2843–2915.

[120] White, H., 1982, "Maximum likelihood estimation of misspecified models," Econometrica **50**, 1–25.

[121] Wiggins, J.B., 1987, "Option values under stochastic volatility: theory and empirical estimates," *Journal of Financial Economics* **19**, 351–372.

[122] Yamada, H. and H. Toda, 1997, "A note on hypothesis testing based on the fully modified vector autoregressions," *Economics Letters* **56**, 27–39.

[123] Yamada, H. and H. Toda, 1998, "Inference in possibly integrated vector autoregressive models: finite sample evidence," *Journal of Econometrics* **86**, 55 95.

[124] Yoshida, T. and R.H. Rasche, 1990, "The M2 demand in Japan: shifted and unstable ?," *Bank of Japan Monetary and Economic Studies* **8**, September, 9–30.

[125] Young, S.D., M.A. Berry, D.W. Harvey, and J.R. Page, 1991, "Macroeconomic forces, systematic risk, and financial variables: an empirical investigation," *Journal of Financial and Quantitative Analysis* **26**, 559–564.