# STOCHASTIC VOLATILITY:

# ESTIMATION AND EMPIRICAL VALIDITY

Thesis submitted to the University of London for the degree of

Doctor of Philosophy

by

Gleb Sandmann

Department of Accounting and Finance

London School of Economics and Political Science

University of London

March 1997

UMI Number: U103601

UMI

Dissertation Publishing

ProQuest

THESES

F
7434

# Abstract

Estimation of stochastic volatility (SV) models is a formidable task because the presence of the latent variable makes the likelihood function difficult to construct. The model can be transformed to a linear state space with non-Gaussian disturbances. Durbin and Koopman (1997) have shown that the likelihood function of the general non-Gaussian state space model can be approximated arbitrarily accurately by decomposing it into a Gaussian part (constructed by the Kalman filter) and a remainder function (whose expectation is evaluated by simulation). This general methodology is specialised to the estimation of SV models. A finite sample simulation experiment illustrates that the resulting Monte Carlo likelihood estimator achieves full efficiency with minimal computational effort. Accurate values of the likelihood function allow inference within the model to be performed by means of likelihood ratio tests. This enables tests for the presence of a unit root in the volatility process to be constructed which are shown to be more powerful than the conventional unit root tests.

The second part of the thesis consists of two empirical applications of the SV model. First, the informational content of implied volatility is examined. It is shown that the in-sample evolution of DEM/USD exchange rate volatility can be accurately captured by implied volatility of options. However, better forecasts of ex post volatility can be constructed from the basic SV model. This suggests that options implied volatility may not be market's best forecast of the future asset volatility, as is often assumed. Second, the regulatory claim of a destabilising effect of futures market trading on stock market volatility is critically assessed. It is shown how volume-volatility relationships can be accurately modelled in the SV framework. The variables which approximate the activity in the FT100 index futures market are found to have no influence on the volatility of the underlying stock market index.

3

**To my parents**

# Table of contents

# List of Tables

# List of Figures

# Chapter 1: Modelling financial time series

## 1.1. Introduction

Understanding the dynamics of stock market returns and other financial time series has been of a considerable interest in the finance literature for a long time. Bachelier (1900) initiated the idea of the stochastic nature of stock returns by first noting, that the geometric Brownian motion may be a reasonable approximation. The body of knowledge has grown immensely since then and some stylised facts have been established.

The evidence suggests that the empirical distribution of most financial time series differs substantially from distributions obtained from sampling independent homoscedastic Gaussian variables. Unconditional density functions exhibit leptokurtosis and skewness; time series of asset returns show evidence of volatility clustering whereas little or no serial dependence can be detected in the return process itself (Fama, 1970, 1991, Lo and MacKinlay, 1988; Pagan, 1996).

As an illustrative example, consider the daily time series of the FTSE100 UK stock market index over the 27/10/86-14/12/95 period. In what follows the first difference of log-prices, $R_t = ln(P_t/P_{t-1})$, will be conventionally referred to as the 'return' series.
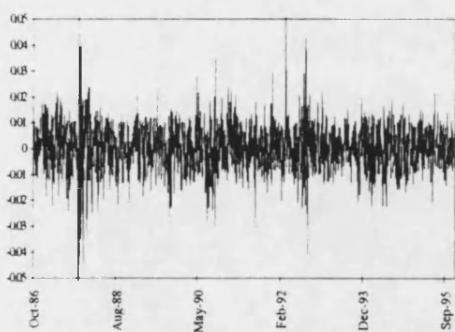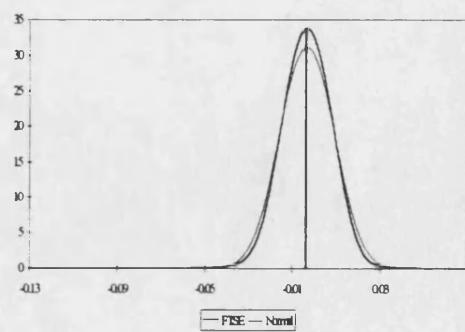


*(a)*                                    *(b)*
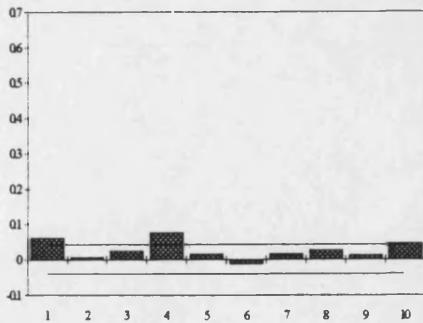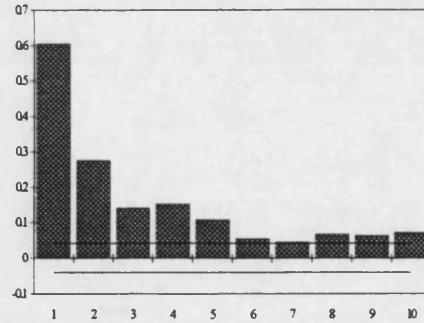
*(c)*                                         *(d)*

**Figure 1.1:** FTSE100: returns, $r_t$; density; correlogram of $r_t$; correlogram of $r_t^2$.

The picture which emerges is typical of most financial data. First, as the time plot **Figure 1.1***(a)* reveals, intervals of large price movements are followed by more quiet periods. This time varying volatility behaviour became known as the "volatility clustering" phenomenon. Secondly, **Figure 1.1***(b)* allows us to compare the estimate of the unconditional density of returns with the corresponding normal density (which is calibrated by equating the first two unconditional moments). The empirical density has fatter tails, and a much higher peak around zero than the corresponding normal distribution. While it may be difficult to detect the fatter tails from the picture above, note that the range was chosen so as to correspond to the minimum and maximum of the observed values, events whose probability of occurrence under the Gaussian assumption would be zero. This indicates that there are many more small and large returns than predicted by the Gaussian approximation. Weak evidence of serial dependence in the mean is provided by **Figure 1.1***(c)* which documents the correlogram of $R_t$ for lags up to 10 together with the corresponding $\pm 2T^{-0.5}$ bounds. And finally, very strong serial dependence in the squares of the process is documented in **Figure 1.1***(d)*. Taken together, these empirical regularities rule out the geometric Brownian motion as a feasible model for asset returns.

What can be done if the Brownian paradigm is rejected? The econometric literature can be seen as being divided into two categories. Models in the first category assume that returns are driven by a process with some fixed, *time homogenous* distribution. More recently,

10

there has been an emphasis on *intertemporally dependent* models. The special feature of models in the latter category is the dependence of the distribution on previous realisations of the process. In particular, the volatility parameter is made to follow an evolution of its own, thus shifting the whole conditional distribution. Since models of latter category form the core of the study, the description of the time-homogenous models will be brief and is included here for the sake of completeness.

The unifying framework which provides the rationale for modelling asset returns by means of time homogenous *as well* as intertemporally dependent models is provided by the Mixture of Distributions Hypothesis (MDH). Here, the central idea is formed by the assertion that the rate information arrival is non-constant over time, and possibly stochastic. As will be seen in the next Section, most of the time series models for asset returns can be regarded as special cases of MDH. Since the MDH forms the *raison d'etre* for the time series models employed in our empirical work, a brief exposition is required before we can proceed any further.

## 1.2. Information arrival and Subordinated Stochastic Processes

The Mixture of Distributions Hypothesis (MDH) appears several times in various disguises in the literature. Rather than following the development of MDH in chronological order, the exposition here will commence with the most recent treatment of the model by Andersen (1996) and provide connections to earlier work, whenever necessary. This is largely due to the fact that this particular derivation is well justified by microeconomic arguments and does not have the *ad hoc* character of previous approaches.

Andersen's model is a version of Glosten and Milgrom (1985) microstructure model in which a single asset with a random liquidation value is traded between three types of

risk-neutral agents: a market maker, and two groups of informed and uninformed investors. During the trading process informed investors obtain private signals as to the value of the asset. The crucial property of the equilibrium is that a sequence of trades fully reveals the content of all available private information leading to a (temporary) equilibrium price. This price prevails until new private information arrives and the dynamic learning process starts anew. Thus the market moves from one temporary equilibrium to the next in response to a large number of information arrivals during each trading period.

To fix ideas, denote by $P_{j,t}$, the transaction price recorded during the $j$-th temporary equilibrium of the trading period $t$ (e.g. day). Let furthermore, the random variable symbolising the total number of information arrivals during each trading period be denoted as $J_t$. Then the return over the full trading period can be decomposed as:

$$R_t = \ln\left(\frac{P_{J_t,t}}{P_{0,t}}\right) = \sum_{j=1}^{J_t}\ln\left(\frac{P_{j,t}}{P_{j-1,t}}\right) = \sum_{j=1}^{J_t}\omega_{j,t} , \qquad \omega_{j,t} \sim IID(0,\sigma_\omega^2)$$

If, $\omega_{j,t}$ are independent of $J_t$ then an appropriate generalisation of the standard central limit theorem (Clark, 1973) delivers the desired result:

$$R_t|J_t \sim N(0,\sigma_\omega^2 J_t) \tag{1.1}$$

which shows that the returns are conditionally normal but have variances that reflect the intensity of information arrival. In some trading periods little news is released and trading is slow, with incremental price movements. When new, possibly lumpy, information arrives, trading is fast and prices fluctuate strongly until a new equilibrium is attained.

Notice, that the argument above delivers almost instantly the justification for the time varying volatility models which are discussed in Section 1.4. In particular, if $J_t$ is a dynamic stochastic process - e.g. a mean reverting Ornsten-Uhlenbeck process (Karlin and Taylor,

12

1981) - one obtains directly the stochastic volatility model. However, time homogenous models can be equally obtained from the MDH by observing that that $J_t$ does not have to be a dynamic variable and can be assumed to be drawn form a particular time invariant distribution.

An interesting alternative point of view can be obtained by regarding (1.1) as a paricular example of a subordinated stochastic process. Denote by $\{X_t\}_{t\geq 0}$ a Wiener process and $\{J_t\}_{t\geq 0}$ some other stochastic process. Then the process $\{R_t = X[J_t]\}_{t\geq 0}$ is called the subordinated and $\{J_t\}_{t\geq 0}$ the directing process (Feller, 1971, Chapter 17). Let the conditional variance of the subordinated process, given the filtration (information structure) $\Im_{t-1}$, be denoted by $Var[R_t|\Im_{t-1}] = \sigma_t^2$. If the directing process $\{J_t\}_{t\geq 0}$ is constant over time, the return process $\{R_t\}_{t\geq 0}$ is itself a Brownian Motion with drift. Its variance only depends on the length of the time interval over which the return is measured, i.e. $Var[R(\Delta t)|\Im_t] = \sigma^2 \Delta t$. However, if the directing process is non-constant, possibly stochastic then the variance of $R_t$ will depend upon the number and importance of events occurring within the time interval: $Var[R(\Delta t)|J(\Delta t), \Im_t] = \sigma^2 J(\Delta t)$.

The essence of subordinated stochastic processes can be best understood in terms of a model of uncertainty. If the 'state of nature' is a complete description of the economy up to time t, $J(\Delta t)$ may be regarded as a stochastic change in state over the interval $[t, t+\Delta t]$. This is due to the change in the stock of information available to the market participants. In the classical theory of consumption under uncertainty knowledge of the state is sufficient for the knowledge of the state dependent variable. Here, only the *distribution* of the state dependent variable $R(\Delta t)$ is known.[1]

---

[1] This argument leads to the idea of *time deformation* (Stock, 1988) where the process is regarded as being driven by the event time scale and not the calendar time scale. An interesting application can be found in e.g. Ghysels and Jasiak (1996).

Thus, non-constant, possibly stochastic information arrival forms the core of the justification for the usage of models which transcend the Brownian terrain. If the latent variable is drawn from a particular time homogenous distribution a number of interesting models for the evolution of asset returns can be obtained. On the other hand, if the latent variable is assumed to follow a dynamic process, intertemporally dependent time series models immediately follow.

## 1.3. Early approaches: time homogenous models

Fama (1965) and Mandelbrot (1963, 1967) considered a class of so-called Stable Paretian distributions as a potential model describing asset returns. The model has two interesting features: it captures the fat tails property of the unconditional densities and comprises the normal distribution as a special case. Mandelbrot and Taylor (1967) show that if the directing process $J(\Delta t)$ follows a strictly positive stable distribution then the subordinated process $R(\Delta t)$ will follow a symmetric stable distribution. The model has been fitted to data and was found to describe data reasonably well (Fama and Roll, 1971). In general, however, its density function is unknown and the variance is infinite thus rendering the model rather unappealing.

A more promising approach was suggested by Blattberg and Godenes (1974) who consider a model for stock returns where the variance parameter of a normal density is drawn from an inverted Gamma distribution, i.e. $\{\sigma^2 J(\Delta t)\}^{-1}$~Gamma. The resulting conditional distribution of the return process $\{R_t\}$ is no longer normal but *Student-t*. The evidence indicates that this process is a better description of returns dynamics than the Stable Paretian model (Blattberg and Godenes, 1974). In particular, the fat tail property of the unconditional density can be addressed by adjusting the number of degrees of freedom. Alternatively, the variance to be drawn directly from the Gamma distribution (Madan and

Seneta, 1990). An analytical representation of the conditional density of the return process could not be found in this case but the empirical analysis suggested that this model is a better fit than the lognormal distribution.

Another interesting model has been proposed by Kon (1984). Here the return process is described by a mixture of normal distributions such that each period return is drawn independently from one of a finite number of independent normal distributions. Kon (1984) argues that this specification is a better description of stock returns than either the *Student-t* model or the Stable Paretian model.

Finally, Press (1967), Merton (1976), and Ball and Torous (1983), among others investigate the applicability of mixed jump-diffusion model: $dP = \mu P dt + \sigma dW + P dq$, where $W$ is the standardised Wiener, $q$ a Poisson process with intensity $\lambda$, and the distribution of the size of the jump is normal $J \sim N(\mu_J, \sigma^2_J)$. In an empirical study of exchange rate futures Tucker, Madura and Marshall (1994) find that this specification dominates the mixture-of-normals model which in turn dominates the Stable Paretian specification. However, they disregard the issues of parsimony and compare models on the basis of the likelihood function value only. Using the Schwarz criterion, which corrects for the presence of additional parameters, Kim and Kon (1994) in an analysis of stock returns, find that the Student-*t* model dominates both, the mixed jump-diffusion and the mixture-of-normals models.

Whatever the relative merits of individual time homogenous models one important empirical characteristic of asset returns cannot be captured by any of the models presented above: the phenomenon termed "volatility clustering". As indicated in the introduction one of the most salient features of financial time series is the fact that intervals of large price movements (i.e. high volatility) are often followed by more quiet periods, with small price fluctuations. This can be interpreted as direct evidence that the latent information flow

variable $J_t$ follows a process whose evolution is time-varying thus giving rise to ARCH and stochastic volatility models described in the following Section.

## 1.4. Time dependent models: ARCH vs. stochastic volatility

Intertemporally dependent models recognise that conditional mean *and* variance of asset returns, $R_t=ln(P_t/P_{t-1})$ may change over time:

$$R_t = \mu_t + \sigma_t v_t , \qquad v_t \sim IID(0,1) \tag{1.2}$$

where the evolution of $\mu_t$ and $\sigma_t$ depend on the previous realisations of the process. Models in this category can be divided into two groups: those imposing a strict functional relationship on the variance evolution (ARCH models) and those treating the variance as an unobserved component following some stochastic process (stochastic volatility, or SV models).

Despite large differences in the set-up, the two approaches can be conceptually reconciled. Notice, that framework of subordinated processes encompasses both, the ARCH and the SV models. Since little is known about the rate of information arrival apart from its time dependence, it is natural to assume that the process is stochastic, possibly mean-reverting. This yields immediately the SV model. Specific assumptions about the information arrival process give the ARCH model (Gallant, Hsieh and Tauchen, 1991). Thus, one resolution to the question about model adequacy is to regard SV models as generalisations of ARCH. Alternatively, one can regard ARCH models as discrete time *approximations* to a diffusion process followed by latent, unobserved volatility, while the SV models are simple *discretisations* of this diffusion. It can be shown (Nelson, 1990a) that the Exponential GARCH model converges to a continuous time SV model as the distance between the observations becomes smaller. By implication, ARCH models can be used to estimate parameters of a continuous time stochastic volatility model (Engle and Lee, 1994; Nelson

and Foster, 1994). Dassios (1992), however, shows that even though both discrete time SV and EGARCH models will converge to the same continuous time limit, the rate of convergence of the SV model is faster.

While ARCH models dominated the finance literature in the past decade, insights into the properties and estimation of SV models were made very recently. In fact, empirical applications considered in this Thesis illustrate that the SV models offer an interesting alternative to GARCH models routinely used in applied empirical research. The empirical question of superior descriptive ability remains unresolved. The prime difficulty with direct comparison is the fact that the models are non-nested so that the standard likelihood ratio tests are inapplicable (Non-nested likelihood ratio tests were examined by Kim, Shephard and Chib, 1996). In addition the conclusions remain largely data dependent. Overall, the evidence seems to suggest that both model specifications perform equally well when fitted to data (Andersen, 1994a; Taylor, 1994; Kim, Shephard and Chib, 1996).

An altogether different approach is to regard the volatility process as being generated by a discrete state Markov chain (Kim, 1993; Naik, 1993; Rockinger, 1994; Billio and Monfort, 1995). Unlike diffusion models, where the variable is allowed to take a continuum of values, here the volatility variable is assumed to be in one of the *discrete* states, say $\sigma_L$ and $\sigma_H$. The motivation is provided by the idea that major events may have drastic (but temporary) effects on asset's risk characteristics, thus periodically shifting the volatility level.[2] While models of this kind might be conceptually appealing they have a number of drawbacks. First, mere two volatility levels are likely to be inadequate. When more states are included the analytical tractability of the framework is lost. On the other hand, it is conceivable that the model converges to the continuous state space SV model as the number of admissible states increases. The analytical tractability is also lost when bivariate

asset return processes are considered or the regression effects of explanatory variables are
to be investigated (Rockinger, 1994). Whatever the merits of volatility switching models,
the present research is focused on the empirical validity of continuous state space time
varying volatility models mentioned above, the statistical properties of which are reviewed
below.

### 1.4.1. Statistical aspects of ARCH models

ARCH (AutoRegressive Conditional Heteroscedasticity) models were first introduced by
Engle (1982) and achieved widespread popularity over the last decade. Generalisations of
the original specification gave rise to GARCH, EGARCH, IGARCH, FIGARCH, ARCH-
M, and many others. It is not the aim of this survey to compare and discuss all formulations
proposed in the literature since several extensive surveys are available (Bollerslev, Chow
and Kroner, 1992; Bera and Higgins, 1993; Bollerslev, Engle and Nelson, 1994). Instead
we concentrate on some selected examples of ARCH models which have either proved to
be more descriptive or provide for a comparison to the stochastic volatility models
discussed in the next section.

*(i) Definition.*

The ARCH(p) model (Engle, 1982) imposes a fixed functional form on the specification of
the variance parameter $\sigma^2_t$ in (1.2) making it dependent on past squared residuals of the
mean equation:

$$\sigma^2_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i r^2_{t-i} , \qquad r_t = R_t - \hat{\mu}_t \qquad (1.3)$$

---

[2] This is different from the model proposed by Hamilton and Susmel (1994) where the parameters of
an ARCH process come from different regimes.

Since many lags are required to describe persistence in conditional variance Bollerslev (1986) and Taylor (1986) proposed a more parsimonious representation, the GARCH(p,q) model:[3]

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2 \qquad (1.4)$$

In an attempt to accommodate the negative correlation between returns and return volatility (Black, 1976; Christie, 1982) into the ARCH framework Nelson (1991) proposed the Exponential GARCH model:

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \, g(r_{t-i}) + \sum_{j=1}^{q} \beta_j \ln \sigma_{t-j}^2 \qquad (1.5)$$

where $g(\cdot)$ is some asymmetric function of $r_t$. Further interesting extensions have been suggested by Engle, Lilien and Robbins (1987) who propose to model the time varying risk premium by including the standard deviation as an explanatory variable in the mean equation:

$$R_t = \delta \sigma_t + \sigma_t v_t, \qquad v_t \sim N(0,1) \qquad (1.2')$$

with $\sigma_t$ following some ARCH parameterisation. Finally, it is well known that the assumption of conditional normality of the error term is often not sufficient to account for the leptokurtosis of the unconditional density observed in real data (Engle and Bollerslev, 1986; Baillie and Bollerslev, 1989; Teräswirta, 1996). Instead, the GARCH-$t$ model (Bollerslev, 1987) may be more appropriate:

$$R_t = \mu_t + \sigma_t v_t, \qquad v_t \sim t_v \qquad (1.2'')$$

Here, the mean equation errors are drawn from a Student-$t$ distribution with $v$ degrees of freedom. Clearly, the conditionally Gaussian model is obtained in the limit, as $v \rightarrow \infty$.

---

[3] Mirroring the relationship between MA($\infty$) and ARMA(p,q) model (1.4) can be represented as

Using the law of iterated expectations the unconditional moments of a GARCH process can be derived. Any GARCH model produces serially independent observations with mean zero. If the sum of the coefficients $\alpha_i$ and $\beta_j$ is less than unity then the unconditional distribution has fat tails (leptokurtosis) and finite variance. Finally, the model has been artificially constructed so as to mimic the volatility clustering, producing serially correlated squared observations. Thus, almost all of the above mentioned stylised facts can be simulated in the ARCH framework.

Numerous extensions have been proposed accommodating more and more empirical regularities, extending the taxonomy of the ARCH literature and making the framework increasingly popular. The framework owns its popularity to primarily two factors. First, the models allow for tractable specification of the exact functional form of the conditional variance function. Secondly, ARCH models are designed to formulate explicitly the likelihood function as well as analytic scores.

*(ii) Schwert models.*

Another ARCH-type model (in the sense that the variance equation is described in terms of some function of lagged residuals) has been proposed by Schwert (1990). Here two regression equations are formulated which describe the evolution of the mean and the volatility of the process in terms of lagged endogenous variables:

$$
\begin{aligned}
R_t &= c + \sum_{i=1}^{I} \phi_i R_{t-i} + \sum_{j=1}^{J} \pi_j \hat{\sigma}_{t-j} + u_t \\
\hat{\sigma}_t &= \alpha + \sum_{m=1}^{M} \beta_m \hat{\sigma}_{t-m} + \sum_{l=1}^{L} \delta_l \hat{u}_{t-l} + \varepsilon_t
\end{aligned}
\tag{1.6}
$$

where $R_t$ represents the return on the asset, $\sigma_t$ is the instantaneous standard deviation of the residuals $u_t$, and circumflexes indicate fitted values from a previous iteration. The residuals from one regression are used as observations in the other, and the system

ARCH($\infty$).

20

is re-estimated until convergence is achieved. Despite simple estimation possibilities, model (1.6) is highly unparsimonious (Engle, 1990) and enjoyed limited popularity (Bessembinder, 1994; Bessembinder and Seguin, 1992, 1993).

*(iii) Estimation.*

We now turn briefly to the estimation of GARCH models. Let $\Psi=(\psi_m, \psi_v)$ denote the set of unknown mean, $\psi_m$ (in (1.2), (1.2'), or (1.2''), above) and variance equation parameters, $\psi_v$ (in either of (1.3), (1.4), or (1.5) above). Then the likelihood function of the model can be written down explicitly since, conditional on the past, the density of the observations is Gaussian with a known volatility parameter. Thus the full likelihood function is given by the product of conditionally Gaussian densities, $p_G(\cdot)$, with means, $m_t$, and variances, $\sigma_t^2$:

$$L(R_t \mid \Psi) = \prod_{t=1}^{T} p_G\left(m_t, \sigma_t^2 \mid \mathfrak{I}_{t-1}\right)$$

One of the crucial features of the model is the asymptotic independence of the mean and variance equation parameters. In other words the information matrix is block diagonal with respect to the two parameter sets. The OLS in the mean equation is thus consistent. The residuals are then used to obtain the estimates of the variance equation parameters by optimising the likelihood function by means of the BHHH algorithm (Berndt, Hall, Hall and Haussman, 1974).

Analytical scores can be obtained by conditioning on initial values $\sigma_0=0$ and $r_0=0$. Assuming the boundedness of conditional fourth moments, $E(r_t^4 \mid \mathfrak{I}_{t-1})$ the maximum likelihood estimator $\psi^*$ is asymptotically normal and consistent (Weiss, 1986). Even if the data are not conditionally normal, maximum likelihood, now called quasi maximum likelihood, is still an efficient procedure. Lee and Hansen (1994) showed that in fact weaker restrictions are sufficient for such estimators to be consistent.

*(iv) Explanatory variables.*

As will be seen in due course, one attractive feature of GARCH (as well as SV models) is the possibility of examining the impact of explanatory variables on the evolution of volatility. Here, in addition to the lagged values of $r_t^2$ the functional form of the conditional variance of an ARCH model can be made dependent on a set of exogenous variables $Z_t=(z_{1t},\ldots,z_{kt})'$. For instance, the GARCH(p,q) model can be amended as:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + Z_t'\gamma \qquad (1.7)$$

where $\gamma$ is a $(k\times 1)$ vector of unknown parameters. Such parameterisation has been advocated by numerous researchers, e.g. Lamoureux and Lastrapes (1990, 1993), with a single notable exception: Baillie and Bollerslev (1989) who estimate the model (1.9) discussed below.

Specification (1.7) imposes a very rigid structure in which the explanatory variables are allowed to affect volatility. In particular the effect of $Z_t$ on the variance of the process is geometrically declining. To see this effect, let $\xi_t = r_t^2 - \sigma_t^2 = \sigma_t^2(v_t^2 - 1)$ be the 'innovation' in the conditional variance process. For illustrative purposes the discussion below will focus on the GARCH(1,1) case. Extensions to GARCH(p,q) and EGARCH(p,q) are straightforward. Equation (1.7) can be reformulated as:

$$(1 - \varphi L)r_t^2 = \alpha_0 + Z_t'\gamma + (1 - \beta_1 L)\xi_t$$

where $L$ is the backshift operator, $Lx_t=x_{t-1}$ and $\varphi=\alpha_1+\beta_1$. Since the conditional expectation of $\xi_t$ is zero by construction, in fact, $\xi_t$ is white noise, taking expectations repeatedly we obtain the expression for the unconditional variance of $r_t$:

$$Er_t^2 = \frac{\alpha_0}{1-\varphi} + \frac{Z_t'\gamma}{(1-\varphi L)} = \frac{\alpha_0}{1-\varphi} + \sum_{i=1}^{\infty} \varphi^{i-1} L^{i-1} Z_t'\gamma \qquad (1.8)$$

Thus, $Z_t$ in model (1.7) is no longer an estimate of the variance of the process at any time. Instead, the variance is influenced by an exponentially weighted average of past values of explanatory variables. This is an important drawback since, as will become evident later, we will be interested in modelling the instantaneous impact of the explanatory variables $Z$ on the evolution of volatility.

In order to avoid this feed-back effect, (1.8) indicates that it is necessary to re-parameterise the model; in the GARCH(1,1) case:

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + (1 - \varphi L) Z_t' \gamma \qquad (1.9)$$

which is the model estimated by Baillie and Bollerslev (1989).[4]

*(v) Multivariate extensions.*

Some difficulties with the ARCH specification arise when the transition from the univariate to multivariate case is made. Even though exact likelihood function can be obtained, the proliferation of parameters poses a serious restriction on the applicability of the models. In addition, elaborate restrictions need to be imposed on parameter values to ensure that the ($N \times N$) matrix of volatilities, $\Sigma_t$ is positive definite for all time (Engle and Kroner, 1995). The general parameterisation of the conditional variance $\Sigma_t = Var(r_t | \Im_{t-1})$ of an $N$-dimensional GARCH(1,1) model is:

$$\text{vech}(\Sigma_t) = \text{vech}(A_0) + A\,\text{vech}(r_{t-1} r_{t-1}') + B\,\text{vech}(\Sigma_{t-1}) \qquad (1.10)$$

where *vech(·)* denotes the column stacking operator of the lower portion of a symmetric matrix, $r_t$ is ($N \times 1$), $A_0$ is ($N \times N$), while A and B are ($N(N+1)/2 \times N(N+1)/2$) matrices. This formulation involves the estimation of 21 (*sic!*) parameters if the case of a bivariate model, $N=2$. A more parsimonious specification is:

---

[4] However, (1.9) makes the gradients highly non-linear in the parameters creating difficulties in estimation.

$$\Sigma_t = A_0 + C' r_{t-1} r_{t-1} ' C + D' \Sigma_{t-1} D \tag{1.11}$$

where $r_t$ is ($N \times 1$), while $A_0$, C and D are all ($N \times N$) matrices. Provided $A_0$ is positive definite specification (1.11) ensures that $\Sigma_t$ is p.d.. The number of unknown parameters can thus be reduced to 12 for the case of $N=2$.

Assuming that the matrix of correlations between individual elements of the return vector, $r_t$ is constant over time, Bollerslev (1990) developed a more tractable and empirically viable (Baillie and Bollerslev, 1990) specification:

$$\Sigma_t = diag(\sqrt{\Sigma_{11,t}}, ..., \sqrt{\Sigma_{NN,t}}) R \, diag(\sqrt{\Sigma_{11,t}}, ..., \sqrt{\Sigma_{NN,t}}) \tag{1.12}$$

where R is an ($N \times N$) correlation matrix and the individual variances, $\Sigma_{ii}$ are standard GARCH(p,q) processes. This model involves the estimation of mere 7 parameters in the GARCH(1,1) case and $N=2$. Finally, factor ARCH models were proposed by Engle, Ng and Rothschild (1990) and are discussed in Lin (1992) and Shephard (1996).

Summing up, the generalisation to multivariate ARCH models is not obvious: the models proposed are either highly constrained or not parsimonious. As will be shown in the next section, SV models generalise to the multivariate case much more naturally.

## 1.4.2. Statistical aspects of stochastic volatility models

An alternative way of modelling changes in conditional variance is to allow it to evolve stochastically over time. Rather than imposing a prespecified functional form on the variance function, $\sigma_t^2$ in (1.2) is assumed to be driven by some unobserved, latent factor. Unlike ARCH models, knowledge of the parameters and past realisations of the process is not sufficient to determine with certainty the value of $\sigma_t^2$ at any one time: the variance is now an unobserved component. The exposition below will focus on some selected aspects

of stochastic volatility (SV) modelling since, again several surveys are available (Taylor, 1994; Ghysels, Harvey and Renault, 1996; Shephard, 1996).

*(i) Definition.*

The most widely used SV model is the lognormal. The statistical model in discrete time is defined by the mean equation (1.2), whence the variance equation is given by:

$$\ln \sigma_t^2 = \alpha + \phi \ln \sigma_{t-1}^2 + \eta_t, \quad \eta_t \sim N\!\left(0, \sigma_\eta^2\right), \quad Corr\!\left(v_t, \eta_t\right) = \rho \qquad (1.13)$$

This specification highlights the similarity to the EGARCH(1,1) model (1.5) but instead of some function of lagged values of $r_t$, the evolution of the variance is now driven by an additional disturbance, $\eta_t$. It is often convenient to re-parameterise (1.13) as:

$$\begin{cases} \ln \sigma_t^2 = \ln \overline{\sigma}^2 + h_t \\ h_t = \phi\, h_{t-1} + \eta_t \end{cases} \quad \eta_t \sim N\!\left(0, \sigma_\eta^2\right), \quad Corr\!\left(v_t, \eta_t\right) = \rho \qquad (1.13')$$

The equivalence is established by noting that the constants capturing the long run (log)variance levels are mapped via $\ln \overline{\sigma}^2 = \alpha(1 - \phi)^{-1}$.

Other definitions have been proposed in the context of option pricing. Various authors have examined SV models formulated in continuous time, taking the general form:

$$\begin{cases} d\ln S = \mu_S dt + \sigma dW_1 \\ df(\sigma) = \mu(\sigma)dt + \upsilon(\sigma)\sigma_\sigma dW_2 \end{cases}, \quad d\langle W_1, W_2 \rangle = \rho$$

where $dW$ denotes a standard Wiener process and the specifications of the functions $f(\sigma)$, $\mu(\sigma)$, and $\upsilon(\sigma)$ are summarised in:

| | volatility process | Study |
|---|---|---|
| *M1* | $d\sigma = \lambda(\kappa - \sigma)dt + \sigma_\sigma dW_2$ | Scott (1987) Stein and Stein (1987) |
| *M2* | $d\sigma^2 = \lambda(\kappa - \sigma^2)dt + \sigma_\sigma \sqrt{\sigma^2} dW_2$ | Heston (1993) |

| M3 | $\dfrac{d\sigma}{\sigma} = \lambda dt + \sigma_\sigma dW_2$ | Hull and White (1987) |
| --- | --- | --- |
| | | Johnson and Shanno (1987) |
| M4 | $d\ln\sigma = \lambda(\kappa - \ln\sigma)dt + \sigma_\sigma dW_2$ | Scott (1987) |
| | | Chesney and Scott (1989) |
| | | Wiggins (1987) |

All these specifications postulate a Markovian structure on the volatility process and ensure the intuitive condition - that the volatility process $\{\sigma_t\}_{t \geq 0}$ be positive almost surely - is satisfied. While these various definitions are admissible, the time series literature focuses almost exclusively on the statistical aspects of the parameterisation (1.13) which is the discrete time analogue of Model *M4*. The correspondence is established by defining $\phi = e^{-\lambda\Delta t}$, the speed of mean reversion, $\ln\bar{\sigma}^2 = 2\kappa$, the long run volatility level, and $\sigma_\eta^2 = \lambda^{-1}\sigma_\sigma^2\left(1 - e^{-2\lambda\Delta t}\right)$, the variance parameter (Gouriéroux and Monfort, 1996, p. 125; Renault and Touzi, 1996). The predominance of (1.13) is due to the fact that the model can be estimated by various methods (Chapter 2) and captures most of the regularities found in financial data.

The statistical properties of the SV model are valid even if $v_t$ and $\eta_t$ are contemporaneously correlated. First, if $\phi$ is less than unity in absolute value the return process, $r_t$, will be stationary with the even moments being given by:

$$E\left(r_t^j\right) = \frac{j!\exp\left(\frac{1}{2}jm_1 + \frac{1}{8}j^2m_2\right)}{2^{\frac{j}{2}}\left(\frac{j}{2}\right)!}, \qquad m_1 = \frac{\alpha}{1-\phi}, \qquad m_2 = \frac{\sigma_\eta^2}{1-\phi^2}$$

while the odd moments are zero. It follows that the variance and the kurtosis are:

$$E\left(r_t^2\right) = \exp\left(\frac{\alpha}{1-\phi} + \frac{\sigma_\eta^2}{2(1-\phi^2)}\right), \qquad \frac{E\left(r_t^4\right)}{\left(\sigma_{r^2}^2\right)^2} = 3\exp\left(\frac{\sigma_\eta^2}{1-\phi^2}\right) \geq 3 \qquad (1.14)$$

As can be seen from (1.14) the kurtosis is larger than that of the normal density, even if $\phi=0$. When $\rho=0$ the model implies further that the squares of the process are autocorrelated in accordance with:

$$E\left(r_t^2, r_{t-i}^2\right) = \exp(2m_1 + m_2)\left(\exp\left(\phi^i m_2\right) - 1\right)$$

which translates into an exponentially decaying autocorrelation function (ACF). More general results concerning the ACF of $|r_t|^c$ for arbitrary constants $c$ and extensions to $t$-distributed mean equation errors - as in (1.2″) - can be found in Ghysels, Harvey and Renault (1996).

Similarly to GARCH models the SV model can be extended in a number of directions likely to be of importance in applied empirical research. Thus, the leverage effect (Black, 1976; Christie, 1982) is captured automatically by a non-zero correlation coefficient between the two disturbances, $v_t$ and $\eta_t$. Furthermore, the time varying risk premium can be modelled by including the volatility as an explanatory variable in the variance equation (1.2′). Finally, additional leptokurtosis may be incorporated in the basic SV model by allowing the mean equation residuals to follow a Student $t_v$ distribution as in (1.2″).

*(ii) Estimation.*

The main drawback of SV models has been the difficulty associated with parameter estimation. Likelihood-based estimation requires the latent volatility process, $\sigma_t$ to be integrated out of the joint density of the observed returns, $r_t$, and latent volatilities, $\sigma_t$. This fundamental difficulty has preserved the widespread popularity enjoyed by GARCH models.

Chapter 2 is entirely devoted to SV estimation methodology where a new estimation technique is proposed. It should be noted here, however, that if the parameter vector $\Psi$ is partitioned into mean, $\psi_m$, and variance equation parameters, $\psi_v$, and the volatility process

is strictly stationary, OLS in the mean equation is inefficient but consistent. Thus, the mean equation parameters, $\psi_m$, can be estimated by OLS but their variance-covariance matrix needs to be adjusted for the effect of heteroscedasticity. Suppose the mean equation takes the form:

$$R_t = x_t'\beta + \sigma_t v_t, \quad v_t \sim N(0,1) \tag{1.2'''}$$

where $x_t$ is a ($k \times 1$) vector of explanatory variables, possibly including lagged values of $R_t$, and $\beta$ (i.e. $\psi_m$ in the previous notation) is a ($k \times 1$) vector of unknown coefficients. Harvey and Shephard (1993) show that the variance of the feasible GLS estimator is given by:

$$Var(\tilde{\beta}) = \left[\sum_{t=1}^{T} x_t x_t' e^{-h_{t|T}}\right]^{-1} \sum_{t=1}^{T} x_t x_t' e^{-2h_{t|T}} \left(R_t - x_t'\tilde{\beta}\right)^2 \left[\sum_{t=1}^{T} x_t x_t' e^{-h_{t|T}}\right]^{-1} \tag{1.15}$$

where $h_t = \ln\sigma_t^2$ and the notation $h_{t|T}$ symbolises the smoothed estimate of $h_t$. Therefore, the estimation of the SV model proceeds in tree steps. First, OLS is applied to the mean equation to obtain the mean adjusted returns, $r_t$. Secondly, the parameters of the stochastic volatility process are estimated, for instance by means of the method proposed in Chapter 2. And finally, the covariance matrix of mean equation coefficients is obtained via (1.15).

*(iii) Explanatory variables.*

Similarly to GARCH models, the basic SV model (1.12) can be extended so as to allow for regression effects in the variance equation. However, unlike GARCH models the SV model is formulated in terms of a time varying deterministic mean which is approximated by a weighted average of explanatory variables $z^k_t$'s and an autoregressive component:

$$\begin{cases} \ln\sigma_t^2 = Z_t'\gamma + h_t \\ h_t = \phi h_{t-1} + \eta_t, \quad \eta_t \sim N(0,\sigma_\eta^2) \end{cases} \tag{1.16}$$

where $Z_t = (z_{1t}, \ldots, z_{kt})'$ and $\gamma$ is a ($k \times 1$) vector. In fact, model (1.13') is obtained as a special case by setting $k=1$, $z_t'=1$, $\forall t$. The statistical properties are not difficult to derive. In

particular, the unconditional variance of the process is given by:

$$E r_t^2 = E \varepsilon_t^2 \; E(\sigma_t^2) = \exp\left( Z_t{}'\gamma + \frac{\sigma_\eta^2}{2(1-\phi^2)} \right)$$

(1.17)

Thus, instantaneous variance is described by the information contained in the explanatory variables at that moment in time, rather than some weighted average of past values. Clearly, if this information is irrelevant the coefficients $\gamma_k$ will be insignificant. On the other hand, if the autoregressive coefficient, $\phi$ is insignificant the explanatory variables capture accurately the dynamics of conditional variance. In this case the model reduces to the multiplicative heteroscedasticity model (Harvey, 1976) with noise: $\ln\sigma_t^2 = Z_t{}'\gamma + \eta_t$, $\eta_t \sim N(0,\sigma_\eta^2)$. If neither hypothesis can be rejected, we are left with a mean reverting volatility process where the mean is approximated by the information encoded in $z_t^k$'s.

If, for some reason, the geometric lag structure is specifically required, then the SV can be adjusted accordingly:

$$\begin{cases} \ln\sigma_t^2 = h_t \\ h_t = Z_t{}'\gamma + \phi\, h_{t-1} + \eta_t \quad \eta_t \sim N(0,\sigma_\eta^2) \end{cases}$$

(1.18)

so that model (1.13) is obtained as a special case by setting $k=1$, $z_t^1=1$, $\forall t$. Equation (1.18) is the SV analogue of (1.7) and the resulting unconditional variance is:

$$E r_t^2 = \exp\left( \frac{Z_t{}'\gamma}{1-\phi L} + \frac{\sigma_\eta^2}{2(1-\phi^2)} \right)$$

(1.19)

which - upon expansion - leads to an expression similar to (1.8).

*(iv) Multivariate extensions.*

Stochastic volatility models generalise to the multivariate case by allowing the logarithm of variance to follow an AR(1) process. This leads to the following parameterisation of the mean adjusted returns, $r_t$:

$$\begin{cases} r_{i,t} = \overline{\sigma}_i \, e^{0.5 h_{i,t}} \, v_{i,t} & i = 1, \ldots, N \qquad v_t \sim NID\!\left(0, \Omega_v\right) \\ h_t = \Phi h_{t-1} + \eta_t & \eta_t \sim NID\!\left(0, \Sigma_\eta\right) \end{cases} \qquad (1.20)$$

Where $r_{i,t}$ are the individual elements of the ($N\times1$) vector $r_t$, $\overline{\sigma}$ is a ($N\times1$) vector capturing the long run volatility levels, $\Phi$ is a ($N\times N$) non-diagonal matrix of mean reversion coefficients, and $\Sigma_\eta$ is a ($N\times N$) covariance matrix of innovations in the ($N\times1$) log-variance vector, $h_t$. Notice, that the vector of mean equation disturbances is multivariate normal with a constant correlation matrix, $\Omega_v$. Thus, similarly to Bollerslev's (1990) GARCH model, (1.20) is a model of changing covariancies but constant correlation.

The closer $\Phi$ is to the identity matrix the more persistence is present in the volatility process, or the slower mean reversion. If the eigenvalues of $\Phi$ are within the unit circle the model is strictly stationary. Differences in volatility adjustment mechanisms will therefore be reflected entirely in the elements of $\Phi$. The diagonal elements $\{\phi^{ii}\}$ give the degree of persistence while the off-diagonal elements $\{\phi^{ij}\}$ indicate cross market dependence.

The formulation allows for common trends in volatility by placing reduced rank restrictions on $\Sigma_\eta$. Thus, if rank($\Sigma_\eta$)=$K<N$ then there are only $K$ components in volatility so that

$$h_t = \Theta \tilde{h}_t \qquad (1.21)$$

where $\Theta$ is an ($N \times K$) matrix of factor loadings and $\tilde{h}_t$ follows a VAR(1). When the volatilities are assumed to follow a multivariate random walk, i.e. $\Phi = I_K$, (1.21) is replaced by

$$h_t = \Theta \tilde{h}_t + \bar{h} \qquad\qquad (1.21')$$

where $\bar{h}$ is a vector of constants. As we can see, despite a parsimonious parameterisation, multivariate SV models have a potential to be empirically viable.

## 1.5. Outline of the thesis

Having established the relative advantages and highlighted the similarities between the ARCH and SV frameworks, we will now focus our attention on estimation of SV models. A novel algorithm for the estimation of SV models is proposed in Chapter 2. It is demonstrated that the finite sample performance of the estimator is on par with the fully efficient Bayesian MCMC method. The extensions of the basic SV model are addressed.

Chapter 3 discusses some aspects of hypothesis testing within the model. The likelihood ratio test for the presence of the unit root in the (log)variance process is considered. However, despite the possibility of model estimation under the null and under the alternative the distribution of the test statistic in unknown. Our Monte Carlo experiments suggest that the distribution can be well approximated by the weighted $\chi^2$ density, critical values of which are readily available. It is also shown that the conventional unit root tests are unreliable in this context.

The empirical validity of the SV model is illustrated in the subsequent Chapters. It is shown how a number of interesting empirical questions can be addressed in this framework by extending the basic SV model to include a set of explanatory variables. The hypothesis that implied volatility of options contain relevant information about the evolution of the

latent return volatility process is examined in Chapter 4. It is found that *in-sample* the implied volatility captures most of the time series dynamics of the conditional volatility of the return process. However, the *out-of-sample* forecasting experiment suggests that the predictions from the basic SV model across all forecasting horizons are at least as accurate as the forecasts obtained from the implied volatility data.

In Chapter 5, the hypothesis that futures trading destabilises the corresponding spot market, leading to an increase in price volatility is examined. It is shown that when the explanatory variables are included as in (1.7) - as has been done on numerous occasions in the literature of volume-volatility relationship - results are potentially misleading. Using the SV model, no evidence for the UK in support of the hypothesis (that futures trading destabilises the spot market) is found.

# Chapter 2: Estimation of stochastic volatility models

## 2.1. Introduction

Despite their intuitive appeal, SV models have been used less frequently than ARCH models in empirical applications. This is due to the difficulties associated with the estimation of SV models. Unlike ARCH models, where the likelihood function can be evaluated exactly, the likelihood function of an SV model is hard to construct. Existing estimation procedures can be subdivided into two groups: (*i*) methods that attempt to build the full likelihood function, and (*ii*) methods which rely on alternative, usually less efficient principles.[1]

Several propositions have been made as to how the likelihood function may be evaluated. Kim, Shephard and Chib (1996) show how the likelihood can be constructed when a mixture of normals is used to approximate the density of the disturbances. Jacquier, Polson and Rossi (1994) have proposed a Bayesian approach to the estimation of SV models using the Markov Chain Monte Carlo (MCMC) technique. Fridman and Harris (1996) show how the extended Kalman filter can be used to perform numerical integration. Finally, Danielsson (1994a) suggested that accurate approximations to the likelihood function can be obtained by means of importance sampling.

Building on the work of Durbin and Koopman (1997a) and Shephard and Pitt (1997) - who have designed importance sampling methods for general state space models - this Chapter shows how the general concept of importance sampling can be used efficiently in the context of SV model estimation. The crucial feature is the formulation of the SV model in a linear state space form with $\ln(\chi_t^2)$ disturbances in the measurement

---

[1] The Quasi-Maximum Likelihood (QML) method of Harvey, Ruiz and Shephard (1994), and GMM methods of Andersen and Sørensen (1996) are examples of this category.

equation. The linear state space form allows very powerful algorithms for filtering and smoothing to be utilised, and more generally, to draw upon a vast body of knowledge on structural time series models. The gain is also due to the fact that the Monte Carlo simulation is only employed to construct that residual part of the likelihood function, which is not already captured by the Gaussian likelihood (QML), which we know to be an inefficient but close approximation.

Apart from reducing the computational effort considerably (while attaining full finite sample efficiency), the algorithm has two distinct advantages. First, the sampling variation can be reduced, giving arbitrarily close approximations to the true likelihood function. Availability of the accurate values of the likelihood function allows for hypothesis testing by means of the likelihood ratio tests. This is likely to be very useful since numerical standard errors of model parameters often leave much to be desired. This area is further explored in Chapter 3 where the likelihood ratio test for the presence of the unit root in the (log)variance equation is shown to be more powerful than the conventional unit root tests. Second, a wide range of extensions can be addressed without any modifications of the estimation procedure due to the fact that the state space form is retained. Thus, the variance of the return process can be examined for the presence of serial correlation, seasonal components and trends, and the effects of dummy and exogenous explanatory variables may be explored in detail. The remaining Chapters are devoted to empirical applications. Furthermore, the method can be extended to multivariate models, an area, where sampling techniques like MCMC as well as variants of the Method of Moments become cumbersome.

This Chapter is organised as follows. Section 2.2 discusses in more detail the various aspects of estimation and inference in the context of SV models. In Section 2.3 we describe the new estimation algorithm while Section 2.4 compares its finite sample

performance with existing techniques by means of a Monte Carlo experiment. Section

2.5 illustrates how the method should be adjusted when the basic SV model is

extended in a number of directions. Here, the question of how to treat zero

observations - which will make $\ln(\chi_1^2)$ ill-defined - is also addressed. Finally, Section

2.6 extends the method into the multivariate context. Section 2.7 concludes.

## 2.2. Existing approaches

The univariate stochastic volatility model was introduced earlier in (1.13) and (1.13')

and is restated here for ease of reference:

$$\begin{cases} r_t = \bar{\sigma} e^{\frac{1}{2}h_t} v_t & v_t \sim N(0,1) \\ h_t = \phi h_{t-1} + \eta_t & \eta_t \sim N(0,\sigma_\eta^2) \end{cases} \quad Corr(v_t, \eta_t) = \rho \qquad (2.1)$$

As before, $r_t$ is the mean adjusted return on an asset, $r_t = R_t - \hat{\mu}_t$. The estimation of $\mu_t$

has been addressed in Section 1.4.2, equation (1.15), and will not be the subject of

interest in the present context. An attractive feature of specification (2.1) is the

possibility of linearising the model. By taking logarithms of the squared mean adjusted

returns one obtains:

$$\begin{cases} \ln r_t^2 = \ln \bar{\sigma}^2 + h_t + \varepsilon_t & \varepsilon_t = \ln v_t^2 \\ h_t = \phi h_{t-1} + \eta_t & \eta_t \sim N(0,\sigma_\eta^2) \end{cases} \quad Corr(\varepsilon_t, \eta_t) = 0 \qquad (2.2)$$

If the original mean equation disturbance, $v_t$, is standard normal, $\varepsilon_t$ follows the $\ln(\chi_1^2)$

distribution whose mean and variance are known to be -1.27 and $\pi^2/2$, respectively.

Notice, that even if the mean and variance equation disturbances $v_t$ and $\eta_t$ are

correlated, the transformed disturbances, $\varepsilon_t$ are uncorrelated with $\eta_t$. Therefore the

information regarding $\rho$ is lost when the transformation is taken (Harvey and Shephard

(1996) show how it can be recovered by conditioning on the signs of $r_t$).

Harvey, Ruiz and Shephard (1994) suggested a Quasi-Maximum Likelihood (QML) method of estimating the model based on the Kalman filter. Assuming joint conditional normality of $(\varepsilon_t, \eta_t)$, equation (2.2) represents the measurement and transition equations of the general linear state space model:

$$
\begin{aligned}
y_t &= Z_t \alpha_t + \varepsilon_t & \varepsilon_t &\sim N(0, H_t) \\
\alpha_t &= T_t \alpha_{t-1} + \eta_t & \eta_t &\sim N(0, Q_t)
\end{aligned}
\qquad t = 1, \ldots, T \qquad (2.3)
$$

where, in general, $y_t$ is $(N \times 1)$ vector of observations, $\alpha_t$ is the $(m \times 1)$ state vector, and the covariance matrices $H_t$ and $Q_t$ are non-singular. Appendix 2 illustrates how the unknown variance equation parameters $\psi_v = (\phi, \sigma_\eta, \bar{\sigma})$, henceforth $\psi$, are placed in the system matrices $H_t$, $Q_t$, $T_t$, $Z_t$. Once the model is in the state space form, the advantages of this approach become evident: *(i)* explanatory variables can be easily incorporated into the variance equation, *(ii)* more complicated process can be assumed for the evolution of the latent variable, *(iii)* missing or irregularly spaced observations can be handled, and *(iv)* generalisations to the multivariate case are straightforward. The QML estimation method involves a numerical optimisation of the Gaussian (log)likelihood function over the set of parameters $\psi$ details of which can be found in Appendix 3.

QML approximates the distribution of $\varepsilon_t$ by $N(-1.27, \pi^2/2)$, while $\varepsilon_t$ is far from being Gaussian. In fact, its density is given in:

**Proposition 2.1**: Let the scalar variable $x$ be standard normal. Then the density of $z = \ln x^2$ is:

$$
p_{\ln \chi_1^2}(z) = \frac{1}{\sqrt{2\pi}} \exp\left[ \frac{z - e^z}{2} \right] \qquad (2.4)
$$

**Proof**: The univariate $\chi^2$ density with $\nu$ degrees of freedom is given by:

$$p_{\chi^2_\nu}(y) = \frac{y^{\frac{1}{2}\nu-1}e^{-\frac{y}{2}}}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}$$

Setting $\nu=1$, noting that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, and making the change of variables $z=\ln y$ gives the desired result. $\square$

**Figure 2.1** shows in how far $\varepsilon_t$ deviates from its normal approximation which implies that the QML estimator is likely to have poor small sample properties even though it is consistent.
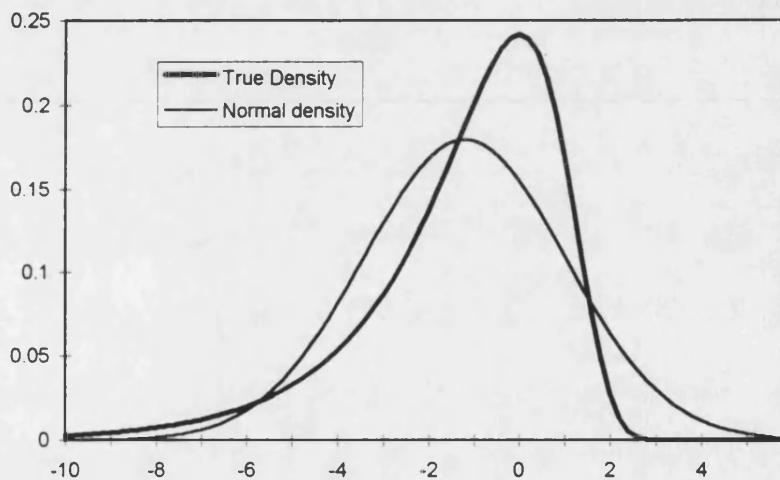


**Figure 2.1**: The $\ln(\chi_1^2)$ density and the Gaussian approximation $N(-1.27, \pi^2/2)$.

Note the high degree of skewness and the long tail on the negative half line. Large negative values reflect small values of $r_t$ - termed inliers - which may arise in empirical applications with high frequency data.

Several other estimation techniques achieved prominent attention in the literature. First, various method of moments (MM) estimators have been suggested by various authors (Taylor, 1986; Melino and Turnbull, 1990; Andersen and Sørensen, 1996).

MM estimators avoid the problems associated with the linearisation of the model as well as the evaluation of the likelihood. They are not difficult to implement and to generalise but the efficiency of these estimators is known to be suboptimal to the likelihood-based method of inference. It is evident in the simulation experiments of Andersen (1994b) and Andersen and Sørensen (1996) that sample sizes of less than $T$=1,000 are insufficient for meaningful estimation by the Generalised MM (GMM). For the sample size of $T$=500, it has been found that the efficiency of MM and QML estimators are very similar (Ruiz, 1994, Andersen and Sørensen, 1996) but both are strictly outperformed by the fully efficient MCMC method (Jacquier, Polson and Rossi, 1994). The QML estimator usually dominates for values of $\phi$ close to unity because for $\phi \approx 1$ the autocorrelations decrease slowly and are captured less well by the moments used in the MM procedure. This simulation evidence on finite sample performance suggests that the full likelihood procedure will be superior to both estimation techniques.

Furthermore, the method of moments is frequently used in econometrics when some variables are restricted to be uncorrelated and no distributional assumptions about the disturbances are made. Neither of these motivations is valid since the SV model is fully specified. Moreover, the estimation methodology does not provide an estimate of the instantaneous volatility $\sigma_t^2$ throughout the sample, $t$=1, .., $T$ so that an additional form of estimation is required. For instance, Andersen (1994a) and Ghysels and Jasiak (1996) use various MM techniques to estimate the parameters and the Kalman filter to obtain the volatility estimates.

Second, Kim, Shephard and Chib (1996) suggest to approximate the distribution of $\varepsilon_t$ by a mixture of normals. Conditional on a particular mixture, the likelihood can be computed via the prediction error decomposition since the linear structure of the model

is essentially retained. An important drawback of this method is that no matter how many mixture components are used, the mixture of normals can never give a good approximation to the tail behaviour of the $\ln(\chi_t^2)$ distribution. Jacquier, Polson and Rossi (1994, p. 416) argue that the convergence of the algorithm is likely to be very sensitive to the number and weight of individual mixture components.

Third, Jacquier, Polson and Rossi (1994), JPR thereafter, have developed a Bayesian approach to the estimation of SV models using a Markov Chain Monte Carlo (MCMC) technique. The likelihood function can be expressed as a mixture of distributions for the observations conditional on the volatilities:

$$L(y|\psi) = \int p(y|\sigma)p(\sigma|\psi)p(\psi)d\sigma \qquad (2.5)$$

where $y$ and $\sigma$ are $(T \times 1)$ vectors of univariate observations, $y_t$, and volatilities, $\sigma_t$, respectively. Writing the expression in terms of $\sigma_t$ rather than $h_t$ avoids the problems associated with the linearising transformation. The first density, $p(y|\sigma)$ is determined by the choice of the distribution for $v_t$, e.g. Gaussian. Second, the marginal distribution of model parameters, $p(\psi)$ is given by a Bayesian prior, e.g. inverse Gamma. The MCMC algorithm is employed to simulate draws from the augmented density $p(\sigma, \psi|y)$ and obtain the marginal distribution $p(\sigma|\psi)$ by averaging the appropriate conditional distributions over simulated draws.

Following this approach, Shephard and Pitt (1997) have constructed a more efficient block MCMC algorithm for performing Bayesian inference on general non-linear and non-Gaussian state space models of which the SV model (2.1) is a special case. They conclude that the performance of the multi-block MCMC methods outperforms the single block approach of JPR in terms of computational efficiency.

JPR have performed extensive simulation experiments in which they have demonstrated that this method is superior to QML and MM estimation techniques across a wide range of parameter values. Particularly in the region of the parameter space where the variance of the volatility process is small, the MCMC technique was found to perform much better. However this is also the region of parameter space in which the MCMC estimates are highly sensitive to the prior used in initialising the Markov Chains (Breidt and Carriquiry, 1996). Furthermore, this technique has some undesirable features. The procedure is quite involved, requiring a large amount of computer intensive simulations. In addition, the method needs to be nontrivially modified for the extensions like the introduction of explanatory variables, alternative processes for the evolution of variance, or multivariate specifications (Jacquier, Polson and Rossi, 1995).

Fridman and Harris (1996) suggest that the non-Gaussianity of the measurement equation disturbances can be handled by means of a "brute force" numerical integration. In a Monte Carlo study - similar to the one presented here - the authors demonstrate how Kitagawa's (1987) extended Kalman filter can be applied in this context. By retaining the state space form, Fridman and Harris' (1996) estimation technique offers the same advantages as the one developed here, and indeed, may prove a strong competitor in applied research.

Some of the disadvantages of this method consist of computational inefficiencies (the extended Kalman filter is known to be rather slow) and the necessity to choose *a priory* a fixed grid, over which the volatility process will be integrated. This creates a trade-off between numerical accuracy on the one hand, and computational efficiency on the other. It is conceivable, that in some instances an optimal grid may not exist. For instance, when estimating the volatility process around the stock market Crash of

'87 the grid selection procedure proposed by Fridman and Harris (1996) will either lead to a very coarse grid over the entire volatility range, or place no probability weight on the high volatility state during the Crash.

Finally, Danielsson (1994a) proposed to estimate the SV model by the Monte Carlo likelihood (MCL) estimation method. At the core of the method is the search for progressively informative sampling densities so that very accurate approximations to the integral (2.5) can be obtained. Once a suitable importance sampling density is found (from which values of the latent volatility variable can be sampled), the latent volatility variable can be integrated out of the joint density of observations and volatilities. This is a powerful technique whose time requirement and precision is similar to MCMC. However, the method is difficult to generalise and remains computationally expensive largely due to the failure of the technique to exploit the linear structure resulting from the transformation (2.2) and the availability of standard simulation algorithms (de Jong and Shephard, 1995).

Durbin and Koopman (1997a) and Shephard and Pitt (1997) have designed importance sampling methods for general state space models. Specialising this approach to the estimation of SV models the next Section demonstrates that the resulting MCL estimator is a viable alternative to the MCMC technique. The finite sample performance of our estimator is as good as MCMC, the computational requirement is smaller by a factor of 10, and the method need not be modified when the basic model is extended in a number of interesting directions.

## 2.3. The Monte Carlo Likelihood (MCL) method

Taking logarithms of the squared residuals in (2.1) gives the linear state space (2.2) but invokes an additional difficulty: the disturbance term in the measurement equation

becomes non-Gaussian. Let $\theta = (\theta_1, \theta_2, .., \theta_T)'$ be a $(T \times 1)$ vector of (unobserved) signals[2] in the measurement equation of (2.3) with individual elements $\theta_t = Z_t \alpha_t$. Then the likelihood function of the general state space model (2.3) can be expressed as:

$$L(y|\psi) = \int p(y, \theta|\psi)d\theta = \int p(y|\theta, \psi)p(\theta|\psi)d\theta \qquad (2.6)$$

where $d\theta$ stands for $\prod_{t=1}^{T} d\theta_t$. The second equality illustrates the necessity to integrate the latent variable $\theta$ out of the joint density of $y$ and $\theta$. Monte Carlo integration is a method which attempts to perform this task. A naïve Monte Carlo estimator of the likelihood function can be obtained by drawing a number of independent random sequences $\theta^{(i)}$ $i=1,...,N$ (each of length $T$) from the unconditional distribution of $\theta$, $p(\theta|\psi)$. This estimator will be given by:

$$\hat{L}(y|\psi) = N^{-1}\sum_{i=1}^{N} p(y|\theta^{(i)}, \psi)$$

While such an estimator will be unbiased, and have a variance $O(N^{-1})$, allowing more accurate approximations with larger $N$, it is likely to be very poor. The reason for this, is the fact that in the $T$-dimensional space of $y$ there will be very few draws of $\theta$ which will be close enough to the particular sample path of $y$ so as to make a meaningful contribution.

The idea of importance sampling (Ripley, 1987) can be best appreciated by noting that the likelihood function (2.6) can also be expressed as:

$$L(y|\psi) = \frac{p(y|\theta, \psi)p(\theta|\psi)}{p(\theta|y, \psi)} \qquad (2.7)$$

---

[2] The signals. $\theta_t$ correspond to the (log)variances, $\ln\sigma_t^2$ in the previous notation.

Now, define a random variable, sometimes called the remainder function (Danielsson, 1994a):

$$X(\theta) = \frac{p(y|\theta,\psi)p(\theta|\psi)}{\tilde{p}(\theta|y,\psi)}$$ 
(2.8)

where the importance density $\tilde{p}(\theta|y,\psi)$ is an arbitrary conditional density of $\theta$ given $y$ and $\psi$. It is simple to verify that for *any* choice of $\tilde{p}(\theta|y,\psi)$, the expectation of the remainder function $X(\theta)$, taken with respect to the distribution of $\theta$ is precisely the conditional likelihood function needed for maximum likelihood estimation:

$$E(X(\theta)) = \int \frac{p(y|\theta,\psi)p(\theta|\psi)}{\tilde{p}(\theta|y,\psi)}\tilde{p}(\theta|y,\psi)d\theta = L(y|\psi)$$ 
(2.9)

Danielsson (1994a) suggested applying this general idea to the estimation of stochastic volatility models. His is an elaborate algorithm for determining the exact form of the importance sampling density $\tilde{p}(\theta|y,\psi)$, which is non-trivial. The central idea of Durbin and Koopman (1997a) is the observation that $X(\theta)$ can be decomposed into two parts: the Gaussian likelihood function and a residual term. If the disturbance term $\varepsilon_t$ were Gaussian, the likelihood function would be:

$$L_G(y|\psi) = \frac{p_G(y|\theta,\psi)p(\theta|\psi)}{p_G(\theta|y,\psi)}$$ 
(2.10)

where $p_G(x|z)$ denotes a Gaussian density function for random vector, $x$, conditional on $z$. This is the likelihood criterion function which is maximised in the QML procedure. Taking the importance density $\tilde{p}(\theta|y,\psi) = p_G(\theta|y,\psi)$ and combining (2.8) and (2.10) one obtains:

$$L(y|\psi) = L_G(y|\psi)E\left[\frac{p(y|\theta,\psi)}{p_G(y|\theta,\psi)}\right]$$ 
(2.11)

where the expectation is now taken with respect to the Gaussian density $p_G(\theta|y, \psi)$. The first component is obtained by the Kalman filter while the second is simulated. Thus, the latent variable can be integrated out of the joint density of observations and volatilities by standard filtering and smoothing algorithms. Importance sampling is only required in the second stage to correct for the non-normality induced by applying the linearising transformation. Defining

$$w^{(i)} = \frac{p(y|\theta^{(i)}, \psi)}{p_G(y|\theta^{(i)}, \psi)} \tag{2.12}$$

where $\theta^{(i)}$ is a particular draw from $p_G(\theta|y, \psi)$. The Monte Carlo estimator of the likelihood function is readily obtained as:

$$\hat{L}(\psi) = L_G(y|\psi)\overline{w} \quad , \qquad \overline{w} = N^{-1}\sum_{i=1}^{N} w^{(i)} \tag{2.13}$$

As before, this estimator is unbiased. Its variance is given by $Var\left[\hat{L}(\psi)\right] = N^{-1}L^2(\psi)\sigma_w^2$ where $\sigma_w^2$ can be consistently estimated by the sample variance of $w^{(i)}$. Thus, the Monte Carlo technique delivers the expectation of that element of the likelihood function which is not captured by the Gaussian term. Because of the one-to-one correspondence between the signals and the noise, the quantities $w^{(i)}$ in (2.12) can be re-expressed as:

$$w^{(i)} = \prod_{t=1}^{t} w_t^{(i)} , \qquad w_t^{(i)} = \frac{p_{\ln \chi_1^2}(\varepsilon_t^{(i)}|\psi)}{p_G(\varepsilon_t^{(i)}|\psi)} \tag{2.14}$$

which is a useful result. It implies that in univariate models, instead of drawing ($T\times m$) signals, $\theta^{(i)}$ from $p_G(\theta|y, \psi)$, we can compute (2.12) by drawing ($T\times 1$) disturbances, $\varepsilon^{(i)} = (\varepsilon_1, \varepsilon_2, .., \varepsilon_T)'$ from $p_G(\varepsilon|\psi)$. This operation is accomplished by means of the simulation smoother of de Jong and Shephard (1995), details of which can be found in Appendix 2.

From our experience, the quantities $w^{(i)}$ in (2.14) are very small numbers and in practice - for reasons of numerical stability - we work with $l_t^{(i)} = \ln w_t^{(i)}$. This has the added advantage of reducing the number of exponential function calculations, as may be seen from (2.16). Furthermore, is more convenient to work with the log-likelihood function, an unbiased estimate of which is given by:

$$\ln \hat{L}(\psi) = \ln L_G(Y|\psi) + \ln \overline{w} + \frac{s_w^2}{2N\overline{w}^2} \qquad (2.15)$$

where $\overline{w}$ and $s_w^2$ are calculated in the following manner:

1. Sample $\varepsilon^{(i)} = (\varepsilon_1, \varepsilon_2, .., \varepsilon_T)'$, $i=1,...N$ from $p_G(\varepsilon|\psi)$.

2. Compute $w^{(i)}$ in (2.14).

3. Calculate $\overline{w}$ and $s_w^2$ as the sample mean and variance of $w_i$.

Notice that quantity $w^{(i)}$ is a ratio of the true density of the disturbances - $\ln(\chi_I^2)$ - to the Gaussian sampling density. Its expectation gives that part of the likelihood surface which is not already captured by the Gaussian approximation. By Proposition 1, a closed form for $w_t^{(i)}$ is given by:

$$w_t^{(i)} = H_t^{1/2} \exp\left(0.5\left(\varepsilon_t^{(i)} - e^{\varepsilon_t^{(i)}} + \left(\varepsilon_t^{(i)}\right)^2 H_t^{-1}\right)\right) \qquad (2.16)$$

where $H_t$ is the variance of the measurement equation noise in the general state space form (2.3).

Durbin and Koopman (1997a) consider a number of devices which improve the accuracy of the MCL function (2.15) which are now discussed.

*(i) Antithetic variables.*

Antithetic variables is a standard technique in importance sampling aimed in reducing the simulation variance. In the present context antithetic samples are constructed as:

$$\tilde{\varepsilon}_t^{(i)} = 2\hat{\varepsilon}_t - \varepsilon_t^{(i)} \quad , \qquad \hat{\varepsilon}_t = E(\varepsilon_t | Y) \tag{2.17}$$

where the $\hat{\varepsilon}_t$'s are obtained from a disturbance smoother (Appendix 2). Equation

(2.17) creates equiprobable draws - since $\tilde{\varepsilon}_t^{(i)} - \hat{\varepsilon}_t = -\left(\varepsilon_t^{(i)} - \hat{\varepsilon}_t\right)$ - which is useful if control

variables are used (as discussed subsequently). When antithetic variables are employed

the sample mean, $\overline{w}$ and sample variance, $s^2_w$ in (2.15) are effectively taken over $2N$

draws.

*(ii) Equalising density slopes.*

So far the importance sampling density $p_G(\varepsilon | \psi)$ has been taken to be the Gaussian

density with constant variance, i.e. we assumed that $\varepsilon_t \sim N(0, H_t)$, $H_t = \pi^2/2$ for $t=1$,

.., $T$. However, the importance sampling density $\tilde{p}(\theta | y, \psi)$ can be chosen arbitrarily in

any way that improves the accuracy of the simulation. In particular, one can use

$\varepsilon_t \sim N(0, \tilde{H}_t)$ where the scalar variances $\tilde{H}_t$'s are chosen so as to make the

differences between the logdensities $\ln p_{\ln \chi_1^2}(\varepsilon | \psi)$ and $\ln p_G(\varepsilon | \psi)$ as constant as possible

in the neighbourhood of $\hat{\varepsilon}_t = E\left(\varepsilon_t | Y\right)$. Intuitively, large negative values of $\hat{\varepsilon}_t$ would

require high values of $\tilde{H}_t$ in order for the slopes of the densities in **Figure 2.1** to be

roughly equal.

The choice of $\tilde{H}_t$'s is determined by equalising the derivatives of the logdensities at $\hat{\varepsilon}$

leading to a set of vector equations:

$$\frac{1}{2}(1 - e^\varepsilon)\bigg|_{\varepsilon = \hat{\varepsilon}} = -\frac{\varepsilon}{\tilde{H}}\bigg|_{\varepsilon = \hat{\varepsilon}} \quad \Leftrightarrow \quad \tilde{H}_t = \frac{2\hat{\varepsilon}_t}{e^{\hat{\varepsilon}_t} - 1} \qquad t = 1, \dots, T \tag{2.18}$$

It is simple to verify that for any choice of $\hat{\varepsilon}_t$ the nominator and denominator of (2.18)

have the same signs, thus ensuring that $\tilde{H}_t$ is positive.

The vector equations (2.18) can be solved for $\tilde{H}_t$ by iteration, starting at $\tilde{H}_t^{(0)} = \pi^2/2 \; \forall$

$t$. Given a parameter vector $\psi$, we iterate $K$ times between computing $\hat{\varepsilon}_t$ and $\tilde{H}_t^{(k)}$,

$k=1, .., K$, effectively running the Kalman filter and the disturbance smoother.

Choosing the metric $d(k) = T^{-1} \sum_{t=1}^{T} \left| \tilde{H}_t^{(k)} - \tilde{H}_t^{(k-1)} \right|$ to describe successive changes in

the variance vector, $\tilde{H}$, we find that after about six to eight iterations the elements of

the variance vector cease to fluctuate, i.e. $\tilde{H}_t^{(k)} \rightarrow \overline{H}_t$, results of which are reported in

**Table 2.1**, at the end of this Chapter. **Table 2.1**. shows rapid convergence across a

range of parameter values for the simulated SV model (2.1).

The individual elements of the variance vector $\overline{H}$ are now different across $t=1, .., T$.
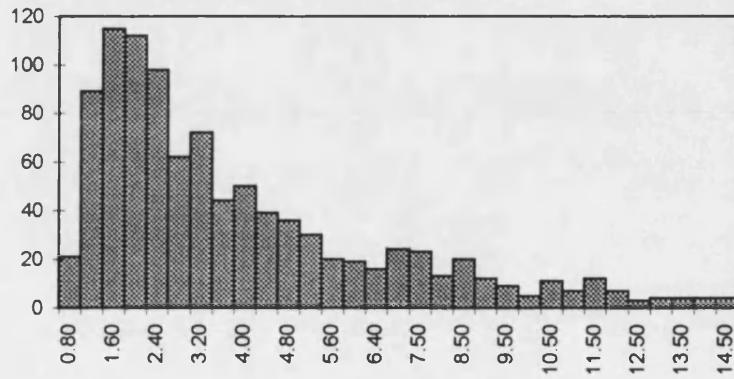


**Figure 2.2:** Effect of equalising density slopes

**Figure 2.2** presents the histogram of $\overline{H}_t$ for one realisation of the basic SV model

with $T=1,000$ observations. It is the mirror image of the density of $\varepsilon_t$ (**Figure 2.1**) and

reconfirms the intuition behind the method: Large negative, but infrequent values of $\varepsilon_t$

require high values of variance parameter $\overline{H}_t$ in order to compensate for the difference

in density slopes in this region. The converse is represented by large probability mass

of $\overline{H}_t$ lower than $\pi^2/2$.

Thus the likelihood construction algorithm - described below equation (2.15) - is modified to include $K$ iterations between the Kalman filter, the disturbance smoother, and the transformation (2.18) before $N$ samples are drawn in Step 1.

*(iii) Control variables.*

Control variables is another variance reduction technique often used in the context of simulation which is based on factoring out the mean of the random variable. In our experience, the contribution of this technique above (*i*) and (*ii*) is not very large but nevertheless deserves some attention. Expanding $l(\varepsilon^{(i)})=\ln w^{(i)}$ as a Taylor series about $\hat{\varepsilon}$ one obtains:

$$l\left(\varepsilon^{(i)}\right) \approx l(\hat{\varepsilon}) + c\left(\varepsilon^{(i)}\right), \quad c\left(\varepsilon^{(i)}\right) \equiv \sum_{r=1}^{q}\sum_{t=1}^{T}\frac{\partial^r l(\varepsilon_t^{(i)})}{\partial \varepsilon^r}\frac{\left(\varepsilon_t^{(i)} - \hat{\varepsilon}_t\right)^r}{r!} \quad (2.19)$$

where the series have been truncated at $q$. Conveniently, the odd $r$ order terms cancel out due to the way in which antithetic samples we constructed in (2.17). This allows us to expand the Taylor series up to the fifth order term ($q=5$) by considering merely the second and fourth derivatives of $l(\varepsilon_t^{(i)})$:

$$
\begin{aligned}
d_{2,t} &= -\frac{1}{2}e^{\varepsilon_t} + \tilde{H}_t^{-1}\\
d_{4,t} &= -\frac{1}{2}e^{\varepsilon_t}
\end{aligned}
\quad (2.20)
$$

Both, $d_{2,t}$ and $d_{4,t}$ are scalar. The contribution of the Taylor expansion terms, $c(\varepsilon^{(i)})$ to $\bar{l} = \ln \bar{w}$ is $\bar{c} = N^{-1}\sum_{i=1}^{N}c\left(\varepsilon^{(i)}\right)$ whose expectation is given by:

$$c = E\left[c(\varepsilon^{(i)})\right] = \sum_{r=1}^{2}\sum_{t=1}^{T}\frac{d_{2r,t}}{(2r)!}E\left(\varepsilon_t^{(i)} - \hat{\varepsilon}_t\right)^{2r} = \sum_{r=1}^{2}\sum_{t=1}^{T}\frac{d_{2r,t}}{2^r r!}C_t^{2r} \quad (2.21)$$

where the quantities $C_t = E(\varepsilon_t - \hat{\varepsilon}_t)^2$ are obtained from the disturbance smoother. Thus, the expectation of $\bar{w}$ in (2.15) may be approximated by $\exp(l(\hat{\varepsilon}) + c)$.

Numerical maximisation of the MCL likelihood (2.15) - details of which can be found in Appendix 3 - gives the estimates of the hyperparameters $\psi$. The choice of the number of draws, $N$ governs the accuracy of the approximation to the likelihood function: as $N$ increases, the approximation becomes more accurate. In the event, the discussion below indicates that $N=5$ is sufficient for most practical purposes. Observe, that once the SV model is formulated in the state space form (2.3) the simulation algorithm and the optimisation procedure are invariant to many extensions of the basic SV model.

## 2.4. Finite sample performance

To assess the performance of the new method we conducted simulation experiments following the design of JPR, thus facilitating direct comparison with the MCMC method. The range of parameter values $\psi=(\bar{\sigma},\phi,\sigma_\eta)$ is selected in the following manner.[3] First, the values of the autoregressive parameter $\phi$ are set to 0.90, 0.95, and 0.98. This choice is motivated by empirical studies which reported the values of the autoregressive coefficient close to unity, ranging between 0.9 and 0.995. Second, for each value of $\phi$, the values of $\sigma_\eta$ are selected so that the coefficient of variation:

$$CV = \frac{\text{var}(h)}{E[h]^2} = \exp\left(\frac{\sigma_\eta^2}{1-\phi^2}\right) - 1 \qquad (2.22)$$

takes the values 10, 1, and 0.1. High values of the ratio of volatility variance to its squared mean indicate pronounced relative strength of the stochastic volatility process while low values of $CV$ signify that the model is close to the one of constant volatility.

In fact, if preliminary exploratory analysis of the data from a model with low $CV$ was based only on the autocorrelation structure of $r_t^2$ or $\ln r_t^2$ the practitioner without a strong prior belief that the SV model is the correct specification will be unable to distinguish between the SV and a homoscedastic model. Nevertheless, the parameter triplets $\psi_7 - \psi_9$ are included for the sake of completeness. The focus of interest is thus centred around parameter triplets $\psi_4 - \psi_6$ which correspond to the coefficient of variation close to unity. Most of the empirical studies surveyed by JPR report parameter estimates in this range.

Finally, the values of long run volatility level, $\bar{\sigma}$ are chosen such that the expected variance:

$$E[h] = \bar{\sigma}^2 \exp\left(\frac{\sigma_\eta^2}{2(1-\phi^2)}\right)$$

(2.23)

is set to 0.0009. If the simulated data are regarded as daily returns, this corresponds to approximately 22% annualised volatility if the data are thought as being sampled at weekly frequency. Note, that JPR chose the parameterisation (1.13) rather than (1.13'). This gives the following range of parameter triplets:

| | $CV=10$ | | | $CV=1$ | | | $CV=0.1$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\psi_4$ | $\psi_5$ | $\psi_6$ | $\psi_7$ | $\psi_8$ | $\psi_9$ |
| $\phi$ | 0.9 | 0.95 | 0.98 | 0.9 | 0.95 | 0.98 | 0.9 | 0.95 | 0.98 |
| $\sigma_\eta$ | 0.675 | 0.484 | 0.308 | 0.363 | 0.260 | 0.166 | 0.135 | 0.096 | 0.061 |
| $\alpha$ | -0.821 | -0.411 | -0.164 | -0.736 | -0.368 | -0.147 | -0.706 | -0.353 | -0.141 |

For each of the nine triplets, $\psi_i$, we generate samples of length $T=500$,[4] estimate the model by various techniques and compute means and root mean squared errors of the

---

[3] The correlation coefficient, $\rho$ is set to zero.
[4] In fact, samples of length $T=600$ are simulated and the first 100 observations deleted. This is a well known technique aiming at reducing the dependence on initial conditions.

parameter estimates over $K=500$ simulated realisations of the process. In all calculations the number of draws, $N$ used to take the expectation in (2.15) is set to 5.

Results from the sampling experiments are presented in **Table 2.2** which is divided into three panels in accordance with the coefficient of variation $CV$. Within each panel the true parameter values are displayed first. The JPR simulation results for the Bayes (MCMC) estimator, their Table 7, are reproduced in the first row, the Quasi-Maximum Likelihood estimator (QML) in the second, followed by the MCL estimator given in equation (2.15). The sampling standard deviations of the parameter estimates are in parenthesis below. The starting parameter values for both, the QML and MCL optimisation routines are obtained from a two dimensional grid search procedure which searches for an optimum across the surface of the Gaussian Likelihood function (2.10).

The figures presented in **Table 2.2** allow several conclusions to be drawn. First, the experiment demonstrates that the Monte Carlo Likelihood (MCL) estimator (2.15) is as efficient as JPR's MCMC estimator across all parameter values. In most cases the standard errors of the MCL estimator, documented in the final row of each panel, are at least as small as those reported for the MCMC estimator. In addition the bias exhibits a similar behaviour; the average parameter estimates obtained by MCL are as close to the true values as the MCMC estimates and significantly closer than the QML estimates.

The performance of the MCL estimator is further illustrated by examining the sampling distributions of the parameter estimates. **Figures 2.3** and **2.4** present the smoothed densities of the estimates of $\phi$ and $\sigma_\eta$ for two triplets $\psi_4$ and $\psi_5$, see Appendix 1 for details of nonparametric kernel density estimation. The MCL estimator is shown to exhibit a much tighter sampling distribution then the QML estimator, a property already indicated by smaller standard errors of the estimates reported in **Table 2.2**.

51

$\phi$=0.9 $\qquad\qquad\qquad\qquad\qquad$ $\sigma_\eta$=0.363
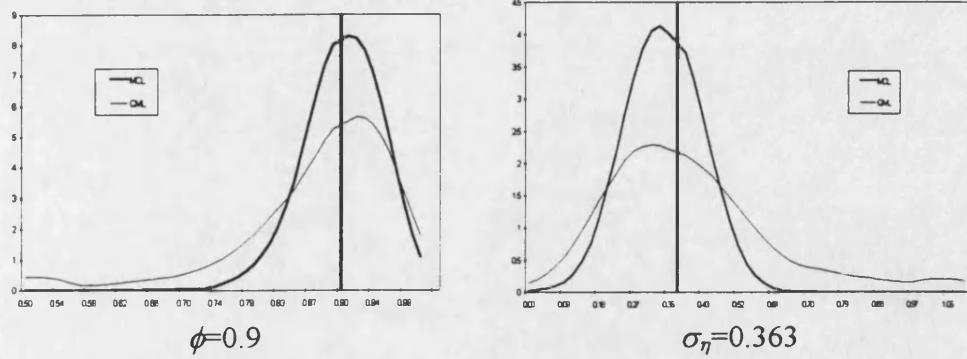
**Figure 2.3:** Sampling distributions of the MCL and the QML estimators; $\psi_4$



$\phi$=0.95 $\qquad\qquad\qquad\qquad\qquad$ $\sigma_\eta$=0.260
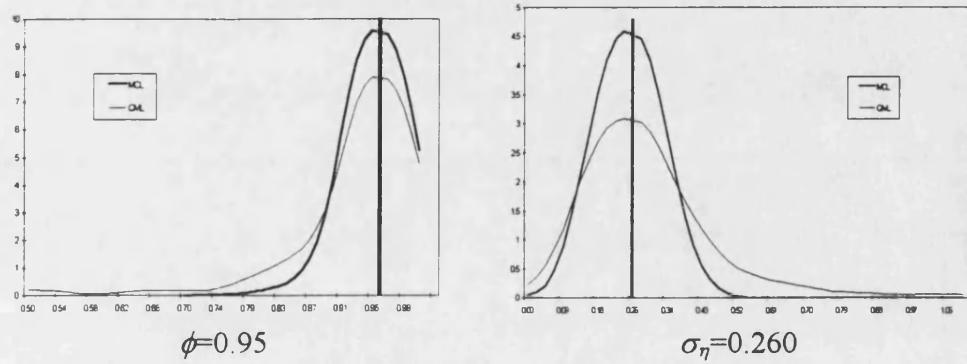
**Figure 2.4:** Sampling distributions of the MCL and the QML estimators; $\psi_5$

Second, confirming the JPR's results, the QML estimator (2.10) is found to be inefficient. Across the entire parameter space the standard errors of the QML estimator are at least twice the size of the fully efficient MCL estimator while the bias is non-negligible. The efficiency of the QML estimator increases as the strength of the SV process becomes more pronounced, i.e. for parameter triplets corresponding to the values of $CV$ equal to ten. For instance, for the case of $CV$=10 the sample standard error on $\phi$=0.95 is 0.046 while in the case of $CV$=1 the standard error on $\phi$=0.95 is twice as large, 0.101.

However, although we find QML to be inefficient, its performance is nowhere as near as bad as reported by JPR. Same conclusion was reached by Breidt and Carriquiry (1996) who also re-examined the finite sample performance of the QML estimator. Since **Figures 2.3** and **2.4** were constructed so as to correspond to JPR's Figures 4 and 5 respectively, direct comparison reveals dramatic differences in the performance

52

of the same estimation technique. This raises the question of possible inefficiencies in JPR's QML estimation method such as poor starting values, different convergence criteria, or inefficient implementation of the algorithm.

Thirdly, the performance of *all three estimators* deteriorates as $CV$ decreases. The standard errors on $\phi$ increase twofold by going from $CV=10$ to $CV=1$ and eight times by going from $CV=1$ to $CV=0.1$. Comparison of the MCL and the MCMC estimators in this region reveals that the MCL estimator exhibits slightly larger standard errors but a considerably smaller bias.

However, as the number of draws, $N$, increases the expectation of the MCL likelihood function in (2.15) can be taken more precisely, thus leading to increased performance. In principle, the approximation can be made arbitrarily close by choosing a large $N$, but the computational burden will render this strategy impractical. In our experience, a very small number of draws is sufficient to produce results comparable with the MCMC estimator. This is illustrated in **Table 2.3**.

The interest is primarily focused upon two factors: the number of draws required, and the benefit of using the device of equalising density slopes as in (2.18). First, observe that already a small number of draws, $N=10$, and without equalising the density slopes, the MCL technique produces results comparable to the fully efficient MCMC estimator. Both in terms of bias and precision the MCL ($N=10$) and MCL ($N=20$) estimators match the performance of MCMC. Second, the precision of the MCL estimator can be significantly improved by employing the density slope equalisation technique, as can be seen by examining the standard errors of the MCL*($N=5$). For instance, employing this device, only 5 draws are sufficient to reduce the standard error on each parameter by a factor of two, as can be seen from the line labelled

MCL*($N$=5) in **Table 2.3**. Increasing $N$ more dramatically reduces the standard errors even further.

The final column of **Table 2.3** reports average time that was required to obtain parameter estimates, given the initial guess: for each realisation of the process the initial guess was obtained from a coarse grid of the quasi-likelihood surface. That same parameter vector was then used by each estimation technique as the starting value. Since time required for the convergence of the algorithm is machine dependent we chose to report the figures in terms of multiples of the QML time (All calculations were performed on a 586, 90mhz, 8 RAM, PC. On average the QML estimation took 0.572 seconds, thus requiring 0.572*27.67=15.8 seconds for the MCL*($N$=5) estimator). The experiment suggests, that full efficiency can be achieved by the MCL with only $N$=5 draws if the device of equalising density slopes is employed. This is our preferred estimator whose performance across the entire parameter space was already reported in **Table 2.2**.

Finally, we investigated the computational requirement for the MCL estimator across different sample sizes. Due to timing considerations we chose to measure the average time for one function evaluation the results of which are presented below:

| $T$ | $QML$ | $MCL$ ($N$=10) | $MCL$ ($N$=20) | $MCL$*($N$=5) | $MCL$*($N$=10) |
|---|---|---|---|---|---|
| 500 | 0.020 | 14.55 | 26.32 | 25.66 | 31.56 |
| 1,000 | 0.037 | 15.62 | 28.33 | 27.56 | 33.90 |
| 2,000 | 0.072 | 16.29 | 29.49 | 28.17 | 34.58 |
| 6,000 | 0.220 | 15.59 | 28.22 | 27.53 | 33.84 |
| 10,000 | 0.359 | 16.00 | 28.96 | 27.94 | 34.39 |

The column labelled QML reports average time for one function evaluation in seconds, which is, of course, slowly increasing in $T$. The following columns give the average time for the different MCL estimators expressed as a multiple of the QML time within

54

each row. While the relative function evaluation time for the MCL\*($N$=5) is less than twice as large (25.66 vs. 14.55) as for MCL($N$=10) estimator, the average time until convergence is almost identical (27.66 vs. 27.66). This indicates that the likelihood surface is better specified and fewer iterations are required by the MCL\*($N$=5) estimation. Thus, despite larger function evaluation times, parameter estimates can be obtained within same time periods for either method.

Finally, we would like to compare the computational requirement for the MCL estimator with that of MCMC. Unfortunately, JPR do not report their average estimation times, not even as multiples of QML, and one has to make an indirect comparison referring to the times reported by Danielsson (1994b).

We re-estimate the basic SV model on the same data set used by JPR and Danielsson (1994b). The data set consists of daily observations on the S&P500 stock index level in the period 2/1/80-30/12/87. The return series are prefiltered to remove the calendar effects as documented in Gallant, Rossi and Tauchen (1992). The sample length is $T$=2,023 observations.

The choice of this particular data set is convenient for two reasons. First, the parameters of the process have already been estimated by MCMC and Danielsson's MCL which gives a reference point. More importantly, this is the only instance in the literature where the times for the MCMC estimation are reported thus allowing us to calibrate the relative time requirement of our method. The results of the estimation were:

| | $\alpha$ | $\phi$ | $\sigma_\eta$ | *time* |
|---|---|---|---|---|
| *MCL* | -0.00 | 0.96 | 0.16 | 1:21 |
| *MCMC* | -0.00 | 0.97 | 0.15 | 7:15 |
| *Danielsson's MCL* | -0.00 | 0.97 | 0.15 | 10:45 |

As in the simulation experiment, the device of equalising density slopes and $N=5$ draws were used. The parameter estimates are almost identical and the time requirement for the new MCL is about five times smaller than either of the of the methods.[5]

These results are very encouraging. They demonstrate that the MCL estimator proposed here exhibits very satisfactory small sample performance which is directly comparable to the fully efficient Bayesian MCMC method. The evidence also suggests that these results can be achieved by using a very small number of draws.

## 2.5. Further issues

Having shown that the MCL estimator exhibits satisfactory finite sample performance we would now like to turn to the practical issues in SV model estimation and indicate some of the interesting extensions of the basic SV model (2.1).

### 2.5.1. Inlier problem

Since our method, as much as QML, relies on the use of the linear state space, taking the logarithms of squared mean adjusted returns becomes a problem when zero, or small values are encountered. In particular, if the drift in of the asset can be assumed to be zero and the series to be investigated is high frequency data, or prices are recorded discretely then it is possible that some returns will be zero. In many practical applications, however, equality of prices at two successive observations in time, leading to zero returns, arise due to data irregularities. For instance, properly accounting for holidays eliminates many "zero" returns in any daily exchange rate series. Deleting such observations from the sample eliminates the inlier problem.

---

[5] Direct comparison of computational requirement is obscured by the difference in computer platforms (SPARC superstations vs. PC), resources (size of RAM, clock speed), as well as differences in

Alternatively, the updating equations of the Kalman Filter can be modified so as to handle missing values (Harvey, 1989).

If the inlier cannot be assumed to be an irregular observation there are three alternatives of dealing with the problem. First, the sample mean of the series may be subtracted from the observations. While the method may be feasible numerically (the resulting series are devoid of entries identically equal to zero) it does not solve the problem conceptually. Second, Breidt and Carriquiry (1996) suggest a transformation of the data which amounts to a truncation of large negative values of $\ln(r_t^2)$ thus shifting some probability mass towards the centre of the distribution:

$$\ln^*\!\left(r_t^2\right) = \ln\!\left(r_t^2 + \lambda s_r^2\right) - \frac{\lambda s_r^2}{r_t^2 + \lambda s_r^2} \quad , \qquad s_r^2 = \frac{1}{T}\sum\left(r_t - \bar{r}\right)^2$$

where $\lambda$ is some subjectively chosen constant, e.g. 0.02. The authors demonstrate that this transformation improves the performance of the QML estimator and mitigates the inlier problem. On the other hand, this transformation alters the entire density of the data, an unsatisfactory solution. Finally, one may cut off the inliers by setting the observation at some value $\kappa$:

$$\ln^*\!\left(r_t^2\right) = \ln\!\left\{r_t^2\, I_{\{r_t \ge \kappa\}}\right\} + \ln\kappa^2\, I_{\{r_t < \kappa\}} \tag{2.24}$$

where $I\{\cdot\}$ is the indicator function, and $\kappa$ is a small number. Invariably, the choice of $\kappa$ is subjective but it is demonstrated below that (2.24) leads to reasonably good MCL estimates for very small $\kappa$.

To assess the performance of the MCL and QML methods across various values of $\kappa$ we designed the following Monte Carlo experiment. For the parameter triplet $\psi_5$ we generated the basic SV model (2.1) as before, except that the $\varepsilon_t$'s have now a 10%

programming languages (C++ vs. GAUSS).

chance of taking the value zero and 90% chance of being drawn from $N(0,1)$. It is rarely the case in practical applications that 10% of the sample are identically equal to zero but the experiment has been designed to illustrate the behaviour of the estimators in extreme situations. The generated series was then transformed according to (2.24) with $\ln(\kappa_i^2)$ taking the values of -20, -30, -100, and -200. The results of the simulations are presented in **Table 2.4** and compared to those of the previous Section. It will be recalled, that in that experiment samples were generated in accordance with (2.1) while none of the draws $\varepsilon_t \sim N(0,1)$ was ever identically equal to zero[6]. The first row of each panel presents true parameter values, the second row reproduces the relevant results from **Table 2.2** and the following rows give simulation results when about 10% of the observations are set to $\kappa_i$.

It is apparent that the performance of QML leaves much to be desired. The bias and the standard errors are very sensitive to the choice of $\kappa$. As $\kappa$ is decreases the performance deteriorates rapidly, leading to enormous biases in all three parameters. However, the decline in precision is not homogenous across the three model parameters. Interestingly, for tiny $\kappa$ (e.g. $\kappa_4 = 3.72*10^{-44}$) the bias in the estimate of the autoregressive parameter $\phi$ disappears, while the biases in the estimates of $\sigma$ and $\sigma_\eta$ remain very large.

The results of the MCL estimator are considerably better. The bias and the standard errors on all three model parameters decrease with the cut off value $\kappa$. Comparison with the estimation results for the full sample (**Table 2.4**, Panel B, row two), reveals that less precision can be achieved when 10% of observations are zero, the standard errors in this case are about twice as large. This is not surprising, and stems from the

---

[6] Which is a property of the random number generator.

fact that the likelihood function is 'flatter' in cases when many zero observations are present.

## 2.5.2. Heavy tails

The unconditional density of many financial series exhibits larger kurtosis than can be captured by simply incorporating conditional heteroscedasticity into a Gaussian process. As has been pointed out before, the basic SV model can be generalised so as to allow the mean equation innovations $v_t$ to be *Student-t* distributed as in (1.2″). The density of the transformed disturbances, $\varepsilon_t = \ln v_t^2$ is given by:

**Proposition 2.2**: Let the scalar variable $x$ have a *Student-t* distribution with $v$ degrees of freedom. Then the density of $z = \ln x^2$ is:

$$p_{\ln t_v^2}(z) = C_v\left(1 + \frac{e^z}{v}\right)^{-\frac{v+1}{2}} e^{\frac{z}{2}} \qquad C_v = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)} \qquad (2.25)$$

**Proof**: The density of a variable which follows a *Student-t* distribution with $v$ degrees of freedom is given by:

$$p_{t_v}(x) = C_v\left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}} \qquad C_v = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}$$

Making first the change of variable $y = x^2$ one obtains:

$$p_{t_v^2}(y) = p_{t_v}\left(\sqrt{y}\right)\left|\frac{dx}{dy}\right| = C_v\left(1 + \frac{y}{v}\right)^{-\frac{v+1}{2}} 0.5 y^{-\frac{1}{2}} 2$$

Now, let $z = \ln y$ and the proposition follows. □

The limit of (2.25), as $v \to \infty$, is of course, the $\ln(\chi_t^2)$ density (2.4) which can be verified by taking logarithms of (2.25), and expanding $\ln(1+x)$ as a Taylor series.

The computation of the MCL likelihood (2.15) involves the quantities $w_t^{(i)}$ in (2.14) which are now constructed as:

$$w_t^{(i)} = \left(2\pi H_t\right)^{1/2} \exp\left(0.5\left(\varepsilon_t^{(i)} + \left(\varepsilon_t^{(i)}\right)^2 H_t^{-1}\right)\right) C_v \left(1 + v^{-1} e^{\varepsilon_t^{(i)}}\right)^{-\frac{v+1}{2}} \qquad (2.16')$$

Furthermore, the first, second and fourth derivatives of (2.25), required for equalising density slopes in (2.18) and Taylor series expansions in (2.19) are:

$$d_1(z) = \frac{1}{2} + \frac{k_v}{v}\left[\frac{e^z}{v + e^z}\right]$$

$$d_2(z) = k_v \frac{e^z}{\left[v + e^z\right]^2}$$

$$d_4(z) = k_v \frac{e^{2z}\left[v^2 - 5v + e^z(v+1)\right]}{\left[v + e^z\right]^4}, \qquad k_v = -\frac{v(v+1)}{2}$$

So that the updating equations for $\tilde{H}_t$ in (2.18) become:

$$\tilde{H}_t = \frac{2\hat{\varepsilon}_t}{e^{\hat{\varepsilon}_t}\left[\dfrac{v+1}{v + e^{\hat{\varepsilon}_t}}\right] - 1} \qquad t = 1,\ldots,T \qquad (2.18')$$

Again, the Gaussian equations (2.18) are obtained in the limit, as $v \to \infty$ in equation (2.18'). Moreover, (2.18') automatically ensures the non-negativity of $\tilde{H}_t$, which can be verified by observing that the signs of the nominator and denominator are the same for any value of $v$ and $\hat{\varepsilon}_t$.

The number of degrees of freedom, $v$ enters the parameter vector $\psi$, over which the likelihood function is maximised. This is different from the Harvey, Ruiz and

Shephard's (1994) QML set-up where the variance of the measurement equation, $H$ in (2.3) is treated as a parameter.

To illustrate the method, we proceed to fitting the SV-$t$ model to the S&P500 return series, which were described earlier in Section 2.4, results of which are presented in **Table 2.5**. For ease of reference the results of the basic SV are reproduced in the upper panel. The estimated number of degrees of freedom is 7.634, well in the range of empirical estimates reported by Bollerslev (1987) using the GARCH-$t$ model: 6.211-13.889. The likelihood ratio test statistic takes the value 26.6 which is significant at the 1% level when compared to the relevant critical value of the $\chi_1^2$ distribution. Similarly, the standard error on $v$ indicates the significance of this parameter. The introduction of the *Student-t* distributed mean equation disturbances reduces the value of the implied coefficient of variation, $CV$ from 0.389 to 0.255: intuitively, lower variance of the latent process is sufficient to account for the variability in the series.
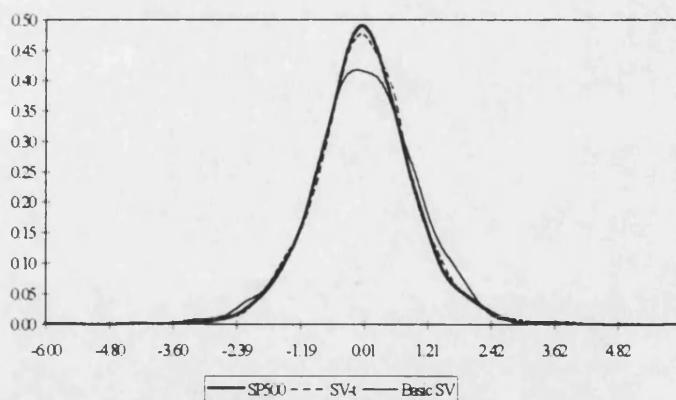


**Figure 2.5:** S&P500: unconditional density and the density of the SV-$t$ model

Finally, **Figure 2.5** demonstrates that the unconditional density of the S&P500 returns is closely approximated by the unconditional density from the SV-$t$ model.[7] By contrast, the unconditional density of the basic SV model (with normal $v_t$) does not capture as well the unconditional distribution of asset's returns. Thus the MCL estimator can be easily adjusted so as to incorporate heavy tailed distributions.

### 2.5.3. SV in the mean

SV in the mean models can be estimated by MCL by rewriting the mean equation (1.2′) as $r_t = \sigma e^{\frac{1}{2}h_t}(\delta + v_t)$ and applying the logarithmic transformation to obtain the residuals $\tilde{\varepsilon}_t = \ln(\delta + v_t)^2$. The new measurement equation disturbances will now have a non-central $\ln(\chi_1^2)$ distribution with the non-centrality parameter $\lambda = \delta^2$. The analytic expression for this density is given in:

**Proposition 2.3**: The non-central $\ln(\chi_1^2)$ distribution with the non-centrality parameter $\lambda$ is:

$$p_{\ln \chi_1^2}(z) = \exp\left[\frac{z - e^z}{2}\right]\sum_{j=0}^{\infty} c_j(\lambda)e^{jz} \qquad (2.26)$$

where

$$c_j(\lambda) = \frac{e^{-\lambda/2}\left(\frac{\lambda}{2}\right)^j}{j!\,\Gamma\left(\frac{1}{2}+j\right)2^{j+\frac{1}{2}}}$$

**Proof**: Starting with the non-central $\chi_1^2$ density (Johnson and Kotz, 1970, p. 130) the change of variables $z = \ln y$ gives the result.□

---

[7] Parameter estimates of **Table 2.5** were used to draw two samples of the SV process the density of which is presented in the figure. The $t$-distributed random numbers were constructed in accordance with the Bailey (1994) algorithm. I am grateful to Jon Danielsson for pointing out this procedure.

The density in (2.26) can be seen as a mixture of central $\ln(\chi_\nu^2)$ densities with varying degrees of freedom. Clearly, when SV in the mean effects are not present ($\lambda=0$), the density (2.26) reduces to (2.4) because $c_{j=0}(0)=(2\pi)^{-1/2}$ and $c_{j\geq1}(0)=0$. Despite the infinite sum in (2.26) it is possible to calculate the derivatives of $\ln p_{\ln z_i^2}(z)$, the first of which - required for the equalising density slopes procedure - is:

$$d_1(z) = \frac{1}{2}\left[1-e^z\right] + g(z,\lambda), \qquad g(z,\lambda) = \sum_{j=0}^{\infty} c_j(\lambda) j e^{jz} \left\{\sum_{j=0}^{\infty} c_j(\lambda) e^{jz}\right\}^{-1}$$

so that equation (2.18) becomes:

$$\tilde{H}_t = \frac{2\hat{\varepsilon}_t}{e^{\hat{\varepsilon}_t} - 1 - 2g(\hat{\varepsilon}_t,\lambda)} \qquad t = 1,\ldots,T \qquad (2.18'')$$

The function $g(\hat{\varepsilon}_t,\lambda)$ is non-negative and increasing in $\hat{\varepsilon}_t$ implying that in practice the modulus of (2.18'') needs to be taken to ensure the non-negativity of $\tilde{H}_t$.

On the other hand it is possible to approximate (2.26) with a multiple of a central $\ln(\chi_\nu^2)$ density by matching the first two moments:

$$\mu = \ln 2 + \overline{\psi}\left(\tfrac{1}{2}\right), \qquad \overline{\psi}\left(\tfrac{1}{2}\right) = \sum_{j=0}^{\infty} c_j(\lambda)\psi\left(\tfrac{1}{2}+j\right)$$

$$\sigma^2 = \sum_{j=0}^{\infty} c_j(\lambda)\left[\psi'\left(\tfrac{1}{2}+j\right) + \left\{\psi\left(\tfrac{1}{2}+j\right) - \overline{\psi}\left(\tfrac{1}{2}\right)\right\}^2\right]$$

where $\psi(a)$ and $\psi'(a)$ are the diagamma and trigamma functions respectively (Benett, 1955). However, it is well known (Johnston and Kotz, 1970, Chapter 28) that the approximation is rather poor for low values of $\nu$ in case of the non-central $\chi_\nu^2$ density which is likely to be also true in the case of $\ln(\chi_t^2)$.

The calculation of the likelihood function (2.15) requires the quantities $w_t^{(i)}$ of (2.16) which are computed as:

$$w_t^{(i)} = \left(2\pi H_t\right)^{1/2} \exp\left(0.5\left(\varepsilon_t^{(i)} - e^{\varepsilon_t^{(i)}} + \left(\varepsilon_t^{(i)}\right)^2 H_t^{-1}\right)\right)\sum_{j=0}^{K} c_j(\lambda)e^{j\varepsilon_t^{(i)}} \qquad (2.16'')$$

where as before, $\varepsilon_t^{(i)}$ are univariate draws from the simulation smoother, $H_t$ are scalar mean equation variances in (2.3), and the weight function $c_j(\lambda)$ is defined in (2.26). The truncation of the infinite sum is made at some finite value $K$. A guide to the choice of $K$ is the surface of the function $c_j(\lambda)$ $\lambda \geq 0$, $j=1,\ldots, K$, which is presented in **Figure 2.6**.
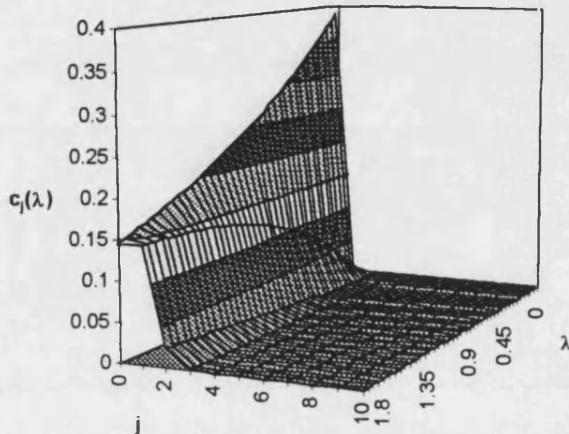


**Figure 2.6:** Weight function $c_j(\lambda)$ in the non-central $\ln(\chi_1^2)$ distribution.

**Figure 2.6** indicates that $K=4$ is a reasonable cut-off value for the range of $\lambda$ likely to arise in applied empirical research. For instance, using the ARCH(p) model, Engle, Lilien and Robbins (1987), obtain estimates of the risk premium, $\delta$, in the range 0.505-0.8, thus giving values of $\lambda$ (0.255-0.64) well in the range of the horizontal axis in **Figure 2.6**.

### 2.5.4. Explanatory variables

As has been mentioned earlier, the basic SV model can be extended to include a set of regressors, $Z_t$ in the variance equation. Chapters 4 and 5 are entirely devoted to the empirical validity of this model. It should be mentioned here, however, that since the explanatory variables enter the state vector, as can be seen in Appendix 2, the MCL

estimation procedure need not be adjusted in order to estimate the coefficients. Moreover, because only $\phi$ and $\sigma_\eta$ enter the hyperparameter vector, the likelihood function needs to be maximised in two directions only, irrespective of the number of explanatory variables. This is very useful since it reduces the dimensionally of the optimisation problem.

## 2.5.5. Smoothing

Once the model parameters have been estimated, interest might centre on obtaining estimates of the volatility process throughout the sample, and beyond. Unlike in the GARCH models, where knowledge of the model parameters is sufficient to construct the volatility figures recursively, in the SV framework the latent volatility can only be estimated.

Several issues are relevant here. First, one may want to construct volatility estimates which take account of the parameter uncertainty. This operation can only be performed within the fully Bayesian MCMC framework where the joint density of latent volatilities and model parameters is readily available. In classical estimation - and in the MCL in particular - the estimated parameters are treated as fixed.

Second, even if the parameters are estimated by the fully efficient MCL, smoothed estimates of the state vector $\alpha_{t|T} = E(\alpha_t|Y)$ in the general state space formulation (2.3) need to be obtained. Conventional state smoothing algorithms do not explicitly take account of the non-Gaussianity in the measurement equation. One solution would be to consider the posterior mode of the state vector, denoted by $\tilde{\alpha}_{t|T}$, rather than its posterior mean $\alpha_{t|T}$. The mode of a random variable is the most probable value, one at which the p.d.f. achieves its maximum. In the present context the posterior mode $\tilde{\alpha}_{t|T}$ is defined as the solution to the vector equation $\partial \ln p(\alpha|Y) / \partial \alpha = 0$ (Durbin and

Koopman, 1997b). The posterior mode is obtained simply by employing the device of equalising density slopes (2.18) and running the state smoother, described in Appendix 2.

The second approach consists of estimating the posterior mean of the signal, $\theta_{t|T} = Z_t \alpha_{t|T}$ rather than of the state vector itself. This is sufficient when only the in-sample estimates of the volatility process are required, as can be seen in (2.28) below. The solution (Durbin and Koopman, 1997b) is given by:

$$\theta_{t|T} = y_t - \sum_{i=1}^{N} f_t \varepsilon_t^{(i)}, \qquad f_t = \frac{w_t^{(i)}}{N \overline{w}_t} \qquad (2.27)$$

where the $\varepsilon^{(i)}$ are drawn from the simulation smoother, $w_t^{(i)}$ are constructed from (2.14). Intuitively, the weights, $f_t$ correct for the non-Gaussianity in the measurement equation.

Finally, if the posterior mean of the state vector itself, $\alpha_{t|T}$ is required, the following procedure may be followed in the case of the basic SV model with explanatory variables. Once the parameters are estimated, $Z_t \hat{\gamma}$ is subtracted from the observations and the state space model reformulated so that the state, $\alpha_t$, the signal, $\theta_t$, and the latent AR(1) process, $h_t$ become identical. Thus equation (2.27) may be applied.

The final difficulty is the fact that the estimation error, $\alpha_{t|T} - \alpha_t$ (where $\alpha_t$ now denotes the true state vector) is $O(1)$. Thus treating $\exp(\alpha_{t|T})$ as lognormal may lead to distortions. This lead Harvey and Shephard (1993) to consider the following estimate of the volatility process. Given a smoothed estimate of the state vector, $\alpha_{t|T}$ throughout the sample, an estimate of the variance of mean adjusted returns, denoted by $\tilde{\sigma}_t^2$ is:

$$\tilde{\sigma}_t^2 = \bar{\sigma}_T^2 \, e^{Z_t \alpha_{t|T}} \,, \qquad \bar{\sigma}_T^2 = T^{-1} \sum_{t=1}^{T} r_t^2 \, e^{-Z_t \alpha_{t|T}} \qquad t = 1, \ldots, T \qquad (2.28)$$

Equation (2.28) demonstrates that for the in-sample volatility estimation the posterior mean of the signal (2.27) is sufficient. Similarly, the $L$-step ahead forecasts of the variance will be given by:

$$\tilde{\sigma}_{\tau+l|\tau}^2 = \bar{\sigma}_\tau^2 \, e^{Z_\tau T_\tau^l \alpha_{\tau|\tau}} \qquad l = 1, \ldots, L \qquad (2.29)$$

where $Z_t$ and $T_t$ are system matrices in the linear state space (2.3), whereas $\tau$ is the final observation time (previously denoted by $T$). In the univariate model with AR(1) dynamics and $k$ explanatory variables, equation (2.29) is specialised to:

$$\tilde{\sigma}_{\tau+l|\tau}^2 = \bar{\sigma}_\tau^2 \, e^{Z_\tau \gamma + \phi^l h_{\tau|\tau}} \qquad l = 1, \ldots, L \qquad (2.30)$$

where $Z_\tau = (z^1_\tau, \ldots, z^k_\tau)'$ now denotes a ($1 \times k$) vector of explanatory variables - including a constant - at the final observation $\tau$, and $\gamma$ is a ($k \times 1$) of vector of coefficients.

## 2.6. Multivariate extensions

The multivariate form of the basic SV model was given in (1.20). Applying the linearising transformation (2.2) yields the linear state space:

$$\begin{cases} Y_t = \ln \bar{\sigma} + h_t + \varepsilon_t & \varepsilon_{i,t} = \ln v_{i,t}^2 \,, \quad v_t \sim NID(0, \Omega_v) \\ h_t = \Phi h_{t-1} + \eta_t & \eta_t \sim NID(0, \Sigma_\eta) \end{cases} \qquad (2.31)$$

where $Y_t$, $\ln \bar{\sigma}$, $h_t$, $\eta_t$ and $\varepsilon_t$ are now ($N \times 1$) vectors, and $N$ is the number of series in the model. However, having assumed the Gaussianity of the $v_t$'s, the new residuals, $\varepsilon_t$, will now have a multivariate $\ln(\chi_1^2)$ distribution. Harvey, Ruiz and Shephard (1994) show that the transformed disturbances, $\varepsilon_{i,t}$, will each have a mean of -1.27 and a

covariance matrix, $\Sigma_\varepsilon$ whose diagonal elements are $\pi^2/2$, while off-diagonal elements are given by:

$$\rho_{i,k}^* = \frac{2}{\pi^2} \sum_{j=1}^{\infty} \frac{\Gamma(\frac{1}{2})(j-1)!}{\Gamma(\frac{1}{2}+j)j} \rho_{i,k}^{2j}, \qquad i,j = 1,\ldots,N \tag{2.32}$$

where $\rho_{i,k}$ are the off-diagonal elements of the correlation matrix $\Omega_v$. Notice, that the

signs of $\rho_{i,k}$ cannot be estimated since the relevant information is lost when the

observations are squared. In QML - as in MCL - we estimate the sign of $\rho_{i,k}$ as positive

if more than half of the pairs $r_{i,t}r_{k,t}$ in (1.20) are positive, and vice versa.

Again, QML yields consistent, alas inefficient estimates of parameter values and full

efficiency can be achieved by MCL. All that is required for the implementation of the

algorithm is the knowledge of the true density of the disturbances. This issue is

addressed in the following two Sections.

### 2.6.1. General case

For simplicity of exposition we focus on the bivariate case here. The analytic

expression for the bivariate $\ln(\chi_i^2)$ density is given in the following proposition:

**Proposition 2.4**: Let $X$ be bivariate standard normal with the correlation coefficient $\rho$.
Then the density of $Y=\ln X^2$ is given by:

$$P_{\ln \chi_i^2}(Y) = \sum_{j=0}^{\infty} K_j(\rho) \prod_{i=1}^{2} \exp\left( y_i(j+0.5) - 0.5(1-\rho^2)^{-1} e^{y_i} \right) \tag{2.33}$$

$$K_j(\rho) = \frac{\rho^{2j}}{\Gamma(j+\frac{1}{2})\Gamma(\frac{1}{2})j! 2^{2j+1}(1-\rho^2)^{2j+\frac{1}{2}}}, \qquad j = 1,\ldots,\infty$$

**Proof**: The bivariate $\chi^2$ density with $\nu$ degrees of freedom is given by:

$$p_{\chi_\nu^2}(Z) = \sum_{j=0}^{\infty} k_j(\rho) \prod_{i=1}^{2} p_{\chi_{\nu+2j}^2}\left(\frac{z_i}{1-\rho^2}\right), \qquad k_j(\rho) = \frac{\Gamma\left(j+\frac{\nu}{2}\right)\left(1-\rho^2\right)^{\frac{\nu}{2}}\rho^{2j}}{\Gamma\left(\frac{\nu}{2}\right)j!}$$

where $p_{\chi_\nu^2}(\cdot)$ denotes a univariate central $\chi^2$ density with $\nu$ degrees of freedom

(Johnson and Kotz, 1972). Setting $\nu=1$ and making the change of variables $Z=\ln Y$ gives

the desired result.$\square$

As in the univariate case it is helpful to visualise the shape of the density, a graph of

which is presented in **Figure 2**.7.
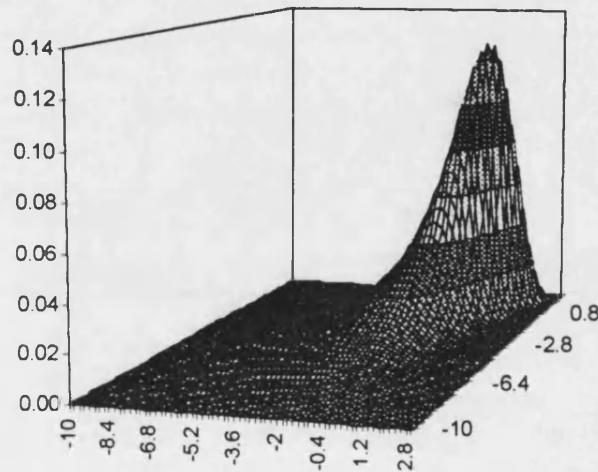


**Figure 2.7:** The bivariate $\ln(\chi_i^2)$ density with a correlation coefficient $\rho=0.9$

Again, the approximation of this density with a bivariate Gaussian, as required by the

QML, will be poor since the true density is highly skewed on the negative half plane.

However, the Gaussian density will provide a reasonable sampling density for the

MCL. The quantities $w_t^{(i)}$ in (2.16) are computed as:

$$w_t^{(i)} = \left|H_t\right|^{1/2} \exp\left\{0.5\left(\varepsilon_t^{(i)'}H_t^{-1}\varepsilon_t^{(i)} - \frac{i'e^{\varepsilon_t^{(i)}}}{\left(1-\rho^2\right)}\right)\right\} \sum_{j=0}^{K} K_j \exp\left(i'\varepsilon_t^{(i)}(j+0.5)\right) \qquad (2.16''')$$

69

where $i$ is a (2×1) unit vector, $\varepsilon_t^{(i)}$ is (2×1), and $K$ is a constant at which the infinite sum in (2.33) is truncated. The structure of the (2×2) mean covariance matrix $H_t = \Sigma_\varepsilon$ was described in (2.32).

Again, a suitable truncation value $\kappa$ needs to be chosen. **Figure 2.7** below depicts the surface of the weight function $K_j(\rho)$ in (2.33). Observe that different values of $\rho$ will allow for different cut-off points $K$. However, $K=10$ appears to be a reasonable upper bound, after which the contribution of further $\ln(\chi_I^2)$ densities in the summation (2.33) is negligible.
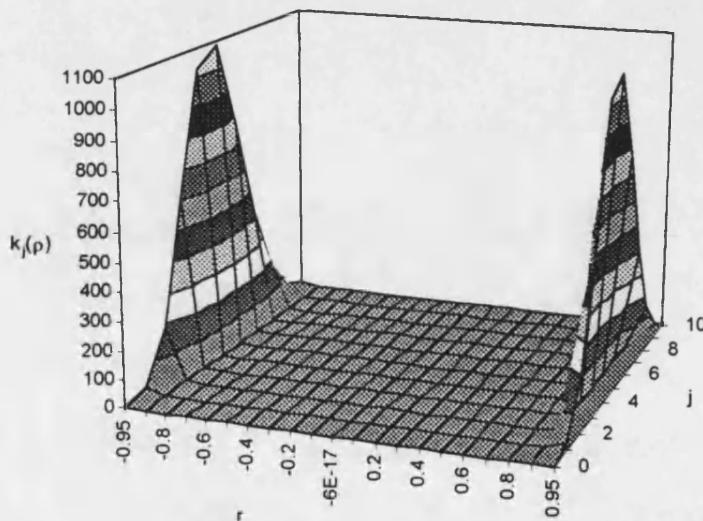


**Figure 2.8:** Weight function $K_j(\rho)$ in the bivariate $\ln(\chi_I^2)$ distribution.

It has been shown earlier that the computational efforts can be reduced when $w_t^{(i)}$ in (2.14) are subjected to Taylor series expansion. Similarly the quality of the sampling density $p_G(\varepsilon|\psi)$ is improved by equalising the slopes of the $p_G(\varepsilon|\psi)$ and $p_{\ln\chi_1^2}(\varepsilon|\psi)$ in the neighbourhood of $\hat{\varepsilon}_t = E(\varepsilon_t|y)$. This amounts to equating the first derivatives of the $\ln p_G(\varepsilon|\psi)$ and $\ln p_{\ln\chi_1^2}(\varepsilon|\psi)$ and solving for the matrix $H_t$ which becomes time varying.

These techniques require analytic expressions for the first, second and fourth derivatives of the logdensity, $\ln p_{\ln \chi_i^2}(Y)$. However, the infinite sum in (2.33) makes it difficult to find tractable analytic expressions for the derivatives of the logdensity. And estimation may therefore proceed directly by calculating (2.16''') and (2.15).

### 2.6.2. Uncorrelated case

Considerable simplification can be achieved when the mean equation disturbances $v_{1,t}$ and $v_{2,t}$ in (1.20) are uncorrelated, i.e. $\Omega_v = I_N$ in (2.31). In this case, the bivariate $\ln(\chi_i^2)$ density, $p_{\ln \chi_i^2}(Y)$ takes a simple form:

$$p_{\ln \chi_i^2}(Y) = \prod_{i=1}^{2} \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{y_i - e^{y_i}}{2}\right\}$$

This follows from Proposition 2.4, by setting $\rho=0$, or directly from Proposition 2.1, by observing that the joint density of two independent variables is a product of their individual densities. Therefore, the first, second and fourth derivatives of the logdensity, $\ln p_{\ln \chi_i^2}(Y)$ are:

$$d_1(Y)' = \frac{1}{2}\begin{bmatrix} 1-e^{y_1} \\ 1-e^{y_2} \end{bmatrix}, \quad d_2(Y) = \frac{1}{2}\begin{bmatrix} -e^{y_1} & 0 \\ 0 & -e^{y_2} \end{bmatrix}, \quad d_4(Y) = \frac{1}{2}\begin{bmatrix} -e^{y_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -e^{y_2} \end{bmatrix}.$$

Equalising the density slopes, the diagonal matrices $\tilde{H}_t$ are chosen so as to satisfy:

$$\tilde{H}_{ii,t} = \frac{2\hat{\varepsilon}_{i,t}}{e^{\hat{\varepsilon}_{i,t}} - 1} \qquad t = 1,\ldots,T \qquad i = 1,2 \tag{2.18'''}$$

which is an extension of (2.18). Letting $\zeta_t = \hat{\varepsilon}_t - \varepsilon_t^{(i)}$ one obtains the multivariate equivalents of the Taylor series expansions (2.19) and (2.20):

$$c(\zeta) = \frac{1}{2!} \sum_{t=1}^{T} \zeta_t' \left( d_2(\hat{\varepsilon}_t) + \tilde{H}_t^{-1} \right) \zeta_t + \frac{1}{4!} \sum_{t=1}^{T} vec(\zeta_t \zeta_t')' d_4(\hat{\varepsilon}_t) vec(\zeta_t \zeta_t')$$

$$Ec = \frac{1}{2} \sum_{t=1}^{T} tr \left[ \left( d_2(\hat{\varepsilon}_t) + \tilde{H}_t^{-1} \right) E\zeta_t \zeta_t' \right] + \frac{1}{8} \sum_{t=1}^{T} tr \left[ d_4(\hat{\varepsilon}_t) E\left[ vec(\zeta_t \zeta_t') vec(\zeta_t \zeta_t')' \right] \right]$$

where the quantities $E\zeta_t \zeta_t' = Var(\varepsilon_t | Y)$ are again obtained from the disturbance smoother. Notice, that matrix of fourth moments, $E\left[ vec(\zeta_t \zeta_t') vec(\zeta_t \zeta_t')' \right]$, need not be evaluated completely since $d_4(Y)$ selects merely the first and fourth diagonal elements of this matrix.

Finally, the univariate quantities $w_t^{(i)}$ in (2.14) and (2.16) can now be computed as:

$$w_t^{(i)} = \left| \tilde{H}_t \right|^{1/2} \exp\left( 0.5 \left( i' \varepsilon_t^{(i)} - i' e^{\varepsilon_t^{(i)}} + \varepsilon_t^{(i)'} \tilde{H}_t^{-1} \varepsilon_t^{(i)} \right) \right) \tag{2.16''''}$$

where $i$ is a (2×1) unit vector, $\varepsilon_t^{(i)}$ is (2×1), and $\tilde{H}_t$ is (2×2).

## 2.7. Conclusion

This Chapter proposes a new method of estimating stochastic volatility models. At the core of the procedure is the representation of the model in a linear state space form. Kalman filter can then be applied to yield the prediction error decomposition which in turn constitutes the Gaussian likelihood function. However, due to the log chi-square disturbances in the measurement equation, the Gaussian likelihood will only make up a part of the true likelihood function. The expectation of the remainder term is simulated. As the number of simulations ($N$) increases, the approximation becomes more accurate. The final sample performance of the MCL algorithm is examined in a simulation study.

The results indicate full efficiency of the estimator across a range of possible parameter values even for moderate simulation sizes, $N=5$.

Apart from computational efficiency, the advantage of the approach lies in the formulation of the model in state space from. This allows the basic model to be extended in a number of directions likely to arise in empirical research. One such extension is the inclusion of explanatory variables in the variance equation which shall be examined in more detail in the following Chapters. Once the model is in the state space form, no modifications of the estimation procedure are required beyond determining the analytical form of the true density function of the disturbances.

**Table 2.1. Equalising density slopes: a recursive solution**

| | $\psi_1$ | $\psi_2$ | $\psi_3$ | $\psi_4$ | $\psi_5$ | $\psi_6$ | $\psi_6$ | $\psi_8$ | $\psi_9$ |
|---|---|---|---|---|---|---|---|---|---|
| d(1) | 3.135 | 3.181 | 3.225 | 3.215 | 3.239 | 3.260 | 3.281 | 3.285 | 3.292 |
| d(2) | 1.309 | 1.417 | 1.516 | 1.490 | 1.546 | 1.602 | 1.634 | 1.654 | 1.676 |
| d(3) | 0.264 | 0.281 | 0.295 | 0.298 | 0.302 | 0.306 | 0.310 | 0.311 | 0.311 |
| d(4) | 0.144 | 0.131 | 0.110 | 0.125 | 0.107 | 0.0090 | 0.0087 | 0.0079 | 0.0070 |
| d(5) | 0.0049 | 0.0042 | 0.0031 | 0.0038 | 0.0029 | 0.00223 | 0.0021 | 0.0018 | 0.0015 |
| d(6) | 0.0023 | 0.0016 | 0.0010 | 0.0014 | 0.0009 | 0.0006 | 0.0005 | 0.0004 | 0.0003 |
| d(7) | 0.00099 | 0.00064 | 0.00034 | 0.00050 | 0.0003 | 0.0002 | 0.0001 | 0.0001 | 0.0001 |
| d(8) | 0.00048 | 0.00027 | 0.00011 | 0.00019 | 0.00009 | 0.00004 | 0.00003 | 0.00002 | 0.00001 |
| d(9) | 0.00026 | 0.00011 | 0.00004 | 0.00007 | 0.00003 | 0.00001 | 8e-6 | 6e-6 | 31e-6 |
| d(10) | 0.00012 | 0.00005 | 0.00001 | 0.00003 | 0.00001 | 3e-6 | 2e-6 | 1e-6 | 1e-6 |

This table reports the values of the metric $d(k) = T^{-1}\sum_{t=1}^{T}\left|\tilde{H}_t^{(k)} - \tilde{H}_t^{(k-1)}\right|$ for $k=1, ..,$ 10 iterations of the simulated SV model with $T=1,000$ and across several parameter triplets $\psi_i = (\alpha, \phi, \sigma_\eta^2)_i$. Small values of $d(k)$ indicate that the individual elements of the variance vector $\tilde{H}$ are not changing considerably across further iterations, i.e. $\tilde{H}_i^{(k)} \to \bar{H}_t$.

# Table 2.2. Comparison between MCMC, QML and MCL estimators

| CV=10 | | $\psi_1$ | | | $\psi_2$ | | | $\psi_3$ | |
|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | $\phi$ | $\sigma_\eta$ | $\alpha$ | $\phi$ | $\sigma_\eta$ | $\alpha$ | $\phi$ | $\sigma_\eta$ |
| **True** | -0.821 | 0.9 | 0.675 | -0.411 | 0.95 | 0.484 | -0.164 | 0.98 | 0.308 |
| **MCMC** | -0.679 (0.22) | 0.916 (0.026) | 0.562 (0.12) | -0.464 (0.16) | 0.94 (0.02) | 0.46 (0.055) | -0.19 (0.08) | 0.98 (0.01) | 0.35 (0.06) |
| **QML** | -0.945 (0.471) | 0.885 (0.057) | 0.703 (0.177) | -0.513 (0.376) | 0.937 (0.046) | 0.506 (0.138) | -0.250 (0.267) | 0.969 (0.032) | 0.324 (0.098) |
| **MCL** | -0.662 (0.19) | 0.907 (0.026) | 0.621 (0.075) | -0.360 (0.139) | 0.949 (0.019) | 0.458 (0.062) | -0.174 (0.112) | 0.975 (0.016) | 0.300 (0.05) |

| CV=1 | | $\psi_4$ | | | $\psi_5$ | | | $\psi_6$ | |
|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | $\phi$ | $\sigma_\eta$ | $\alpha$ | $\phi$ | $\sigma_\eta$ | $\alpha$ | $\phi$ | $\sigma_\eta$ |
| **True** | -0.736 | 0.9 | 0.363 | -0.368 | 0.95 | 0.26 | -0.147 | 0.98 | 0.1657 |
| **MCMC** | -0.87 (0.34) | 0.88 (0.046) | 0.35 (0.067) | -0.56 (0.34) | 0.92 (0.046) | 0.28 (0.065) | -0.22 (0.14) | 0.97 (0.02) | 0.23 (0.08) |
| **QML** | -1.002 (0.91) | 0.864 (0.122) | 0.410 (0.228) | -0.591 (0.756) | 0.920 (0.101) | 0.302 (0.176) | -0.330 (0.619) | 0.955 (0.083) | 0.200 (0.142) |
| **MCL** | -0.598 (0.274) | 0.904 (0.044) | 0.336 (0.076) | -0.327 (0.186) | 0.947 (0.030) | 0.249 (0.059) | -0.163 (0.155) | 0.974 (0.025) | 0.163 (0.046) |

| CV=0.1 | | $\psi_7$ | | | $\psi_8$ | | | $\psi_9$ | |
|---|---|---|---|---|---|---|---|---|---|
| | $\alpha$ | $\phi$ | $\sigma_\eta$ | $\alpha$ | $\phi$ | $\sigma_\eta$ | $\alpha$ | $\phi$ | $\sigma_\eta$ |
| **True** | -0.706 | 0.9 | 0.135 | -0.353 | 0.95 | 0.0964 | -0.1412 | 0.98 | 0.0614 |
| **MCMC** | -1.54 (1.35) | 0.78 (0.19) | 0.15 (0.082) | -1.12 (1.15) | 0.84 (0.16) | 0.12 (0.074) | -0.66 (0.83) | 0.91 (0.12) | 0.14 (0.099) |
| **QML** | -1.007 (1.33) | 0.858 (0.19) | 0.165 (0.203) | -0.890 (1.40) | 0.875 (0.20) | 0.153 (0.19) | -0.688 (1.21) | 0.903 (0.17) | 0.129 (0.179) |
| **MCL** | -0.884 (1.51) | 0.848 (0.21) | 0.102 (0.097) | -0.641 (1.34) | 0.890 (0.23) | 0.088 (0.079) | -0.453 (1.11) | 0.922 (0.19) | 0.067 (0.061) |

This table reports the results of the simulation experiments. For each set of parameter triplets $\psi_i=(\alpha,\phi,\sigma^2_\eta)_i$, samples of length $T$=500 of the basic SV model (2.1) are generated $K$=500 times. The model is then estimated by various techniques and the average estimated parameter values (and their standard deviations) are presented in each row. The results for the MCMC estimator are reproduced from Jacquier, Polson and Rossi (1994), Table 7.

**Table 2.3. Performance of the MCL estimator**

| CV=1 | | $\psi_s$ | | |
|---|---|---|---|---|
| | $\alpha$ | $\phi$ | $\sigma_\eta$ | *time* |
| **True** | -0.368 | 0.95 | 0.26 | |
| **MCMC** | -0.56 (0.34) | 0.92 (0.046) | 0.28 (0.065) | - |
| **QML** | -0.591 (0.756) | 0.920 (0.101) | 0.302 (0.176) | 1 |
| **MCL (N=10)** | -0.388 (0.378) | 0.947 (0.051) | 0.234 (0.126) | 27.66 |
| **MCL (N=20)** | -0.390 (0.378) | 0.947 (0.051) | 0.232 (0.109) | 49.49 |
| **MCL\* (N=5)** | -0.327 (0.186) | 0.947 (0.030) | 0.249 (0.059) | 27.67 |
| **MCL\*(N=10)** | -0.325 (0.180) | 0.947 (0.029) | 0.249 (0.059) | 35.08 |

This table reports the results of the simulation experiment on a single set of parameter values, $\psi_s$. Samples of length $T=500$ of the basic SV model (2.1) are generated $K=500$ times and estimated by MCL. Values 10, 20, and 5 in parenthesis signify the number of draws, $N$, employed by taking the expectation in (2.15). When the device of equalising density slopes (Equation (2.18)) has been used, a * appears. The final column reports the average relative time until convergence as a multiple of the QML speed.

## Table 2.4. Inlier problem: Sensitivity of the QML and MCL estimators

**Panel A:** Sensitivity of QML w.r.t. cut-off value $\kappa$

| | $\alpha$ | $\phi$ | $\sigma_\eta$ |
|---|---|---|---|
| **True $\psi_s$** | -0.368 | 0.95 | 0.26 |
| **Full sample** | -0.591 | 0.920 | 0.302 |
| | (0.756) | (0.101) | (0.176) |
| $\kappa_1 = 4.54*10^{-5}$ | -1.393 | 0.836 | 0.966 |
| | (1.144) | (0.132) | (0.292) |
| $\kappa_2 = 3.06*10^{-7}$ | -0.835 | 0.912 | 0.888 |
| | (0.877) | (0.093) | (0.121) |
| $\kappa_3 = 1.93*10^{-22}$ | -0.851 | 0.948 | 0.848 |
| | (0.649) | (0.040) | (0.099) |
| $\kappa_4 = 3.72*10^{-44}$ | -1.360 | 0.949 | 0.849 |
| | (1.731) | (0.063) | (0.104) |

**Panel B:** Sensitivity of MCL w.r.t. cut-off value $\kappa$

| | $\alpha$ | $\phi$ | $\sigma_\eta$ |
|---|---|---|---|
| **True $\psi_s$** | -0.368 | 0.95 | 0.26 |
| **Full sample** | -0.327 | 0.947 | 0.249 |
| | (0.186) | (0.030) | (0.059) |
| $\kappa_1 = 4.54*10^{-5}$ | -0.845 | 0.869 | 0.396 |
| | (0.698) | (0.107) | (0.142) |
| $\kappa_2 = 3.06*10^{-7}$ | -0.854 | 0.867 | 0.411 |
| | (0.548) | (0.085) | (0.131) |
| $\kappa_3 = 1.93*10^{-22}$ | -0.601 | 0.906 | 0.370 |
| | (0.286) | (0.044) | (0.116) |
| $\kappa_4 = 3.72*10^{-44}$ | -0.575 | 0.910 | 0.365 |
| | (0.280) | (0.043) | (0.111) |

This table reports the results of the simulation experiment on a single set of parameter values, $\psi_s$. Samples of length $T=500$ of the basic SV model (2.1) with approximately 10% zero values are generated $K=500$ times and estimated by QML and MCL. Inliers are cut-off at $\kappa_i$ in accordance with (2.24) where the cut-off constants $\kappa_1$-$\kappa_4$ were chosen so as to correspond to $\ln(\kappa_i^2)=$ -20, -30, -100, and -200.

**Table 2.5. Estimates of the SV model with fat-tailed disturbances**

| | $\phi$ | $\sigma^2_\eta$ | $\ln\bar{\sigma}^2$ | $v$ | $CV$ | $LogLik$ | $LR$ |
|---|---|---|---|---|---|---|---|
| $\hat{\psi}$ | 0.960 | 0.026 | -9.313 | - | 0.389 | -4311.6 | 26.6 |
| $s.e.(\hat{\psi})$ | (0.018) | (0.009) | (0.094) | - | - | - | - |
| $\hat{\psi}$ | 0.984 | 0.007 | -9.498 | 7.634 | 0.255 | -4298.3 | - |
| $s.e.(\hat{\psi})$ | (0.010) | (0.003) | (0.122) | (0.003) | - | - | - |

This table reports the estimation results of the SV model where the mean equation disturbances follow a *Student-t* distribution with $v$ degrees of freedom. The standard errors of $(\phi, \sigma_\eta^2, v)$ are obtained from the numerical approximation to the Hessian, while the standard errors of the estimate of $\ln\bar{\sigma}^2$ are taken from the corresponding diagonal element of $P_T$. The likelihood ratio test statistic follows the $\chi_1^2$ distribution.

# Chapter 3: Inference in the SV model

## 3.1. Introduction

It was shown in Chapter 2 how efficient parameter estimates can be obtained by correcting the Gaussian (QML) likelihood function for the $\ln\chi^2$ distribution. This difficulty arises when the original observations are transformed to give a linear state space. The true likelihood function is decomposed into the Gaussian part and a remainder term, which is computed by simulation. A finite sample simulation experiment suggested that the performance of the estimator is at least as good - and often better - than other fully efficient estimation procedures available in the literature.

One further aspect of the method needs more detailed attention. Apart from efficient parameter estimates the procedure delivers arbitrarily accurate approximations to the likelihood function itself. This opens the possibility of likelihood ratio hypothesis testing.

In general, the object of hypothesis testing is to derive test statistics which indicate the reasonableness of some hypothesis ($H_0$) being true. If the data fall into a particular region of the sample space (critical region) then the test is said to reject the null hypothesis, otherwise the test fails to reject. Because there are only two possible outcomes, there are only two ways to make incorrect inferences. *Type I* errors arise when the null hypothesis is falsely rejected. *Type II* errors occur when the null is incorrectly not rejected. Comparison and evaluation of tests is based upon the notions of *size* (the probability of rejecting the null, when it is true) and *power* (the probability of rejecting the null, when it is false). A test is preferred if it has maximum power among all tests with size less than or equal to some particular level.

This Chapter is organised as follows. Section 3.1.1. briefly reviews significance tests and suggests how confidence intervals of model parameters can be constructed. Section 3.2. shows how the volatility process can be tested for the presence of a unit root. The new estimation method can be used to construct likelihood ratio tests of the hypothesis $H_0$: $\phi=1$. However, the asymptotic distribution of the test statistic is unknown. We simulate this density and tabulate critical values. The power of the test is compared to that of the augmented Dickey-Fuller test. Section 3.3. illustrates how the AR(1) specification for the (log)variance process can be tested against higher order dynamics. Section 3.4 concludes.

### 3.1.1. Significance tests and confidence intervals

The object of the maximum likelihood procedure is to obtain an efficient point estimate of the parameter vector, denoted by $\hat{\psi}$. After the estimation, the focus is usually centred around testing of hypotheses concerning individual elements of $\hat{\psi}$. This is conventionally accomplished by means of $t$-tests:

$$t_i = \frac{\hat{\psi}^i - \tilde{\psi}^i}{SE(\hat{\psi}^i)}$$

where $SE$ stands for the estimated standard error, the superscript, $i$, on $\psi$ points to a particular element of the parameter vector, and $\tilde{\psi}^i$ is a fixed value of that parameter, e.g. zero. In large samples the limiting distribution is $N(0,1)$. For reasons that will become apparent in the following Section, for now, we concentrate on $\tilde{\psi}^i$ strictly *within* the parameter space.

Thus, for instance the significance of the explanatory variables in (1.16) or (1.18) can be tested by dividing the parameter estimate by the standard deviation and comparing

to the critical value of the limiting normal distribution.[1] In the general state space model, the standard deviation is obtained from the diagonal element of the state covariance matrix, $P_T$ (Appendix 2). This is a test of $H_0$: $\psi^j=0$ against a two-sided alternative $H_1$: $\psi^j\neq0$.

We can also test whether $\phi$ or $\sigma_\eta$ take specific values, say $\tilde{\phi}<1$ or $\tilde{\sigma}_\eta>0$, by the same method. This is justified because the ML estimators $\hat{\phi}$ and $\hat{\sigma}_\eta$ are normally distributed, even though the measurement equation noise is non-Gaussian (Dunsmuir, 1979). The standard deviations of the hyperparameters can be obtained from the numerical estimate of the information matrix. Asymptotic $100(1-\varepsilon)$ percent confidence intervals for the estimates of $\phi$ and $\sigma_\eta$ can be constructed in accordance with:

$$\hat{\psi}^i \pm z_{0.5\varepsilon} SE(\psi^i) \qquad (3.1)$$

where $z_{0.5\varepsilon}$ is the $0.5\varepsilon$ point of the normal distribution. This gives a pair of values $\psi^{i'}$, $\psi^{i''}$ which are symmetrically centred around $\hat{\psi}^i$. Notice, however, that when the parameter estimate is close to the boundary of the parameter space (e.g. $\hat{\phi}=0.98$) this may not be desirable. First, it may well happen that one of the numerically computed values will be outside the parameter space (i.e. greater than unity). Second, it has been shown in Chapter 2 (**Figures 2.3** and **2.4**) that in finite samples the posterior density of parameter estimates is skewed away from the boundary. One would therefore expect the confidence interval to be described by a pair of values which are *asymmetrically* placed around the estimated coefficient. For a given sample size $T$ this asymmetry will be more pronounced, the closer the coefficient estimate is to the boundary.

---

[1] However, because of the non-Gaussianity of the measurement equation disturbances but linear filtering the standard errors obtained in this fashion will be consistent but inefficient. A better test statistic can be constructed based on the likelihood ratio principle since the MCL delivers arbitrarily accurate approximations to the likelihood function value.

During the estimation the general problem of constrained likelihood optimisation is translated to the unconstrained optimisation problem by transforming the parameter space (Hamilton, 1994, p.146; Koopman *et al*, 1995, p. 209). Specifically, unrestricted optimisation is performed on a set of parameters $\theta$ which are related to the true parameters via $\theta = h^{-1}(\psi)$. The vector function, $h(\cdot)$ which incorporates the necessary restrictions is given by: $\psi_\phi = \theta_\phi \left(1 + \theta_\phi^2\right)^{-0.5}$, $\psi_\sigma = \exp(\theta_\sigma)$ where $\theta_\phi$ and $\theta_\sigma$ denote the elements of $\theta$ corresponding to $\phi$ and $\sigma_\eta$. More details can be found in Appendix 3.
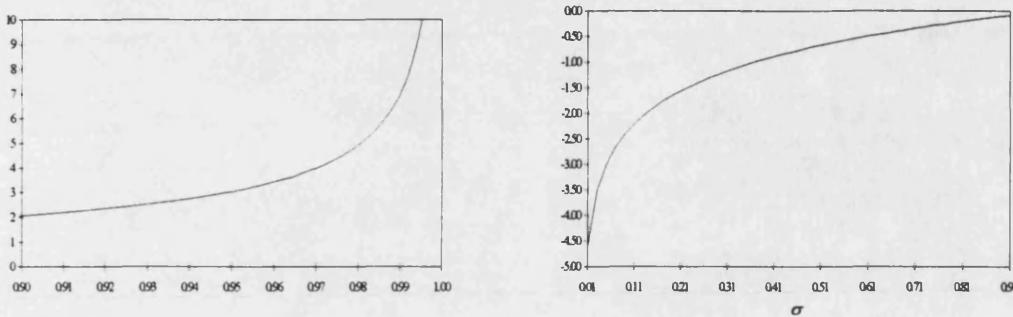


**Figure 3.1:** Imposition of parameter restrictions.

**Figure 3.1** shows how, in the case of the basic univariate SV model, the restricted parameters $\psi = (\phi, \sigma_\eta)'$ - on the horizontal axis - are mapped into unrestricted ones via $\theta = h^{-1}(\psi)$. Optimisation delivers $\hat{\theta}$ and $\text{var}(\hat{\theta})$, from which $\hat{\psi}$ and its covariance matrix are constructed in accordance with:

$$\hat{\psi} = h(\hat{\theta}), \qquad \text{var}(\hat{\psi}) = \frac{\partial h(\hat{\theta})}{\partial \theta'} \text{var}(\hat{\theta}) \frac{\partial h(\hat{\theta})}{\partial \theta}$$

Now, instead of constructing the confidence interval on $\hat{\psi}$ according to the equation above, we may choose to construct a confidence interval on $\hat{\theta}$ giving a pair $\theta''$, $\theta'''$ for each element of the parameter vector, $i$. It can be seen from **Figure 3.1** that the non-linearity of $h(\cdot)$ induces the corresponding pairs $\psi''$, $\psi'''$ to be asymmetric

82

around the point estimate $\hat{\psi}$. The resulting confidence intervals on $\phi$ and $\sigma_\eta$ will be skewed away from the respective boundaries (1 and 0). This may prove useful in empirical research.

This method, however, is not applicable when the true parameter is posed to be on the boundary of the parameter space, e.g. $\tilde{\phi}=1$ or $\tilde{\sigma}_\eta=0$, and different procedures are required. The latter case is a test for the presence of the time varying volatility against a fixed level (Harvey and Streibel, 1997). The former is a test for the unit root in the volatility process and is addressed in the following Section.

## 3.2. Testing for a unit root in the volatility process

Empirical applications of SV models document one striking regularity: the autoregressive parameter, $\phi$ is often found to be close to unity. Similarly, in the GARCH literature the sum of the parameters on the lagged squared residuals, $\alpha_i$, and on lagged conditional variances, $\beta_i$, are often found to sum up close to unity (Engle and Bollerlev, 1986).[2]

Estimates of SV model parameters reported in previous studies are reproduced in Table 3.1. With empirical estimates of $\phi$ well above 0.9 a natural question arises as to the stationarity of the volatility process. While estimation and statistical analysis of SV models with $\phi=1$ is possible, the stationary version is more appealing on several grounds. First, the mean of the process can be consistently estimated by OLS. Even though, as has been mentioned in Chapter 1, the dynamic properties in the mean are less pronounced, some interest may still be focused on regression effects, as in e.g. equation (1.2'''). Second, a model of asset prices with infinite unconditional variance,

---

[2] Nelson (1990b), however, showed that the long run behaviour of such models is unsatisfactory: depending on the value of the intercept term, the variance process tends to 0 or $\infty$.

which can be seen by setting $\phi=1$ in (1.14), poses difficulties for the framework of most of the modern finance theory. In particular, the portfolio theory and the CAPM are based upon the existence of unconditional second moments. And finally, there are a number of empirical observations which indicate that the true volatility process is mean reverting. Thus, for instance, the implied volatility of options - in as far as it can be regarded as an indicator of the true volatility process - is strongly mean reverting, as we discuss in Chapter 4.

The non-stationarity in the mean of the process has been studied at some length (Stock, 1993; Hamilton, 1994). Testing for the presence of a unit root against a stationary alternative usually proceeds by means of augmented Dickey-Fuller tests. The discussion in Kim and Schmidt (1993) indicates that the tests are valid under quite general conditions regarding the distribution of the error term. Most importantly, only the independence of the disturbances - and not their Gaussianity - is required.

In principle therefore, the tests are applicable to the problem of testing for the presence of the unit root in the (log)variance process. However, with the empirical estimates of the signal-noise ratio, $q$ being relatively small, the reduced form of the linearised stochastic volatility model (2.2) - under $H_0:\phi=1$ - will resemble an ARIMA(0,1,1) model with the moving-average parameter close to the non-invertibility region. The isomorphism is established (Harvey, 1989, p. 68) by:

$$\theta = 0.5\left[\left(q^2 + 4q\right)^{\frac{1}{2}} - 2 - q\right], \qquad q = 2\sigma_\eta^2 / \pi^2$$

In such circumstances, the Dickey-Fuller tests are known to be oversized (Phillips and Perron, 1988; Schwert, 1989; Pantula, 1991) rendering the tests based on the tabulated critical values unreliable. Their performance is examined in Section 3.2.1.

Alternative testing procedures for the presence of the unit root were proposed by Nyblom and Makelainen (1983) and Kwiatowski *et al.* (1992).[3] In both cases the process has a unit root under the alternative hypothesis. However, under the null, the process is level stationary in Nyblom and Makelainen (1983), and trend stationary in Kwiatowski *et al.* (1992). Thus despite the validity of the tests under non-Gaussianity of the disturbances neither test procedure is applicable if a test for a unit root against an autoregressive alternative is required.

Summing up, on the methodological level there is clearly a need for a test of a unit root $(H_0:\phi=1)$ in the volatility process against a stationary alternative $(H_1:\phi<1)$. On the empirical level, on the other hand, estimates of the autoregressive coefficient close to unity necessitate formal testing of the unit root hypothesis. Such a test can be constructed based on the likelihood ratio testing principle. Its empirical distribution is investigated in the remainder of the Section and contrasted with the augmented Dickey-Fuller tests.

### 3.2.1. Distribution of the Dickey-Fuller tests

Dickey-Fuller tests are based on the regressions of the form

$$y_t = \alpha + \varphi y_{t-1} + \sum_{i=1}^{p} \psi_i \, \Delta y_{t-1} + \varepsilon_t \qquad (3.2)$$

The process $y_t$ has a single unit root in the autoregressive polynomial under the null.[4] Consequently, $\hat{\varphi}$ will tend to be negative but close to zero. Under the null the distribution of $\hat{\varphi}$ is given by a Brownian Bridge (Hamilton, 1994, p. 486) and therefore the test statistics

---

[3] See also Tanaka (1983) and Watson and Engle (1985).
[4] Alternatively. $v_t=v_{t-1}+\varepsilon_t$ is a random walk starting at $y_0=\alpha$, in which case lagged differences were included in an attempt to account for more general error distribution.

$$ADF_1 = \frac{\hat{\varphi} - 1}{\hat{\sigma}_\varphi} \qquad (3.3a)$$

$$ADF_2 = \frac{T(\hat{\varphi} - 1)}{1 - \sum_{i=1}^{p} \hat{\psi}_i} \qquad (3.3b)$$

will have the non-standard distributions tabulated in Fuller (1976). The tests may be applied to the linearised SV model (2.2) by setting $y_t = \ln r_t^2$ in (3.2), where $r_t$ is the mean adjusted return on the asset.

Monte Carlo evidence in Phillips and Perron (1988) and Schwert (1989) suggests that the tests in (3.3) exhibit non-trivial size distortions when the data generating process is Gaussian ARIMA(0,1,1) with the MA parameter, $\theta$ close to the non-invertibility region. Intuitively, a large number of lags will be required in the autoregression (3.2) to achieve a reasonable approximation to the mixed process, even if the errors were Gaussian. For instance, Said and Dickey (1984) advocate values of $p$ of the order $T^{1/2}$. We expect similar behaviour in the present context, possibly exacerbated by the $\ln(\chi_1^2)$ distribution of the error term.

The size of the tests (3.3) in the context of SV model is examined in a Monte Carlo experiment. Data samples of variable lengths, $T_1 = 500$ and $T_2 = 1,000$, are drawn from the basic SV model (2.1). The choice of the sample length, $T$ is motivated by the trade-off between the sample sizes likely to arise in empirical applications and the computational constraints of the Monte Carlo experiment. While far larger sample sizes are feasible in the context of augmented Dickey-Fuller tests, the likelihood ratio test - considered in the subsequent Section - requires the MCL estimation of the SV model under the null as well as under the alternative leading to a very large computational effort.

Since the interest is focused on the distribution of the test statistic under the null, the autoregressive parameter in the (log)variance process is set to unity. The volatility parameter, $\sigma_\eta$ takes the values 0.1 and 0.3 corresponding roughly to upper and lower bounds of the empirical estimates reported in **Table 3.1**. The values of $\sigma_\eta$ imply reduced form MA parameter values of -0.96 and -0.87 respectively.[5] Under the null the unconditional variance of the return process is not defined and the value of the long run volatility level, $\bar{\sigma}$ is undetermined. However, since $\bar{\sigma}$ is merely a scale parameter it will be irrelevant to the distribution of the test statistic. We set $\bar{\sigma} = 0.014$ corresponding to a value likely to arise in empirical applications. This choice of parameters gives four cases:

|            | $\sigma_\eta=0.1$ ($q=0.002$) | $\sigma_\eta=0.3$ ($q=0.018$) |
|------------|:-------------:|:-------------:|
| $T=500$    | **C1**        | **C2**        |
| $T=1000$   | **C3**        | **C4**        |

For each case, $K=10,000$ sample paths were drawn. For each realisation of the process, the Dickey-Fuller test statistics were calculated based on variable lengths of the autoregressive polynomial in (3.2), $p=10, 20, 30, 40, 50$. Since deterministic time trends in the (log)variance process do not constitute a sensible model for asset returns the null hypothesis is a random walk without the drift. The critical values for the nominal sizes of 10, 5, and 1% are given by Fuller (1976, pp. 371,373):

|          |           | 0.10   | 0.05   | 0.01   |
|----------|-----------|:------:|:------:|:------:|
| **ADF₁** | $T=500$   | -2.57  | -2.87  | -3.44  |
|          | $T=1,000$ | -2.57  | -2.86  | -3.43  |
| **ADF₂** | $T=500$   | -11.2  | -14.0  | -20.5  |
|          | $T=1,000$ | -11.3  | -14.1  | -20.7  |

---

[5] The process is non-invertible with $q=0$ and $\theta=-1$

The null hypothesis is rejected if the test statistic is *less* than the critical value and is not rejected otherwise. The rejection frequencies are reported in **Table 3.2**. Three conclusions can be drawn. First, the distribution of the $ADF_2$ test (3.3b) is more adversely affected than the distribution of the $ADF_1$ test (3.3a). Irrespective of the value of $\sigma_\eta$ and the sample length, $T$ the size distortions are uniformly more severe for the $ADF_2$ test. This is not mitigated by higher lags of the autoregressive polynomial. The same conclusion was reached by Schwert (1989). By contrast, the size distortions of the $ADF_1$ test can be corrected by increasing the number of lags, $p$. Second, for low $\sigma_\eta$ (0.1) a large number of lags, $p$ (50) is required to induce the size of the test statistic to approach its nominal value. Even then the test remains slightly oversized irrespective of the length of the time series. Finally, for higher values of $\sigma_\eta$ (0.3), lower orders of the polynomial are required ($p$=30), again, irrespective of the length of the time series. This is not surprising, since lower values of $\sigma_\eta$ imply MA models with the coefficient closer to the non-invertibility region.

Overall, the $ADF_1$ test (3.3a) appears to exhibit tolerable size distortions for very large lag lengths ($p$=50 for $\sigma_\eta$=0.1, and $p$=30 for $\sigma_\eta$=0.3). This is the preferred test whose power will be compared with the likelihood ratio test developed in the next Section.

### 3.2.2. Distribution of the likelihood ratio test

In general, the likelihood ratio test, $\xi_{LR}$, is concerned with testing the validity of restrictions of the form:

$$f(\psi) = 0 \qquad\qquad (3.4)$$

where $\psi$ is a ($p\times 1$) vector of model parameters, and $f(\cdot)$ denotes some - possibly nonlinear - matrix function such that the matrix of first derivatives, $\partial f(\cdot)/\partial\psi$, has rank $m$. When the restrictions are linear, (3.4) takes the form: $R\psi - r = 0$ where $R$ and $r$

are ($m \times p$) and ($m \times 1$) matrices of fixed values respectively (Engle, 1984; Harvey, 1989; p. 234; Gouriéroux and Monfort, 1995, pp. 82).

Under the null hypothesis, $H_0$, the parameter vector, $\psi$ satisfies the restriction (3.4). Denote the estimate of the parameter values under the null, by $\hat{\psi}_0$ and the value of the maximised likelihood function by $L(\hat{\psi}_0)$. By contrast, under the alternative, $H_1$, the unrestricted parameter estimate and the corresponding likelihood function value are denoted by $\hat{\psi}_1$ and $L(\hat{\psi}_1)$ respectively.

The basis of the Likelihood Ratio test is the fact that under some regularity conditions (Silvey, 1975; Godfrey, 1988; Harvey, 1989), the statistic

$$\xi_{LR} = 2\left(\ln L(\hat{\psi}_1) - \ln L(\hat{\psi}_0)\right) \tag{3.5}$$

is asymptotically distributed as $\chi_m^2$ under $H_0$. The intuition behind the LR test is that whenever the maximised likelihood function under $H_0$, $L(\hat{\psi}_0)$ is much smaller than the unrestricted maximised likelihood $L(\hat{\psi}_1)$, there is evidence against the null hypothesis. The construction of the test statistic in the context of SV models is, however, computationally demanding since the model needs to be estimated twice, under the null and under the alternative. On the other hand, the test statistic does not require the knowledge of the information matrix and is based exclusively on the values of the likelihood function. Thus MCL estimation under the null and under the alternative is sufficient to construct $\xi_{LR}$.

In the proceeding discussion it was assumed that the restricted parameter vector, $\hat{\psi}_0$ was strictly in the interior of the parameter space. When one (or several) parameters are constrained to lie on the boundary of the parameter space the issue becomes more involved. In general, *if the order of differencing required to achieve stationarity is the*

*same* under the null and under the alternative, then the distribution of the test statistics under $H_0$ is given (Gouriéroux, Holly and Monfort, 1982) by a weighted sum of $\chi^2$ densities:

$$\xi_{LR} \sim \sum_{i=0}^{m} w(m,i)\chi_i^2$$

In particular, when $H_0$ involves a test of a single parameter lying on the boundary, $\xi_{LR}$ is known to be distributed as:

$$\xi_{LR} \sim \tfrac{1}{2}\chi_0^2 + \tfrac{1}{2}\chi_1^2 \tag{3.6}$$

The density in (3.6) has a concentration at the origin since $\chi_0^2$ is a degenerate distribution with all its mass at zero.

However, when $\phi=1$ the process is I(1) under the null and I(0) under the alternative and the standard theory does not apply. The distribution of the test statistic is unknown even in the case of Gaussian measurement equation noise. Analysis, of the kind performed in e.g. Dickey and Fuller (1981), is hindered here by the fact that closed form expressions for the test statistic in terms of sums and partial sums of observations and disturbances are not available. In addition, the distribution of the unrestricted ML estimates $\hat{\psi}_1$ under the null hypothesis is difficult to derive. Nevertheless, progress can be achieved by Monte Carlo experimentation. The density is simulated by calculating $\xi_{LR}$ from random samples of the basic SV model drawn under the null hypothesis. The resulting critical values are tabulated for several nominal sizes.[6]

To achieve compatibility with the Dickey-Fuller tests examined earlier, the experimental design of the previous Section is retained. The choice of $\sigma_\eta$ and $T$ is motivated by our incentive to determine

*(i)* whether or not the asymptotic distribution of $\xi_{LR}$ depends on $\sigma_\eta$,

*(ii)* how accurately can it be approximated in finite samples, and

*(iii)* the extent to which it differs from the weighted chi-squared density (3.6).

Clearly, MCL estimation will be most accurate in case C4 and least precise in C1 since the strength of the stochastic volatility process is increasing in $\sigma_\eta$, and the likelihood function is always better specified when larger sample sizes are available. However, it is interesting to see whether this will be reflected in the finite sample distribution of the likelihood ratio test statistic. Differences in sample densities of $\xi_{LR}$ for cases C3 vs. C4 and C1 vs. C2 will shed light on the first research objective, *(i)*. By contrast, the comparison of C2/C4 and C1/C3 will provide indicative evidence regarding the question *(ii)* above. And finally, difference to the density (3.6) will be highlighted by the size distortions of all four cases, thus addressing *(iii)*.

Estimation was performed in the following manner. Since $\sigma_\eta$ is the only hyperparameter ($\bar{\sigma}$ enters the state vector as described in Appendix 2), under the null hypothesis a one-dimensional grid search provided the starting values. Next, parameter estimates were obtained by the MCL method and the value of the restricted likelihood function, $L(\hat{\psi}_0)$ stored. The $\hat{\sigma}_\eta^2$ obtained under $H_0$ provides a useful starting value for the estimation under the alternative. The starting value for the autoregressive parameter was set to a value close (but not equal) to unity (0.981) since under the alternative, $\phi$ is constrained to be less than unity. The model was then re-estimated by MCL giving $L(\hat{\psi}_1)$.

---

Attention needs to be paid to the initial conditions with which the Kalman recursions are initialised. When the autoregressive parameter is constrained to be less than unity in absolute value the starting values for the Kalman filter may be taken as the mean and covariance matrix of the unconditional distribution of the state vector. The likelihood function will then be based on the prediction error decomposition with the summation starting at $t=2$ (Appendix 2, 3). However, when $\phi=1$ the unconditional variance of the state is not defined. Usage of the diffuse prior, $P_0=k\mathbf{I}_m$, where $k$ is a large number ($10^5$) and $\mathbf{I}_m$ is an ($m \times m$) identity matrix is problematic since in this case *(a)* the likelihood will have to be based on a summation starting at $t=3$, and, more importantly, *(b)* the constant level is not identified. In order to overcome this problem we initialise the Kalman filter for the non-stationary model with the diagonal element of $P_0$ corresponding to $h_t$ equal to zero. This implies that this element of the state vector is fixed. The likelihood for both models is therefore based on prediction errors starting at $t=2$.

Summary statistics of the sampling densities of model parameters were obtained as a by-product of the experiment and are reported in **Table 3.3**, Panel A. First, estimation of the scale parameter $\bar{\sigma}$ is equally accurate for all cases, C1-C4 and under both hypothesis. As expected, $\sigma_\eta$ is more precisely estimated when larger sample sizes are available. This is illustrated by smaller sample standard deviations for the cases C1 vs. C3 and C2 vs. C4. Second, the posterior density of $\hat{\phi}$ under the alternative of $\phi<1$, has a sample mean very close to unity (0.994-0.997) and a tiny sample standard deviation (0.008-0.003). This is encouraging and provides additional evidence for the efficiency of the MCL estimation technique discussed in Chapter 2.

Now turn to the main object of the investigation, the estimation of the density of the likelihood ratio test statistic. First, the following regularity was encountered: many

values of $\xi_{LR}$ were negative. This implied that the value of the maximised likelihood function was lower under the alternative than under the null hypothesis. However, examination of a scatter plot of $\hat{\phi}$ (vertical axis) against $\xi_{LR}$ (horizontal axis) e.g. for the case C3 presented in **Figure 3.2** resolves the contradiction.
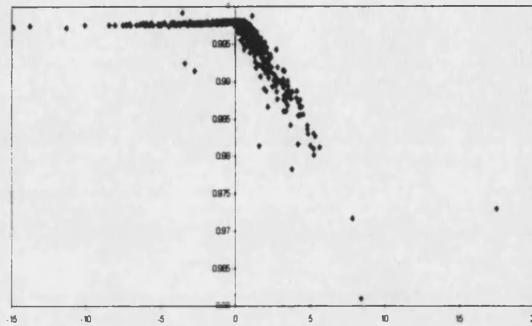


**Figure 3.2:** Estimates of $\phi$ vs. values of the $LR$ statistic

The graph documents that values of $L(\hat{\psi}_1)$ smaller than $L(\hat{\psi}_0)$ - leading to negative $\xi_{LR}$ - are likely to arise when the estimate of $\phi$ is very close to unity. Of course, in these circumstances the estimate of $\phi$ tends to unity, leading to the equality of the function values (and $\xi_{LR}=0$). However, under the stationary alternative, $H_1:\phi<1$, the estimation is performed in such as way that the autoregressive coefficient cannot become unity (Appendix 3). Therefore all negative values of $\xi_{LR}$ are henceforth set to zero. The resulting mass at the origin is reported in the final column of **Table 3.3**, Panel B.

Given a set of sample values $\{\xi^k_{LR}\}_j$, $k=1, .., 1,000, j=1, .., 4$ for each of the four cases, the densities are estimated non-parametrically, details of which can be found in Appendix 1. One density estimate, $\hat{f}_{\xi_{LR}}(x)$ for the case C2 - deliberately undersmoothed, and excluding the mass at the origin is presented in **Figure 3.3**. For comparison the graph of the $\chi_1^2$ density, $f_{\chi_1^2}(x)$ is also drawn.
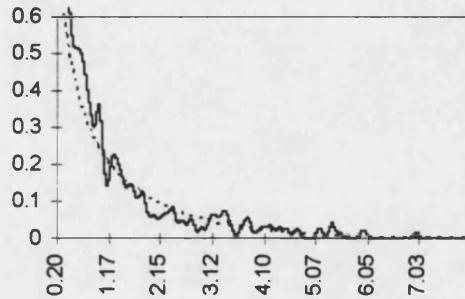
**Figure 3.3:** Estimated density of the $LR$ statistic and the $0.5\chi_1^2$ approximation

It is evident that the estimated density of the likelihood ratio test statistic (solid line) resembles the continuous part of the weighted $\chi^2$ density (3.6), (dotted line). Roughly half the mass is concentrated at the origin and the remainder of the density is highly skewed agreeing closely with $f_{\chi_1^2}(x)$.

This assertion is further supported by the results reported in **Table 3.4**, Panels B and C. The rejection frequencies - where the critical values were taken from (3.6) - are given in Panel B. The empirical sizes are very close to the nominal values, except, perhaps in the extreme right hand tail of the distribution. This may be explained by the fact the tail of the distribution may be difficult to approximate with $K=1,000$ draws. In general the $\xi_{LR}$ test based on the critical values obtained under (3.6) appears to be slightly undersized, i.e. rejecting less often than it should, leading to a *conservative* test statistic.

Conversely, Panel C presents the critical values obtained by numerical integration. Since the density estimates depend weakly on the bandwidth parameter, $h$, we present summary statistics across a range of smoothing parameters, $h_i$, $i=1, .., 10$ equally spaced between 0.05 and 0.5. The critical values thus obtained allow several conclusions to be drawn. First, when the number of observations in the series is large, $T=1,000$, the difference in estimated cut-off points across values of $\sigma_\eta$ is comparably small (comparing C4 vs. C3). For instance, the absolute difference in average cut-off

values across all sizes ranges between 0.01 and 0.20, as rows 11 and 8 illustrate. This evidence is suggestive of the claim that the asymptotic distribution of the likelihood ratio test does not depend on $\sigma_\eta$. By contrast, for smaller sample sizes, $T=500$ - comparing the entries in rows five (C2) and two (C1) - the absolute difference in the cut-off values across $\sigma_\eta$ is larger, ranging between 0.04 and 0.70. Not surprisingly, the difference becomes larger as the size decreases. This is explained by the fact that the estimate of the extreme right tail of the distribution is based on progressively few observations. Sample lengths of $T=500$ observations are not sufficient to achieve accuracy in the tail of the distribution as the comparison of average cut-off values across the cases with the same $\sigma_\eta$ but varying $T$ illustrates. For instance, the difference between rows 8 and 2 ranges between 0.35 and 0.75.

And finally, the simulated critical values do not differ substantially from those obtained under the mixture of $\chi^2$ densities (3.6). Slightly less weight is concentrated in the extreme tail of the distribution leading to lower cut-off values for the simulated distribution for any nominal size less than 0.05. Summing up, the distribution of the likelihood ratio test statistic is resembles very closely the weighted $\chi^2$ density: roughly half the mass is concentrated at the origin and the continuous part of the distribution is similar. Critical values of the weighted $\chi^2$ density lead to a conservative test statistic.

### 3.2.3. Power of the tests

The second stage of the Monte Carlo experiment compares the power of the test statistics $ADF_l$ in (3.3a) and $\xi_{LR}$ in (3.5). The process is now simulated under the alternative hypothesis with the autoregressive coefficient strictly less than unity. Six parameter triplets $\psi_i=(\phi,\sigma_\eta,\bar{\sigma})_i$ are constructed such that for each of the two values of $\sigma_\eta$ (0.3, 0.1) there are three corresponding values of the autoregressive coefficient $\phi$

(0.90, 0.95, 0.98). The long run volatility level, $\bar{\sigma}$ is calibrated each time so as to yield 23% annualised volatility. The resulting parameter values and the corresponding values of the coefficient of variation, $CV$ are recorded in the first three rows of **Table 3.4**. To reduce the computational burden the sample length was fixed at $T=500$ but the rejection frequencies are based on $K=1,000$ replications. While the critical values of the $ADF_l$ test were taken from Fuller (1976), the cut-off points for the $\xi_{LR}$ test were taken from (3.5), see **Table 3.3**, Panel C. This implies that the $\xi_{LR}$ test is a conservative test.

As expected, the power of the Dickey-Fuller test is low. More importantly, the power decreases as the autoregressive parameter approaches unity. For instance, for the nominal size of 0.05 and $\sigma_\eta=0.1$ the power is 0.559 for $\phi=0.9$ and only 0.245 for $\phi=0.98$. Second, the power is not a monotone function of the coefficient of variation, $CV$ which implies that the mean reversion - captured by $\phi$ - is the pivotal quantity and not the strength of the stochastic volatility process per se. The power increases slightly with $\sigma_\eta$ for a given $\phi$ which is not surprising given that the discussion in Section 3.2.1. And finally, the power decreases with the nominal size which may indicate that the tail behaviour is approximated less accurately with finite number of draws, $K$.

Turning now to the power of the likelihood ratio test statistic $\xi_{LR}$ we observe the following regularities. The power of the likelihood ratio test is higher than the power of $ADF_l$ test at each nominal size level; except in the last two cells, ($\sigma_\eta=0.1$, $\phi=0.95$) and ($\sigma_\eta=0.1$, $\phi=0.90$) which will be discussed shortly. The behaviour is similar, in as much as the power increases as the true value of the autoregressive coefficient decreases. For instance, for $\sigma_\eta=0.3$ at the 10% level the power of the $\xi_{LR}$ is 0.987 when $\phi=0.90$ and 0.765 when $\phi=0.98$, despite the $CV$ taking a relatively high value of 8.71. The results of this Section can be best summarised by a graph, comparing the

relative powers of the two tests. In **Figure 3.4**. the power is plotted against the

coefficient of variation, $CV$ for the nominal size of 10%.
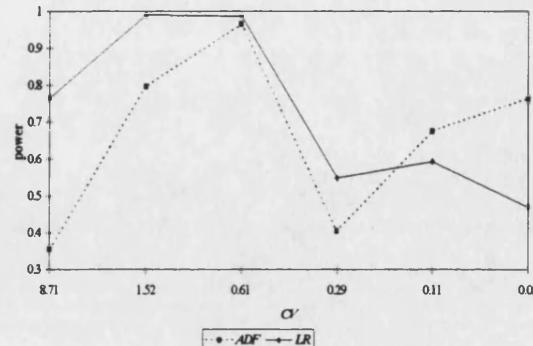


**Figure 3.4:** Power of $\xi_{LR}$ and $ADF_1$ at 10% nominal size.

It is interesting to observe that the strict dominance over the Dickey-Fuller test is not

present in the two final cells. Here, however, the $CV$ values indicate that the strength

of the stochastic volatility process is almost negligible ($CV$=0.11, $CV$=0.05), making

the likelihood "flat" over the relevant region of the parameter space. This behaviour is

illustrated in **Figure 3.5**.



*(a)*



*(b)*
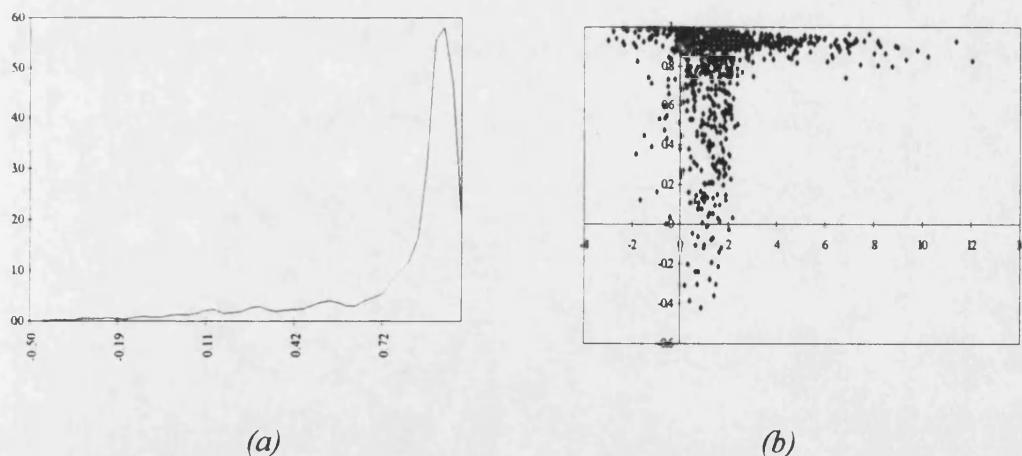
**Figure 3.5:** Sampling distribution of $\phi$ and the scatter plot of $\hat{\phi}$ vs. $\xi_{LR}$.

The sampling distribution of $\phi$, presented in **Figure 3.5***(a)*, exhibits a long tail

corresponding to a non-negligible mass of $\hat{\phi}$ away from the true value of 0.9. **Figure**

**3.5***(b)* presents a scatter plot of $\hat{\phi}$ (vertical axis) vs. $\xi_{LR}$ (horizontal axis). The graph

demonstrates three key features. First, for $\hat{\phi}$ close to unity $\xi_{LR}$ may become negative; this phenomenon has been discussed earlier. Second, when the likelihood is well specified, $\hat{\phi}$ is close to the true value of 0.9 and one obtains the intuitive behaviour: $\xi_{LR}$ increases, as $\hat{\phi}$ falls. However, when the likelihood is "flat" $\hat{\phi}$ is away from the true value and the likelihood function values under the null and under the alternative do not differ much, leading to low values of $\xi_{LR}$.

Summing up, it has been demonstrated that testing for a unit root in the (log)variance process against a stationary alternative by means of augmented Dickey-Fuller tests is not a reliable procedure. The tests are ether oversized (when the lag of the autoregressive polynomial in (3.2) is chosen too small) or have low power (when the lag is chosen so as to approximate the correct size). Moreover, the power declines as the true value of the autoregressive coefficient approaches unity, which is arguably the most interesting case from the point of view of applied empirical analysis. Furthermore, it is shown that the likelihood ratio test based upon the estimation of the SV model by MCL is more powerful. The distribution of the likelihood ratio test statistic is unknown but the Monte Carlo experiment presented here suggests that it can be well approximated by the weighted $\chi^2$ density, critical values of which are readily available.

## 3.3. Testing for higher order dynamics

The likelihood ratio testing principle can be also be applied in other situations. The basic SV model (2.1) can be extended in a number of directions, some of which were indicated in Section 1.4.2. A further generalisation might allow $h_t$ to follow a more complicated ARMA process, for instance an AR(2) process:

$$\begin{cases} r_t = \bar{\sigma} e^{0.5h_t} v_t & v_t = N(0,1) \\ h_t = \phi_1 h_{t-1} + \phi_2 h_{t-2} + \eta_t & \eta_t \sim N\left(0, \sigma_\eta^2\right) \end{cases} \tag{3.7}$$

Model (3.7) can be easily represented in the state space form, as can be seen in Appendix 2. Similar models were briefly discussed by Shephard (1996) and Kim, Shephard and Chib (1996). GARCH counterparts were proposed by e.g. Engle and Lee (1992).

The hypothesis of interest is the validity of the restriction $\phi_2 = 0$. Thus under the null, (log)variance is a stationary AR(1) process, $H_0: \phi_2 = 0, \phi_1 < 1$ while under the alternative, it is stationary AR(2), $H_1: \phi_2 \neq 0, |\lambda_i| < 1$ where

$$\lambda_{1/2} = \frac{1}{2}\left(\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}\right) \tag{3.8}$$

are the roots of the autoregressive polynomial, $1 - \phi_1 L - \phi_2 L^2 = 0$. Again, estimation by MCL allows the likelihood ratio test, $\xi_{LR}$ to be easily constructed. Moreover, the asymptotic distribution of the test is $\chi_1^2$ since the restricted parameter is not on the boundary of the parameter space, $\phi_2 \in (-1, 1)$.

In order to investigate the finite sample properties, and shed more light upon the estimation of SV models with higher order dynamics, a small scale Monte Carlo experiment was conducted. The process was generated under the null of stationary AR(1) with the parameter values[7] $(\phi, \sigma_\eta, \bar{\sigma}) = (0.95, 0.26, 0.025)$. The length of the time series was set to $T_1 = 500$ and $T_2 = 1,000$ while the number of draws was $K = 1,000$.

Since the process is stationary under the null and under the alternative, a proper stationary distribution exists and can be used to initialise the Kalman recursions

---

[7] This is $\psi_5$ of Chapter 2.

(Appendix 2). Furthermore, parameter estimates under the null provide good starting values for the estimation under the alternative, while $\phi_2$ is initialised at zero.

The results are reported in **Table 3.5**. Summary statistics of the posterior distribution of parameter estimates are reported in Panel A while the rejection frequencies of the likelihood ratio test are to be found in Panel B. First, parameter estimation under the true null is more accurate when the sample size is large. The sampling standard deviation on the estimates of $\phi_1$ and $\sigma_\eta$ decreases as $T$ grows from 500 to 1,000. This is not surprising since the likelihood becomes more informative as $T$ increases. More interesting are the results of the estimation under the alternative: parameter estimates appear to be on average *nowhere near the true values*. For instance the mean on $\phi_1$ is 1.478 and on $\phi_2$ is -0.502 for $T$=1,000. More insightful is the posterior distribution of the roots of the autoregressive polynomial, $\lambda_1$ and $\lambda_2$ which is presented in **Figure 3.6**.



$\lambda_1$                                                        $\lambda_2$
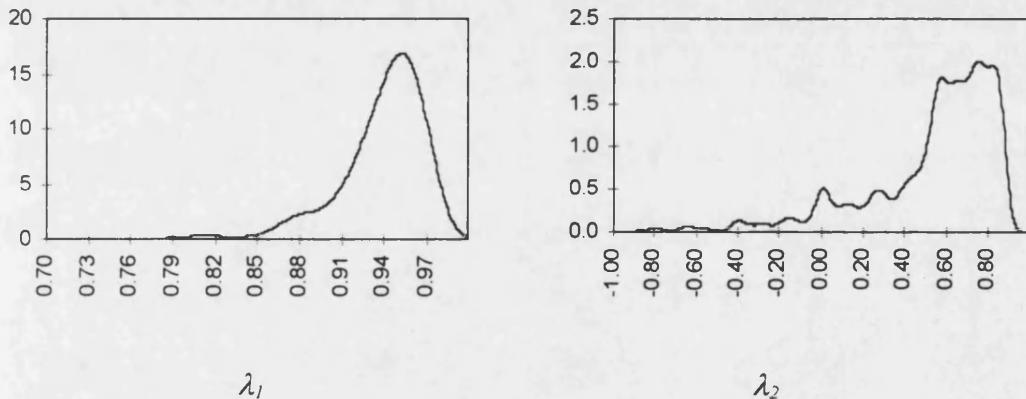
**Figure 3.6:** Estimation of the AR(2) process: sampling densities of the roots $\lambda_i$.

As expected, the density of the first root, $\lambda_1$ is centred around the true value of 0.95. By contrast, the posterior density of $\lambda_2$, is at best bi-modal, with some mass concentrated at the origin, and the remainder being close to unity. This implies that the information is often not sufficient to allow for a distinction between the two roots; they are estimated as $\lambda_1 \approx \lambda_2$. The autoregressive parameters are related to the roots via

$$\phi_1 = \lambda_1 + \lambda_2 \qquad\qquad \phi_2 = -\lambda_1\lambda_2 \qquad\qquad (3.9)$$

In consequence, roughly equal roots will induce estimates of $\phi_1$ to be around 1.8 and $\phi_2$ around -0.8, which is well documented in the summary statistics of **Table 3.5**, Panel A. The mean of the posterior density of $\phi_2$ is -0.502 which is between 0 and -0.8. Similarly the mean of $\phi_1$ (1.4) is between 0.9 and 1.8.

Additional evidence in support of this interpretation is given in **Figure 3.7**(a) where a scatter plot of $\hat{\phi}_2$ (horizontal axis) vs. $\hat{\phi}_1^2 + 4\hat{\phi}_2$ (vertical axis) [which - by virtue of (3.8) - is a measure of closeness of the roots of the autoregressive polynomial] is presented. It illustrates that negative values of $\phi_1$ are likely to occur when the diference in the roots is small (i.e. a single root is found).
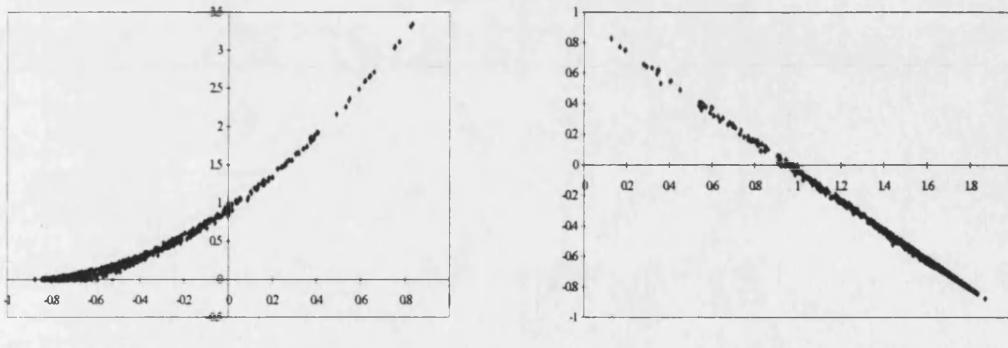


(a): $\hat{\phi}_2$ vs. $\hat{\phi}_1^2 + 4\hat{\phi}_2$            (b): $\hat{\phi}_1$ vs. $\hat{\phi}_2$

**Figure 3.7:** Aspects of the posterior density of $\phi_1$ and $\phi_2$ in AR(2).

On the other hand, estimates of the two coeffcients are strongly related. **Figure 3.7**(b) illustrates that the relationship between $\hat{\phi}_1$ (horizontal axis) and $\hat{\phi}_2$ (vertical axis) is linear. The likelihood will therefore be "flat" along the line sketched in **Figure 3.7**(b). Finally, note that while the posterior density of $\sigma_\eta$ is centred around the true value under the null, it is shifted towards the origin under the alternative. This can be explained by noting that whenever higher order dynamics are present, lower values of the signal-noise ratio are sufficient to produce an acceptable fit to the data.

The resulting distribution of the likelihood ratio test is summarised in **Table 3.5**, Panel B. Here the rejection frequencies of the test are reported for several nominal sizes. The critical values are taken from the $\chi_1^2$ distribution. Two conclusions can be drawn. First, small sample sizes ($T$=500) are not sufficient to approximate the distribution of the test statistic. The empirical sizes are far away from the nominal levels. Second, for $T$=1,000 the approximation is much better with the empirical sizes being reasonably close to the nominal levels. More importantly, the size distortions are not uniform across the nominal sizes which indicates that the difference may well be due to sampling variation.

Summing up, estimation of a higher order autoregressive process in (log)variance is feasible alas the procedure for constraining the autoregressive coefficients adopted here leads to parameter estimates which need to be interpreted with caution. The likelihood ratio test of stationary AR(1) dynamics can be constructed but appears to be valid for large sample sizes only.

## 3.4. Conclusion

This Chapter considers some aspects of inference within the SV model. Confidence intervals on model hyperparameters are briefly discussed. At the centre is the discussion of the test of the unit root in the (log)variance process. Two methods are juxtaposed: the augmented Dickey-Fuller test and the likelihood ratio test. It is shown that large number of lags are required in the augmented Dickey-Fuller test so that the distribution of the test statistic is approximated reasonably close. This leads to the loss of power. By contrast, the distribution of the likelihood ratio test is unknown but is shown to resemble closely the $0.5\chi_1^2$ density. The likelihood test is shown to be more powerful than the augmented Dickey-Fuller test except in the unrealistic case where

the strength of the stochastic volatility process is so low as to make the likelihood surface ill-conditioned.

As we consider only particular data generating processes - and the distribution of the tests may depend upon the particular parameter values chosen - the results may only be illustrative. Nevertheless, the results are informative and conclusive within the bounds of Monte Carlo experimentation. This translates into a recommendation for applied empirical research: whenever the value of the likelihood ratio test statistic is found to be in the region of 2-4 the $p$-value can be simulated by means of Monte Carlo methods similar to the ones adopted here.

**Table 3.1. Empirical estimates of SV model parameters**

| Study | Time series | $\hat{\phi}$ | $\hat{\sigma}_\eta$ | $CV$ | $q$ |
|---|---|---|---|---|---|
| Melino&Turnbull (1990) | CAD/USD, $T$=3,011 | 0.91 | 0.14 | 0.14 | 0.005 |
| JPR* | portfolios, $T$=1,540 | 0.93 | 0.32 | 1.07 | 0.02 |
| | stocks, $T$=1,540 | 0.89 | 0.36 | 0.84 | 0.03 |
| | exchange rates, $T$=2614 | 0.95 | 0.20 | 0.77 | 0.01 |
| Danielsson (1994a) | S&P500, $T$=2,202 | 0.96 | 0.17 | 0.37 | 0.005 |
| HRS* | exchange rates, $T$=946 | 0.97 | 0.14 | 0.58 | 0.004 |
| Taylor (1994) | DEM/USD, $T$=3,283 | 0.94 | 0.22 | 0.50 | 0.01 |

This table reproduces the parameter estimates of the basic SV model (1) reported in several empirical application of the SV model. For reasons of brevity only average parameter values for Jacquier, Polson and Rossi (1994) and Harvey, Ruiz and Shephard (1994) are reported. The entries are denoted by JPR* and HRS* respectively. The coefficient of variation,

$$CV = \mathrm{var}(\sigma_t)E[\sigma_t]^{-2} = \exp\left(\sigma_\eta^2\left(1-\phi^2\right)^{-1}\right) - 1$$

is conventionally used to describe the strength of the volatility process (Jacquier, Polson and Rossi, 1994) and is reported in the penultimate column. The final column reports the empirical estimates of the signal-noise ratio, $q = 2\sigma_\eta^2/\pi^2$ .

**Table 3.2. Distribution of the Dickey-Fuller tests**

**Panel A.** Size of the $ADF_1$ test

| Size | 0.05 | | | | | 0.01 | | | | |
|------|------|------|------|------|------|------|------|------|------|------|
| p | 10 | 20 | 30 | 40 | 50 | 10 | 20 | 30 | 40 | 50 |
| C1 | 0.909 | 0.453 | 0.196 | 0.103 | 0.070 | 0.786 | 0.233 | 0.069 | 0.028 | 0.014 |
| C3 | 0.246 | 0.075 | 0.051 | 0.044 | 0.039 | 0.104 | 0.018 | 0.010 | 0.009 | 0.007 |
| C2 | 0.938 | 0.544 | 0.275 | 0.145 | 0.092 | 0.851 | 0.346 | 0.117 | 0.044 | 0.024 |
| C4 | 0.267 | 0.080 | 0.056 | 0.048 | 0.044 | 0.126 | 0.020 | 0.011 | 0.009 | 0.008 |

**Panel B.** Size of the $ADF_2$ test

| Size | 0.05 | | | | | 0.01 | | | | |
|------|------|------|------|------|------|------|------|------|------|------|
| p | 10 | 20 | 30 | 40 | 50 | 10 | 20 | 30 | 40 | 50 |
| C1 | 0.961 | 0.684 | 0.466 | 0.358 | 0.301 | 0.903 | 0.528 | 0.327 | 0.246 | 0.205 |
| C3 | 0.371 | 0.175 | 0.163 | 0.184 | 0.207 | 0.217 | 0.088 | 0.088 | 0.109 | 0.131 |
| C2 | 0.968 | 0.685 | 0.443 | 0.309 | 0.246 | 0.911 | 0.521 | 0.285 | 0.182 | 0.145 |
| C4 | 0.355 | 0.139 | 0.112 | 0.115 | 0.132 | 0.196 | 0.051 | 0.042 | 0.050 | 0.062 |

This table reports the rejection frequencies of the augmented Dickey-Fuller tests (3.3) for two nominal sizes (0.05 and 0.01). $K$=10,000 samples of the basic SV model (2.1) are drawn for each case, C1-C4, and the test statistics are computed based on $p$ lags in the autoregression (3.2).

**Table 3.3. Distribution of the likelihood ratio test against $H_0: \phi=1$**

**Panel A:** Distribution of the MCL parameter estimates

| | | $\sigma_\eta=0.1$ | $\phi=1$ | $\bar{\sigma}$ | $\sigma_\eta=0.3$ | $\phi=1$ | $\bar{\sigma}$ |
|---|---|---|---|---|---|---|---|
| **T=500** | **Under $H_0$** | 0.093 | - | 0.014 | 0.277 | - | 0.014 |
| | | (0.027) | - | (0.008) | (0.038) | - | (0.005) |
| | **Under $H_1$** | 0.103 | 0.994 | 0.014 | 0.286 | 0.995 | 0.015 |
| | | (0.029) | (0.008) | (0.003) | (0.038) | (0.005) | (0.006) |
| **T=1000** | **Under $H_0$** | 0.095 | - | 0.014 | 0.278 | - | 0.014 |
| | | (0.016) | - | (0.003) | (0.027) | - | (0.005) |
| | **Under $H_1$** | 0.104 | 0.997 | 0.014 | 0.286 | 0.997 | 0.015 |
| | | (0.018) | (0.003) | (0.003) | (0.036) | (0.003) | (0.006) |

Panel A reports sample means and standard deviations (in parentheses) of the posterior densities of the parameters. The basic SV model (2.1) is simulated $K=1,000$ times and estimated twice, under the null hypothesis, $H_0: \phi=1$, and under the alternative, $H_1: \phi<1$.

**Panel B:** Size of the likelihood ratio test

| *Size* | **0.1** | **0.05** | **0.025** | **0.01** |
|---|---|---|---|---|
| *critical value* | [1.642] | [2.706] | [3.841] | [5.412] |
| *C1* | 0.085 | 0.033 | 0.016 | 0.003 |
| *C2* | 0.087 | 0.048 | 0.021 | 0.004 |
| *C3* | 0.125 | 0.059 | 0.022 | 0.003 |
| *C4* | 0.110 | 0.051 | 0.034 | 0.019 |

Panel B reports the rejection frequencies of the likelihood ratio test, $\xi_{LR}$. The nominal size and the critical values of the weighted $\chi^2$ density (3.6) are reported in the first two rows.

**Panel C:** Critical values of the likelihood ratio test statistic

| | | Size | 0.1 | 0.05 | 0.025 | 0.01 | *mass at 0* |
|---|---|---|---|---|---|---|---|
| | 1 | *Min* | 1.443 | 2.193 | 2.963 | 4.402 | |
| *C1* | 2 | *Avg* | 1.500 | 2.263 | 3.030 | 4.421 | 0.448 |
| | 3 | *Max* | 1.590 | 2.339 | 3.129 | 4.433 | |
| | 4 | *Min* | 1.486 | 2.602 | 3.701 | 4.716 | |
| *C2* | 5 | *Avg* | 1.545 | 2.641 | 3.726 | 4.790 | 0.461 |
| | 6 | *Max* | 1.644 | 2.688 | 3.763 | 4.860 | |
| | 7 | *Min* | 1.855 | 2.930 | 3.717 | 4.745 | |
| *C3* | 8 | *Avg* | 1.914 | 2.946 | 3.777 | 4.770 | 0.500 |
| | 9 | *Max* | 1.990 | 2.965 | 3.851 | 4.804 | |
| | 10 | *Min* | 1.689 | 2.431 | 3.504 | 4.759 | |
| *C4* | 11 | *Avg* | 1.712 | 2.483 | 3.534 | 4.778 | 0.518 |
| | 12 | *Max* | 1.769 | 2.567 | 3.568 | 4.807 | |
| $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ | 13 | - | 1.642 | 2.706 | 3.841 | 5.412 | 0.5 |

Panel C tabulates the critical values of the $\xi_{LR}$ test. The nominal size is reported in the column header. Each density was estimated with 10 different bandwidth parameters, $h_i$, equally spaced between 0.05 and 0.5. Averages (Min/Max) across $i$ are reported in rows 1 to 12. For comparison, the final row reports the critical values of the weighted $\chi^2$ density (3.6) which are taken as the $2\alpha$ significance points of the $\chi^2_1$ density for the size $\alpha$. The final column gives the percentage of sample values at the origin.

**Table 3.4. Power of the Dickey-Fuller and the likelihood ratio tests**

| | $\sigma_\eta$ | 0.3 | | | 0.1 | | |
|---|---|---|---|---|---|---|---|
| | $\phi$ | 0.98 | 0.95 | 0.9 | 0.98 | 0.95 | 0.9 |
| | $\bar{\sigma}$ | 0.008 | 0.011 | 0.013 | 0.013 | 0.014 | 0.014 |
| | $CV$ | 8.71 | 1.52 | 0.61 | 0.29 | 0.11 | 0.05 |
| | [0.1] | 0.355 | 0.796 | 0.965 | 0.404 | 0.676 | 0.764 |
| $ADF_l$ | [0.05] | 0.213 | 0.614 | 0.896 | 0.245 | 0.457 | 0.559 |
| | [0.01] | 0.043 | 0.247 | 0.582 | 0.063 | 0.151 | 0.221 |
| | [0.1] | 0.765 | 0.989 | 0.987 | 0.549 | 0.594 | 0.470 |
| $\xi_{LR}$ | [0.05] | 0.506 | 0.931 | 0.966 | 0.335 | 0.372 | 0.199 |
| | [0.01] | 0.130 | 0.618 | 0.806 | 0.063 | 0.097 | 0.05 |

This table tabulates the rejection frequencies of the augmented Dickey-Fuller $ADF_l$ (3.3a) and the likelihood ratio test, $\xi_{LR}$ (3.5). The nominal size is given in parenthesis. The lag length, $p$ in the autoregression (3.2) is selected so that the size distortions are minimised. This requires $p=50$ for $\sigma_\eta=0.1$ and $p=30$ for $\sigma_\eta=0.3$. The basic SV model (2.1) is simulated with the values of the parameters reported in the first column $K=1,000$ times. The length of the time series is $T=500$.

## Table 3.5. Distribution of the likelihood ratio test against $H_0: \phi_2 = 0$

**Panel A.** Distribution of MCL parameter estimates

|  |  | $\phi_1$ | $\phi_2$ | $(\lambda_1)$ | $(\lambda_2)$ | $\sigma_\eta$ | $\bar{\sigma}$ |
|---|---|---|---|---|---|---|---|
|  | *True* | 0.95 | 0 | 0.95 | 0 | 0.26 | 0.025 |
| *T=500* | **Under $H_0$** | 0.947 | - | 0.947 | - | 0.249 | 0.025 |
|  |  | (0.030) | - | (0.030) | - | (0.059) | (0.003) |
|  | **Under $H_1$** | 1.398 | -0.426 | 0.934 | 0.464 | 0.132 | 0.025 |
|  |  | (0.358) | (0.345) | (0.047) | (0.373) | (0.075) | (0.003) |
| *T=1,000* | **Under $H_0$** | 0.948 | - | 0.948 | - | 0.252 | 0.024 |
|  |  | (0.020) | - | (0.020) | - | (0.034) | (0.002) |
|  | **Under $H_1$** | 1.478 | -0.502 | 0.939 | 0.539 | 0.117 | 0.025 |
|  |  | (0.308) | (0.297) | (0.031) | (0.320) | (0.064) | (0.002) |

Panel A reports sample means and standard deviations (in parentheses) of the posterior densities of the parameters. The basic SV model (2.1) is simulated $K=1,000$ times and estimated twice, under the null hypothesis, $H_0: \phi_2 = 0$, and under the alternative, $H_1: \phi_2 \neq 1$.

**Panel B.** Rejection frequencies of the likelihood ratio test

| *Size* | **0.1** | **0.05** | **0.025** | **0.01** |
|---|---|---|---|---|
| *critical value* | [2.706] | [3.841] | [5.024] | [6.635] |
| *T=500* | 0.053 | 0.013 | 0 | 0 |
| *T=1,000* | 0.173 | 0.071 | 0.02 | 0.002 |

Panel B tabulates the rejection frequencies of the $\xi_{LR}$ test. The nominal size and the corresponding critical value - taken from a $\chi_1^2$ distribution - are reported in the first two rows.

# Chapter 4: Implied volatility revisited

## 4.1. Introduction

GARCH and SV models attempt to explain the time series properties of second conditional moments of asset returns by taking into account the price history of the process. Often additional information is available from the implied volatility of options. Implied volatility is defined as that value of the diffusion coefficient, $\sigma_{impl}$ which equates the theoretical Black and Scholes (1973) price, BS thereafter, to the market price of the option. Given a price, the formula is inverted to give a value of implied volatility as a function of exercise, or strike price, $E$ and maturity, $\tau$.

$$\sigma_{impl}(\tau,E) = BS^{-1}(\tau,E) \qquad (4.1)$$

This can be represented in a matrix of implied volatilities sorted by $E$ and $\tau$. For a fixed maturity, the graph of the mapping $E \to \sigma_{impl}$ is typically U-shaped: implied volatilities 'smile'; but sometimes this smile is more or less lopsided, or 'skewed'. For a fixed strike the mapping $\tau \to \sigma_{impl}$ defines the term structure of implied volatility, or the 'volatility curve'. Finally, each individual element of the volatility matrix, $\sigma_{impl,t}(\tau,E)$ exhibits pronounced time series dynamics.

In principle, therefore, two sources of information are available: the time series of the underlying asset and the time series of the matrix of implied volatilities. One question which arises naturally in this context is that of the information content of implied volatilities. Can the information contained in the volatility matrix be employed to further the understanding of the true dynamics of the underlying instrument? Important implications for the construction of optimal volatility forecasting rules arise if the answer to this question is affirmative. The present Chapter addresses this research objective by adding the information contained in implied volatilities as a set of

explanatory variables to the basic SV model. It is shown that implied volatilities capture accurately the time series dynamics of the latent asset volatility in that the autoregressive component is often rendered redundant.

The second research objective is a joint test of market efficiency and unpriced volatility risk. Under these two assumptions, implied volatilities can be shown to be the options market's subjective forecast of the future volatility of the underlying asset even if the volatility is stochastic. In these circumstances the history of the price process should have no incremental explanatory power. By implication, volatility forecasts constructed from time series models should be less accurate than those obtained from implied volatilities. Our out-of-sample forecasting experiment indicates that while volatility is very hard to predict, forecasts obtained from the SV model are at least (and often *more*) accurate that those given by implied volatilities.

The present Chapter addresses these issues in the light of uniquely available over-the-counter data. Section 4.2 contains a review of related work and identifies the measurement problems associated with previous tests of the hypothesis. Section 4.3 introduces the data, and contains the empirical analysis of the volatility curve. The main results are discussed in Section 4.4, while Section 4.5 concludes.

## 4.2. Overview of previous work

One of the main assumptions of the BS model is that the underlying instrument follows a diffusion process with a constant volatility parameter. The matrix of implied volatilities constructed by equating the theoretical BS prices to the market prices of European call options[1] as in (4.1) should be flat across all maturities and strikes.

---

[1] A call option gives the holder the right to purchase one unit of the underlying asset at a predetermined exercise, or strike price, $E$. A European (American) option can be exercised only on (anytime prior to) the specified expiry, or maturity date. The time remaining to the expiry date is called the maturity of the option, $\tau$ (Hull, 1993).

However, one of the most striking and widely documented regularities of observed option prices is that the matrix of implied volatilities obtained by inverting the BS formula is neither flat nor constant over time (Latane and Rendelman, 1976; Schmalensee and Trippi, 1978; Rubinstein, 1985).

This empirical observation motivated three distinct research areas. Firstly, the 'smile' effect inspired the development of more sophisticated option pricing models.[2] In these models (Hull and White, 1987; Johnson and Shanno, 1987; Scott, 1987; Wiggins, 1987; Renault and Touzi, 1996) the volatility is itself a stochastic process following a diffusion:

$$\frac{dS_t}{S_t} = \mu(t, S_t, Y_t)dt + \sigma_t dW_t^1$$
$$d \log \sigma_t^2 = k(c - \log \sigma_t^2)dt + \sigma_\eta dW_t^2$$

(4.2)

where $S_t$ is the price of the underlying asset, and the instantaneous (log)variance follows a mean reverting process. Hull and White (1987) show that when the volatility risk is unpriced and the two Wiener processes are uncorrelated, the price of the option can be expressed as the expectation of the BS price taken with respect to the distribution of the sample paths of the volatility process.[3] By making the volatility parameter follow a continuous process it is possible to match the observed volatility skew (Hull and White, 1987; Chesney and Scott, 1989; Melino and Turnbull, 1990; Heynen, 1994). This research suggests the empirical validity of stochastic volatility option pricing models.

---

[2] But see also models of volatility switching (Naik, 1993), mixed jump diffusions (Merton, 1976; Amin, 1993), and state dependent volatility models (Derman and Kani, 1994; Dupire, 1994; Rubinstein, 1994) as well as evidence of empirical validity of such models (Dumas, Fleming and Whaley, 1995; Malz, 1996; Bates, 1996).

[3] Willard (1996) shows that this result is valid even if the volatility and spot processes are instantaneously correlated.

Second, the term structure of implied volatility has been tested for consistency with rational expectations hypothesis, according to which long term implied volatility is some average of expected future short term volatilities. Stein (1989) showed how mean reversion in volatility and rational expectations can be combined to impose testable restrictions on the evolution of the term structure of implied volatility. In particular, under rational expectations, the long term implied volatility should equal some weighted average of expected future short term implied volatilities. Using two (short and long maturity) daily time series of S&P100 index options he found evidence to the contrary, interpreting it as market 'overreaction'. Campa and Chang (1995) examined a richer dataset consisting of implied volatilities of six maturities and were unable to reject the expectation hypothesis.

Recognising that this is a test of a joint hypothesis of correct time series specification and a particular rational expectations model, Heynen, Kemma and Vorst (1994) re-examined the issue by considering alternative processes describing changes in asset price volatility. Comparison of an SV model, a GARCH, and an EGARCH specification showed that the joint hypothesis could be rejected in the first two cases but not in the latter. This research suggested that a major factor in the understanding of the dynamics of volatility term structure is the ability of the model to capture the time series dynamics and in particular, the mean level of volatility. Xu and Taylor (1994) proposed an alternative model of expectation generation in which expectations are assumed to revert some time dependent long run average level. Their model is well supported by the data and no evidence of 'overreaction' could be found. The authors document significant term structure dynamics with the slope of the volatility curve changing approximately once every two to three months. Thus, depending on the process which is assumed to govern the evolution of volatility and the model of

expectation generation one may conjecture that implied volatility may or may not be a poor predictor of future *implied* volatilities.

The final research avenue is the question of whether the information encoded in implied volatilities is helpful in forecasting realised, *ex post* variance of returns. It is here that this Chapter aims to make a contribution. The information content of implied volatilities has been investigated by several authors. Canina and Figlewski (1993) analyse daily data on the S&P100 stock index options for the period 1983-1986. The hypothesis is tested by means of regressions of the form:

$$\tilde{\sigma}_{t,t+L} = \alpha + \beta_1 \sigma_{impl,t}(\tau,E) + \beta_2 \hat{\sigma}_{t,t+L} + u_t \qquad (4.3)$$

where $\tilde{\sigma}_{t,t+L}$ is the realised return volatility over some future horizon $t,t+L$, $\sigma_{impl,t}(\tau,E)$ is the implied volatility at $t$ with maturity $\tau$, and strike $E$, and $\hat{\sigma}_{t,t+L}$ is the predictor obtained from the time series of the underlying asset using data up to the time $t$. In a joint test of market efficiency and optimality of the implied volatility as a forecast variable one expects $\alpha=0$, $\beta_1=1$, and $\beta_2=0$ (Fair and Shiller, 1990; Pagan and Schwert, 1990). Canina and Figlewski (1993) find that implied volatility is a poor predictor of future ex post volatility, and better forecasts can be obtained from simple historical moving averages.

Day and Lewis (1992) and Lamoureux and Lastrapes (1993) approach the problem by adding the time series of implied variance as an exogenous explanatory variable into the variance equation of a GARCH (and an EGARCH) model as in (1.7):

$$\begin{cases} r_t = \sigma_t v_t, \qquad v_t \sim N(0,1) \\ \sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma \sigma_{impl,t-1}^2(\tau,E) \end{cases} \qquad (4.4)$$

where, as before, $r_t$ is the mean adjusted return on the asset. The hypothesis, that contemporaneous prices of options contain information beyond of what can be

obtained from the historic price series of the underlying asset is tested by examining the significance of the coefficient $\gamma$. Both studies report that the hypothesis of no influence, $\gamma=0$ and the hypothesis of exclusive influence, $\alpha_i=\beta_i=0$ can be simultaneously rejected. This evidence led the authors to conclude that implied volatility has some informational content but by itself, it is not sufficient to account for all time series variation if the conditional variance.

However, the formulation (4.4) is unduly restrictive in that the coefficient on the regressor does not capture the instantaneous impact of implied volatility on unconditional variance. Instead, equation (1.8) shows that

$$Er_t^2 = \frac{\alpha_0}{1-\varphi} + \sum_{j=1}^{\infty} \gamma \, \varphi^{j-1} \, \sigma^2_{impl,t-1-j}(\tau,E), \qquad \varphi = \alpha_1 + \beta_1 \qquad (4.5)$$

Thus the unconditional variance of returns is given by an exponentially weighted average of past implied volatilities but not the current value, $\sigma^2_{impl,t}(\tau,E)$, arguably the most relevant of all observations.

By contrast, Xu and Taylor (1995) who examine four exchange rate series together with the corresponding implied volatilities taken from exchange traded contracts during the period 1985-1991 find the opposite result. The evolution of the conditional volatility in-sample can be best captured by the time series of $\sigma^2_{impl}$.

The comparison of the out-of-sample predictive power of forecasts obtained from time series models and implied volatilities has been examined by Day and Lewis (1992), Lamoureux and Lastrapes (1993), Xu and Taylor (1995) and Jorion (1995). The results of the first two papers suggest that the time series models, which only utilise the information contained in the price history of the process, perform better than implied volatilities. By contrast, Xu and Taylor (1995) and Jorion (1995) find the opposite result: the forecasts constructed from GARCH models are shown to be inferior to the

implied volatility figures. Jorion (1995) suggests that measurement problems may be responsible for the discrepancy. Stale, non-synchronous quotes for the underlying stock market index, arbitrage restrictions due to transaction costs, and bid/ask distortions in the recorded option prices may contribute to measurement errors, particularly in stock index implied volatilities. Further distortions are induced by attempting to invert American options on dividend paying stocks in order to obtain values of implied volatilities. For instance, Canina and Figlewski (1993) report several cases where implied volatility calculated by their method is negative. The empirical analysis presented in this Chapter demonstrates that even in the case of foreign currency options - which minimise the measurement problems mentioned above - better forecasts may be constructed from the history of the return process.

An altogether different approach to testing the informational content of implied volatilities is advocated by Noh, Engle and Kane (1994). Here option trading strategies are devised based upon the predictions derived from time series models of the underlying (GARCH) and compared to the predictions obtained solely from the history of the implied volatilities. In an application to S&P500 stock index options for the period 1985-1992 the authors find that the cumulative return generated by the trading strategy based on GARCH forecasts is much higher than when the rules are derived form implied volatilities. The results, however, are unconvincing; for example, Exhibit 9, pg. 27 reveals that the superiority of the GARCH based trading rule is driven by a single observation: during the Crash of 1987 the GARCH portfolio generates an abnormal profit, while the implied volatility portfolio generates an abnormal loss. Moreover, prior to this event the cumulative return grows at a faster rate for the implied volatility based strategy than for the GARCH portfolio. This highlights some important limitations of the methodology. Apart from measurement problems discussed earlier, sensitivity to outliers, execution risk, various assumptions regarding

transaction costs, opportunity cost of capital etc., and subjectivity of the trading rules contribute to the shortcomings of this method.

One difficulty associated with all previous studies is the unavailability of "time homogenous" implied volatility data. A time series of implied volatility, $\sigma_{impl,t}(\tau,E)$, $t=1, .., T$ is called time homogenous if the maturity, $\tau$ and the moneyness, $E/F_t$ of the option are set relative to the observation date, $t$ and are thus constant throughout the sample. Exchange traded options do not allow for a construction of a time homogenous implied volatility data-sets because such options have a fixed *calendar* expiration date and a fixed *nominal* strike.

For instance, in order to construct a univariate time series of daily implied volatility from exchange traded contracts, at each observation date an option with some strike and some maturity has to be selected. On the following date this particular strike will still be available but because the underlying will have moved the moneyness of the option will now be different. Similarly, the calendar expiry is fixed so that the maturity of the option is now shorter by one period. In the presence of pronounced term structure dynamics of the kind documented in this Chapter, and elsewhere (Xu and Taylor, 1994, 1995) this data construction procedure will induce maturity mismatch which may affect inference.

When the out-of-sample forecasting experiments are performed, the horizon of the time series model may be synchronised with the maturity of the option (Lamoureux and Lastrapes, 1993) at each observation date. As a consequence, however, the quality of the forecast from the time series model will vary throughout the sample because the predictive ability of time series models depends crucially on the length of the forecasting horizon. The $l$-step ahead forecast function of the GARCH(1,1) model (4.5) with no explanatory variables ($\gamma=0$) is:

$$\hat{\sigma}_{t,t+l}^2 = \overline{\omega} + \varphi^{l-1}\left(\sigma_t^2 - \varpi\right), \qquad \overline{\omega} = \alpha_0\left(1-\varphi\right)^{-1}$$

For large lead periods $l$, the forecast will tend to the long run mean, $\overline{\omega}$ so that the analysis is reduced to the comparison of the predictive power of implied volatility to that of the long run variance. This explains why Day and Lewis (1992) and Lamoureux and Lastrapes (1993) (who study lead periods between $l$=64 and $l$=129 trading days) find that a naive forecast, the sample variance, performs at least as well as the GARCH/EGARCH forecast.

Finally, tests based on (4.3) are invalidated by econometric difficulties. First, the forecasts from time series models and implied volatilities have typically a correlation coefficient above 0.85 potentially leading to the multicollinearity problem. More importantly, the volatility persistence parameter is typically well above 0.9 leading to highly autocorrelated volatility forecasts. The strong autocorrelation in the dependent and the exogenous variables will induce small sample biases in the coefficient estimates of (4.3). This is a well known problem in the context tests for rationality (Mankiw and Shapiro, 1986).

Summing up, the evidence reported in previous studies suggests that ATM implied volatilities are rich in time series and term structure dynamics and have some informational content as to the evolution of the conditional variance of the process. However, the difficulties associated with the data, the measurement techniques, and the experimental set-up render the results open to discussion.

## 4.3. Empirical study

### 4.3.1. Data description

One of the major difficulties of the empirical research in this area has been the unavailability of a contemporaneously recorded prices of the underlying and a time homogenous implied volatility matrix. Our data set comprises daily bid and ask quotes for over-the-counter (OTC) foreign exchange options on the DEM/USD exchange rate[4] together with the corresponding, contemporaneously recorded quotes of the spot exchange rate.[5] The sample period is April, 1, 1992 until November 10, 1996 totalling $T=1,169$ observations.

The market convention is to quote prices for at-the-money (ATM) forward[6] straddles, i.e. a call and a put with an identical strike equal to the current forward exchange rate. At each observation date we have a full term structure of implied volatility as defined by the one, two, three, six, nine, and twelve months maturities, $\tau=1,..,6$. The strike on these options is set relative to the instantaneous price of the underlying, while the maturity is a fixed period relative to the observation date. Unlike exchange traded options with fixed nominal strike and a calendar maturity date, these OTC data are devoid of maturity and strike mismatch encountered in previous research. For notational convenience the dependence of implied volatilities on the strike price is henceforth suppressed. Similarly, the bid/ask distortions discussed earlier are avoided by averaging bid and ask quotes to give the mid market observation.

---

[4] Quoted on REUTERS Page TRDO and Telerate Page 4720. I am grateful to Banker's Trust for collecting the data, and making them available for this research.

[5] Similar data-set, alas devoid of the contemporaneous price of the underlying, was investigated by Campa and Chang (1995).

[6] So that the strike, $E$ equal to the forward price of the underlying, $F_t$.

The market convention is to quote prices in terms of BS implied volatilities (4.1) rather than option premia. Provided the *volatility risk is unpriced*, this method is justified even under the stochastic volatility dynamics (4.2). In this case the price of the call option, $C_{SV}$ is given by the expectation of the BS price over the distribution of the average variance (Hull and White, 1987):

$$C_{SV} = E\left[BS(\overline{V})\right] \tag{4.6}$$

where $\overline{V}$ is the average variance until contract expiry. Furthermore, because the BS formula is approximately linear in volatility (Cox and Rubinstein, 1985) and roughly linear in variance (Lamoureux and Lastrapes, 1993) the expectation operator and the function in (4.6) may be interchanged, to give:

$$C_{SV} \approx BS\left[E(\overline{V})\right] \tag{4.7}$$

So that the BS implied volatility obtained via (4.1) is reinterpreted as the expectation of the average volatility over the remaining life of the option. From the practical point of view, (4.7) eliminates the need to filter option prices through a more sophisticated pricing model in order to obtain values of implied volatilities. This task is unfeasible for two reasons. First, most stochastic volatility models involve costly numerical simulations. Semi-closed form solutions (Heston, 1993; Stein and Stein, 1991) are only valid for specific processes which may be mis-specified. Second, the parameters which govern the evolution of the volatility process have to be estimated (Renault and Touzi, 1996) thus inducing an additional source of error. Instead, our results are conditioned upon the validity of the pricing model (4.6) and thus present a joint test of model validity and market efficiency.

The underlying instrument is the DEM/USD exchange rate and summary statistics are given in **Table 4.1**, Panel A. The data are strongly consistent with many earlier studies

of exchange rates (Baillie and Bollerslev, 1989; Diebold and Nason, 1990) in that the unconditional density of $r_t$ (as characterised by the first four sample moments) is: (*i*) centred around zero ($m_1$=-7.0·$10^{-5}$); (*ii*) negatively skewed ($m_3$=-0.13); and (*iii*) exhibits large kurtosis ($m_4$=7.58). The minimum (maximum) observation is 6.88 (4.96) standard deviations away from the mean.

More importantly, returns appear to have little time series dynamics while their squares and log squares do. This is captured by the Box-Ljung $Q$-statistic which is barely significant for $r_t$ ($Q_{10}$=21.38, $Q_{40}$=51.25), but is highly significant, both for $r_t^2$ ($Q_{10}$=64.42, $Q_{40}$=156.00) and $\ln r_t^2$ ($Q_{10}$=209.96, $Q_{40}$=440.69) with the relevant $\chi^2_{10}$ and $\chi^2_{40}$ 5% critical values being 18.3 and 55.8. The correlogram of $r_t$, *(a)* and $\ln r_t^2$, *(b)* together with the $\pm 2T^{-0.5}$ bands[*] is presented in **Figure 4.1**. The graphs are indicative of little autocorrelation in the return series and some time series dynamics for $\ln r_t^2$.



*(a)*          *(b)*
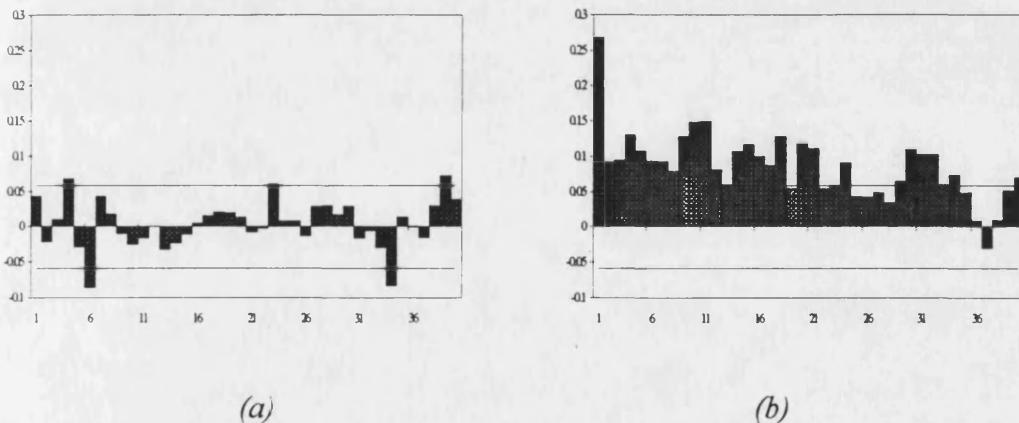
**Figure 4.1:** DEM/USD: Correlogram of $r_t$, and $\ln r_t^2$

We proceed by fitting the basic SV model:

$$\begin{cases} r_t = \sigma_t v_t & v_t \sim N(0,1) \\ \ln \sigma_t^2 = \ln \bar{\sigma}^2 + h_t & \\ h_t = \phi h_{t-1} + \eta_t & \eta_t \sim N(0, \sigma_\eta^2) \end{cases} \quad (4.8)$$

---

[*] In large samples the autocorrelations from a white noise process are approximately uncorrelated and normal with mean zero and variance $T^{-1}$ (Harvey, 1993, p.42).

to the entire sample period, results of which are presented in **Table 4.1**,Panel B. Volatility is found to be persistent with the estimate of the autoregressive parameter taking the value of $\phi$=0.911. The coefficient of variation ($CV$=1.5) is well within the region documented in earlier studies (and summarised in **Table 3.1**). The graph of the contours of the likelihood function, presented in **Figure 4.2***(a)*, suggests a unique maximum in the interior of the parameter space. Using the results of Chapter 3, the hypothesis of a unit root in the (log)variance process is rejected at the 1% level by comparison of the likelihood ratio test statistic, $\xi_{LR}$=25.2 with the critical value of 5.412. **Figure 4.2**(b) shows the in-sample estimates of the volatility process (light line) obtained from the basic SV model (4.8). For comparison the one month implied volatility (heavy line) is also presented.



*(a)*             *(b)*
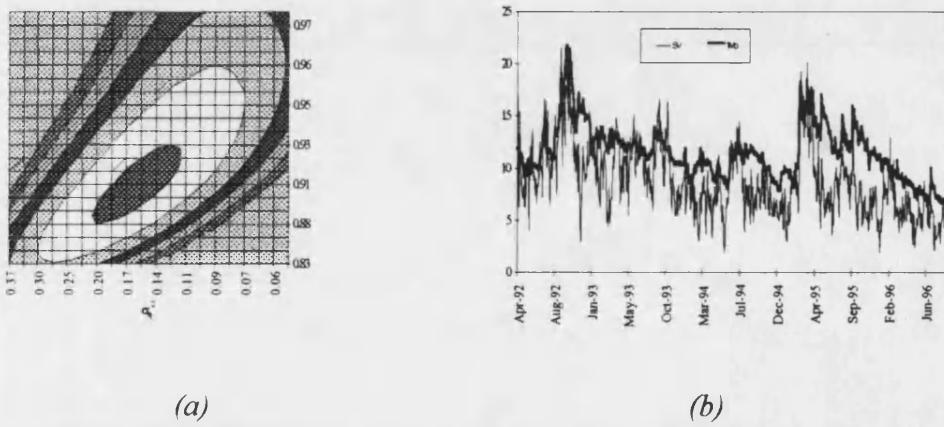
**Figure 4.2**: DEM/USD: Likelihood contours and volatility estimate

The graph suggests that the implied volatility can be regarded as a slowly moving component around which the daily conditional volatility evolves. This motivates the use of implied volatility data as proxies for the time varying mean of the volatility process. Notice, that in our sample the implied volatility is generally above the volatility estimated from the price process[8], the implications of which will become evident in Section 4.4.2. Several suggestions may be put forward as to why options implied volatility is above historical volatility. A premium associated with small but

non-negligible possibility of a market crash may lead to this phenomenon. Alternatively, counterparty risk premium may be priced in this OTC market leading to higher option premia and hence, implied volatilities.

Summary statistics of implied volatility data are presented in **Table 4.2**. While the mean levels across maturities are equal[9] implied volatilities of shorter maturities fluctuate considerably more than those with longer time to expiry: the standard deviation is a strictly declining function of maturity. This leads to periodical inversion of the slope of the term structure, a feature also captured by the fact that the minimum observation is lower and the maximum observation is higher the shorter the maturity. As can be seen in Panel C, the correlogram and partial correlogram for all maturities resemble closely the ACF and PACF of a first order autoregressive process: the ACF decreases slowly and the PACF has a peak at the first lag and is negligible otherwise. Thus each individual series is well described by a mean reverting process. Panel B also presents the correlation matrix of implied volatilities. As expected, correlation is a decreasing function of the distance in maturities: the implied volatilities of neighbouring maturities have higher correlation than those further apart.

### 4.3.2. Analysis of the volatility curve

The term structure of implied volatility is defined as a collection of six implied volatilities with varying maturities, $\tau=1,..,6$. The whole data set is represented in a matrix, **X**, of size $(T \times p)$ so that each row gives the volatility curve at time $t$ while each column gives the time series of $\sigma_{impl,t}(\tau)$ for a given maturity, $\tau$. It is clear from the summary statistics of **Table 4.2** that the evolution of the entire volatility curve can be represented by a set of variables with dimension strictly less that six.

---

[8] The same regularity is reported by Jackwerth and Rubinstein (1996, p.1613).

One way of analysing the dynamics of the term structure of implied volatility is principal component analysis (Basilevski, 1994). This is a general technique aimed at determining the factors which account for most of the co-variation in a set of variables, $X$. The objective of the method is to find a unique orthonormal linear decomposition $Z=XP$ which is obtained by diagonalising the $(p \times p)$ sample covariance matrix, $\Sigma = X'X$.[10] The principal component analysis is performed on the covariance matrix rather than the correlation matrix since ($i$) all series are measured in same units, and ($ii$) it allows us to assign meaningful interpretations to individual components (Basilevski, 1994). The eigenvectors of $\Sigma$, denoted by $v_i$ form columns of $P$ and are ordered in such a way that the first corresponds to the largest eigenvalue, $\lambda_i$, the second to the second largest etc. where $\lambda_i$ are the diagonal elements of $L=R'X'XP$.

The resulting transformation of $X$ to a new set of regressors, $Z$ is such that the elements of $Z$ are pairwise uncorrelated and of which the first will have the maximum possible variance, the second the maximum possible variance among those uncorrelated with the first, and so forth. The columns of $Z$ - denoted by $z_i$ - sorted in order of their contribution to the covariance matrix $\Sigma$ are called the principal components of $X$. The proportion of the variance explained by each $z_i$ is measured by the ratio of each $\lambda_i$ to the total sum of the eigenvalues. The advantage of the technique is that a small number, $q < p$ of principal components, $z_i$ $i=1,..,q$ can explain a large proportion of the variability in the data. On the other hand, the method does not deliver any information about the dynamic properties of $z_i$.

Principal component analysis is sufficient for present purposes since the aim is to construct easily interpretable proxy variables for the in-sample evolution of the

---

[9] Comparing the distribution of the test statistic in the lower diagonal part of Panel B with the critical value of 1.96 leads to a rejection of the hypothesis of a difference in means for all maturities.

[10] The multiple $(T-1)^{-1}$ may be omitted w.l.o.g.

volatility curve. When compared to static factor analysis (Everitt, 1984; Bartholomew, 1987) the uniqueness of the decomposition facilitates the interpretation of the components. By contrast, dynamic factor analysis (Engle and Watson, 1981) while delivering a structural representation of the implied volatility curve requires additional assumptions, e.g. regarding the expectation generation mechanism (Xu and Taylor, 1994).

Assuming that the observed volatility curve results from the asset price dynamics of the kind (4.2), let $X$ be a ($T \times 6$) matrix of logarithms of squared implied volatilities.[11] The results of the principal component analysis, as reported in **Table 4.3** allow three conclusions to be drawn. First, column 3 indicates that 99.6% of the variability in the term structure of implied volatility can be explained by the first two principal components of which the leading one captures 97.3%. This is not surprising given the high degree of correlation between individual implied volatility series. Second, the first eigenvector, $v_1$ is strictly positive. This allows us to reinterpret the leading component as the *level* of the term structure, since it is some average of implied volatilities of different maturities with positive weights, $z_i = Xv_i$. By contrast, the second eigenvector $v_2$ has negative weights on the shorter and positive weights on the longer maturities effectively forming differences between the short and long maturities. We interpret it as a measure of the *slope* of the volatility curve.

---

[11] The log transformation does not significantly alter the covariance structure of the variables, the composition of the eigenvectors, and relative weights. See also the discussion in Bartholomew (1987, pg. 40) regarding the indeterminacy with regard to the log transformation.

## 4.4. Informational content of implied volatility

### 4.4.1. In-sample explanatory power

In this Section the hypothesis that implied volatility contains relevant information as to the evolution of the (log)variance of the underlying asset is examined. The richness of the data-set allows for a detailed investigation of the following questions:

*(i)* Can in-sample return volatility be exclusively captured by implied volatility?

*(ii)* Is information regarding the slope of the volatility curve relevant?

*(iii)* Is information encoded in the entire volatility curve relevant?

The research objectives are addressed by extending the basic SV model (4.8) to include a set of $K$ explanatory variables:

$$\begin{cases} r_t = \sigma_t v_t & v_t \sim N(0,1) \\ \ln\sigma_t^2 = Z_t'\gamma + h_t & \\ h_t = \phi h_{t-1} + \eta_t & \eta_t \sim N\left(0,\sigma_\eta^2\right) \end{cases} \qquad (4.9)$$

where $Z_t = (1, z^1{}_t, \ldots, z^K{}_t)'$ and $\gamma$ is a $(K+1\times1)$ vector. The first element of $Z_t$ is unity so that the corresponding element of $\gamma$ is $\ln\bar{\sigma}^2$. Model (4.9) collapses to (4.8) if the remaining elements of $\gamma$ are zero. In this case the explanatory variables are seen as having no informational content. By contrast, if $\phi$ is zero, the autoregressive component is redundant and the explanatory variables capture accurately the asset's volatility dynamics (apart from an *IID* disturbance, $\eta_t$).[12]

Formulation (4.10) is more appropriate than the GARCH counterpart (4.5) employed by Day and Lewis (1992) and Lamoureux and Lastrapes (1993) since the *instantaneous*

---

[12] Testing for $\sigma_\eta^2 = 0$ is more problematic since the autoregressive parameter is not identified under the null hypothesis (Harvey and Streibel, 1997).

impact of the explanatory variables is modelled, rather than some weighted average of past values, as equation (4.5) demonstrates.

In order to verify the working hypothesis the following variables were selected:

$$1MoVol = \ln\sigma^2_{impl,t}(1),$$

$$12MoVol = \ln\sigma^2_{impl,t}(6),$$

$$Slope = \sigma^2_{impl,t}(1) - \sigma^2_{impl,t}(6),$$

$$PC1, PC2$$

where $\sigma^2_{impl,t}(\tau)$ is the (squared) implied volatility of maturity $\tau$, and $PC1$ and $PC2$ are first two the principal components obtained in Section 4.3.2. The first research objective is addressed by comparing the explanatory power of $1MoVol$ and $12MoVol$ and the extent to which the autoregressive component in (4.9) remains significant after the introduction of explanatory variables. The $Slope$ variable (defined as the difference between implied volatilities of short and long maturities) gives a measure of the steepness of the volatility curve. Inclusion of this variable in conjunction with $1MoVol$ allows inferences about the second working hypothesis to be made. And finally, question (iii) is addressed by examining the explanatory power of the entire volatility curve, which - due to the multicollinearity issues - is approximated by the first two principal components.

The results of the estimation for the entire sample of $T=1,169$ observations are presented in **Table 4.4**, Panel A. For ease of comparison the column labelled $M0$ reproduces the results of the estimation of the basic SV model, already presented in **Table 4.1**, Panel B. The general finding is that the coefficients on all explanatory variables are significant, which can be verified by comparing the ratio of the coefficient estimate to the standard deviation (reported in parenthesis below) to the limiting

normal distribution. Similarly, the likelihood ratio tests (reported in the row labelled *LR1*) indicate the significance of the explanatory variables when compared to the relevant critical values of the $\chi^2$ distribution.

The estimates of the hyperparameters $\phi$ and $\sigma^2_\eta$ change considerably depending on the set of explanatory variables. For instance, estimates of the autoregressive coefficient decline from $\phi$=0.911 in the basic SV model, *M0* to $\phi$=0.569 in model *M3*. However, the hypothesis $H_0$: $\phi$=0 is rejected in all cases by observing that *(a)* the *t*-statistics, and *(b)* the likelihood ratio tests (reported in the row labelled *LR2*) are significant at 1% level, except for model *M4* where the *LR2* test is significant at 5% level. Day and Lewis (1992) and Lamoureux and Lastrapes (1993) also found that the estimates of the components describing the evolution of conditional variance declined, but remained statistically significant after the introduction of explanatory variables. This finding led the authors to the conclusion that implied volatilities are "insufficient statistics" (Day and Lewis, 1992, p. 278) for the conditional variance specification. Recall, that both of these studies employ a single implied volatility measure which is not time homogenous, with varying maturity of three to six months.

By contrast, our data allow a more detailed examination of this issue, leading ultimately to a different interpretation. Consider the estimation results of models *M1* and *M2* and recall that in *M1* only the *1MoVol* is used as the sole explanatory variable while only *12MoVol* is used in *M2*. The important observation is that the change in volatility persistence is not identical across the two models: the autoregressive coefficient is found to be lower for the model *M1* ($\phi$=0.578) than for *M2* ($\phi$=0.892). The in-sample explanatory power of implied volatility is thus related to the maturity of the option: implied volatilities of shorter maturities lead to lower degrees of volatility persistence.

Notice, that the autoregressive component in (4.9) appears relevant in either model. However, this finding is not robust across sub-samples. Both, *rolling* and *updating* sample construction procedures are considered. With 500 observations reserved for the minimum sample size[13], our sample length, $T$=1,169 permits 669 estimations of each model. Sample means and standard deviations of parameter estimates are reported in **Table 4.4**, Panel B. Irrespective of the sample construction procedure the volatility persistence parameter in model *M1* declines sharply. On average, estimates of $\phi$ are centred around 0.100 (0.223) for the rolling (updating) samples. Notice, that the sampling standard errors indicate a large number of sub-samples for which estimates of $\phi$ are, in fact, negative. By contrast, estimates of $\phi$ in model *M2* are centred around 0.544 (0.817) for the rolling (updating) samples. Thus, in many sub-samples implied volatilities of 1Mo maturity render the autoregressive component redundant, which cannot be achieved by 12Mo volatility figures. Xu and Taylor (1995) who also examined foreign currency options obtained a similar, but stronger result: in their data-set the time series of implied volatility is by itself sufficient to describe the evolution of conditional volatility.

The second research objective (importance of the slope information) is addressed by examining the significance of the *Slope* variable in model *M3*. First, the $t$-statistic associated with the estimated coefficient (2.88) is significant at 1% level. Second, the likelihood ratio test - constructed by reference to the maximised likelihood function values in models *M3* and *M1* of **Table 4.4**, Panel A - takes the value 7.80 which is significant at 1% when compared to the critical value of $\chi_1^2$. When sub-samples are examined (**Table 4.4**, Panel B), the estimates of the volatility persistence parameter are centred at -0.031 (0.230) for the rolling (updating) samples, indicating that the two

---

[13] Chapter 2 showed that sample sizes of 500 observations are large enough to successfully estimate SV models. Fewer observations may lead to the likelihood function being ill-conditioned, or "flat"

variables together capture the evolution of the DEM/USD volatility even more accurately, than the level of the implied volatility on its own.

Very similar results are obtained when the information encoded in the entire volatility curve (as approximated by the first two principal components, *PC1* and *PC2*) is inserted into the variance equation, as has been done in model *M4*. Since models *M4* and *M3* are non-nested, likelihood ratio tests cannot be performed. Examination of the hyperparameter estimates for the entire sample and across sub-samples indicates little difference between the two specifications. We interpret this as weak evidence against the hypothesis that the information encoded in the entire volatility curve is required to explain the in-sample evolution of return volatility.

Summing up, the introduction of implied volatility data into the variance equation of the SV model shifts the estimate of the volatility persistence parameter, $\phi$ towards zero. By implication, contrary to the claims made in previous research, most of the time series dynamics of conditional volatility can be explained by implied volatilities. We interpret the discrepancy in the light of measurement problem associated with selecting the most informative volatility measure, which we find to be given by options of very short maturities.

## 4.4.2. Out-of-sample forecasting power

The preceding Section indicated that the information encoded in the term structure of implied volatility explains most of the in-sample time series dynamics of the conditional variance. However, the question regarding the predictive ability of implied volatilities should be addressed in the context of an out-of-sample forecasting experiment. If the volatility risk is unpriced, equations (4.1) and (4.7) show that implied volatility is the expectation of the average volatility of the underlying instrument until option

over the relevant region of the parameter space.

130

expiration. In these circumstances no time series model should exhibit better forecasting performance than implied volatility.

The predictive power of various models is examined in its ability to forecast the *realised, or ex post volatility* over the remaining contract life. The realised volatility is conventionally taken as the average of squared returns over the remaining life of the contract:

$$\tilde{\sigma}_{t,t+L} = \sqrt{L^{-1}\sum_{l=1}^{L} R_{t+l}^2} \qquad L = 20,40,60,130 \tag{4.10}$$

where, as before, $R_s$ is the return on the asset. The lead period, $L$ corresponds to the approximate number of trading days in respectively one, two, three and six months period. Thus the forecasting horizon and the maturity of the option are not only matched but also constant throughout the sample. This is important, since it warrants the comparison of the forecasts across various forecast horizons.

The figure is annualised by a factor of $\sqrt{260}$ for ease of comparison with implied volatility figures, which are scaled by the same factor. The measure (4.10) is widely used in the context of volatility forecasting experiments (Day and Lewis, 1992; Lamoureux and Lastrapes, 1993; Jorion, 1995; West and Cho, 1995; Xu and Taylor, 1995). The justification of which is given by the following argument. If the mean of the return process is zero (or the mean is negligibly small as is often the case with financial time series) then the squared return is a proxy for the realisation of $\sigma_t^2$. The measure (4.10) will therefore give an estimate of the average realised volatility over the time period, $L$.

Clearly, if the true volatility process is (4.2), then (4.10) can only be considered to be a noisy proxy for the long run average volatility level. The obvious shortcomings of this measure necessitate the consideration of an alternative proxy for the ex post volatility.

Under the null hypothesis of the SV model being true, an alternative measure of the ex post variance is given by the in-sample volatility estimate from the SV model estimated at a later time period:[14]

$$\tilde{\sigma}_{t,t+L} = \sqrt{L^{-1}\sum_{l=1}^{L}\hat{\sigma}_{t,t+l|s}^2} \qquad s \geq t + L \qquad (4.11)$$

The measure (4.11) will be subject to estimation error as well as model misspecification error. In the event, both ex post volatility measures lead to identical results when contrasted with the implied volatility value $\sigma_{impl,t}(\tau)$, and the forecasts, $\hat{\sigma}_{t,t+L}$ from the basic SV model (4.8).

Again, rolling and updating samples are considered. At each consecutive time step $l$-days forecasts from the basic SV model (4.8) are generated in accordance with the forecast function (2.30) reproduced here for ease of reference:

$$\tilde{\sigma}_{T+l|T}^2 = \overline{\sigma}_T^2 \, e^{Z_{T'}\gamma + \phi^l h_{T|T}}, \qquad l = 1,...,L \qquad (4.12)$$

where the scale parameter $\overline{\sigma}_T^2$ is defined in (2.28). The volatility forecast is taken as the average of one-step ahead forecasts over the forecasting horizon, $L$, and annualised by the factor $\sqrt{260}$. The results of the out-of-sample forecasting experiment are reported in **Table 4.5**. Summary statistics of the forecast errors, $u_t(L) = \hat{\sigma}_{t,t+L} - \tilde{\sigma}_{t,t+L}$, where the volatility proxy, $\tilde{\sigma}_{t,t+L}$ is calculated as in (4.10) are reported in Panel A and allow three conclusions to be drawn. First, in terms of average errors (*AVG*), mean absolute deviation (*MAD*), and root mean squared error (*RMSE*) the forecasts constructed from time series models are at least a good *and often more accurate* than the forecasts given by implied volatility. This applies to all forecasting horizons. In general, implied volatilities are upward biased forecasts (AVG>0) of future realised

---

[14] The results are robust to the alternative specification: $\tilde{\sigma}_{t,t+L} = L^{-1}\sum\hat{\sigma}_{t,t+l|s}$ .

volatility, indicative evidence of which was represented in **Figure 4.2**. This does not however, imply that the SV model forecasts are excellent predictors of future ex post volatility. On the contrary, particularly over short forecasting horizons, the forecasts can be extremely inaccurate, as the *MIN/MAX* statistics indicate.

Second, forecasts constructed from updating samples are better than those constructed from rolling samples. This is not surprising, since larger sample sizes lead to more accurate parameter estimation. Hence the forecasting ability is improved.

Furthermore, all *RMSE*s and *MAD*s decline as the forecasting horizon increases. This property is due to two facts. First, the variability in the ex post variance proxy (4.10) stabilises as the forecasting horizon increases. Second, the predictors from time series models revert to the long run mean levels. In both instances the comparison of the forecasting power becomes less meaningful and reduces to the comparison of the forecasting power of the long run mean.

The results concerning the sub-optimality of implied volatility forecasts are unaltered when an alternative ex post volatility measure (4.11) is considered, as can be seen from Panel B of **Table 4.5**. Again, the forecasts obtained from the time series models are more accurate (in terms of *AVG*s, *MAD*s and *RMSE*s) than implied volatility figures. Finally, the accuracy of all predictions across all horizons is increased when (4.11) is taken as the ex post volatility proxy. This is indicative of the decreased noise component in the second ex post volatility measure.

This forecasting experiment lends further support to the claim that better volatility forecasts can be constructed based upon time series specifications of conditional variance than simply upon implied volatility figures. Similar conclusion was reached by Day and Lewis (1992) and Lamoureux and Lastrapes (1993) who examined stock index options. On the other hand Xu and Taylor (1995) and Jorion (1995) show that

opposite results can be obtained in the foreign exchange options market where measurement problems associated with stale prices, transaction costs and bid/ask bounce are absent. This Chapter demonstrates that even within the foreign exchange options market the question regarding the predictive ability of implied volatility remains open.

## 4.5. Summary and conclusion

This Chapter re-examined the hypothesis that implied volatility of options contain relevant information about the evolution of the latent volatility process of asset returns. Contrary to the results reported in earlier studies (Day and Lewis, 1992; Lamoureux and Lastrapes, 1993), we find that *in-sample* the implied volatility captures most of the time series dynamics of the conditional volatility of the return process. The discrepancy is related to measurement problems encountered in previous research associated with selecting the most informative implied volatility measure. It is shown that the explanatory power of implied volatility varies with the maturity of the contract: implied volatilities of short maturity (1Mo) are more informative.

However, the *out-of-sample* forecasting experiment suggests that the predictions from the basic SV model across all forecasting horizons are at least as accurate as the forecasts obtained from the implied volatility data. This finding augments the conclusions reached in some studies (Day and Lewis, 1992; Canina and Figlewski, 1993; Lamoureux and Lastrapes, 1993) in that implied volatility of options may not be the best forecast of the average realised volatility. The data-set employed here allows our experiment to be set up in such a way that the criticisms raised in response to earlier work by Jorion (1995) do not apply. Specifically, all measurement problems

associated with stale prices, bid/ask quotes, and different forecast horizons are carefully avoided.

To conclude, the results indicate that - if the options market is efficient - the assumptions underlying the SV option pricing model (4.7) may not hold. In particular, a volatility risk premium may be required by the market. Alternatively, demand and supply circumstances caused by preferences of options' market participants for particular risk patterns may invalidate the pricing relationship.

# Table 4.1. Preliminary analysis of DEM/USD exchange rate

**Panel A:** Summary statistics, time period 1/4/1992-10/10/1996, $T$=1,169.

|          | $m_1$          | $m_2$          | $min/m_2$ | $max/m_2$ | $m_3$  | $m_4$   | $Q_{10}$ | $Q_{40}$ |
|----------|----------------|----------------|-----------|-----------|--------|---------|----------|----------|
| $r_t$    | $-7.0 \cdot 10^{-5}$ | $6.5 \cdot 10^{-3}$ | 6.88 | 4.96 | -0.13 | 7.58 | 21.38 | 51.25 |
| $r^2_t$  | $4.2 \cdot 10^{-5}$ | $1.1 \cdot 10^{-4}$ | -0.39 | 18.11 | 8.24 | 112.23 | 64.42 | 156.00 |
| $\ln r^2_t$ | -11.65 | 1.93 | 1.73 | -2.82 | -0.04 | 2.19 | 209.96 | 440.69 |

Panel A reports the sample means $(m_1)$, standard deviations $(m_2)$, skewness $(m_3)$, and kurtosis $(m_4)$, as well as the modulus of the minimum (and maximum) deviation from $m_1$ expressed as a multiple of $m_2$ for the returns $(r_t)$, their squares $(r^2_t)$, and log squares $(\ln r^2_t)$. The Box-Ljung statistic

$$Q_p = T(T+2)\sum_{\tau=1}^{p}(T-\tau)\hat{\rho}^2(\tau)$$

is evaluated at 10 and 40 lags, where $\hat{\rho}(\tau)$ is the sample autocorrelation at lag $\tau$.

**Panel B:** Results of the estimation of the basic SV model, $T$=1,169.

|              | $\phi$  | $\sigma^2_\eta$ | $\ln \bar{\sigma}^2$ | $q$  | $CV$ | $LogLik$ | $\xi_{LR}$ |
|--------------|---------|-----------------|----------------------|------|------|----------|------------|
| $\hat{\psi}$ | 0.911   | 0.155           | -10.634              | 0.03 | 1.50 | -2633.5  | 25.20      |
| $s.e.(\hat{\psi})$ | (0.001) | (0.002)   | (0.137)              | -    | -    | -        | -          |

Panel B reports the parameter estimates of the basic SV model (4.7). The standard deviations of the hyperparameters $(\phi, \sigma^2_\eta)$ are taken from the numerical approximation to the Hessian, while the standard deviation of the estimate of $\ln \bar{\sigma}^2$ is taken from the relevant diagonal element of $P_T$. The remaining columns give the signal-noise ratio, $q$ and the coefficient of variation, $CV$ defined by

$$q = 2\sigma^2_\eta/\pi^2, \qquad CV = \text{var}(\sigma_t)E[\sigma_t]^{-2} = \exp(\sigma^2_\eta(1-\phi^2)^{-1})-1$$

as well as the value of the maximised likelihood $(LogLik)$. The final column reports the likelihood ratio test, $\xi_{LR}$ of $\phi$=1, the distribution of which was shown to be closely approximated by the $0.5\chi_1^2$ density.

# Table 4.2. Summary statistics of the DEM/USD volatility curve

**Panel A:** Summary statistics, time period 1/4/1992-10/10/1996, *T*=1,169.

|        | *1Mo*  | *2 Mo* | *3 Mo* | *6 Mo* | *9 Mo* | *12 Mo* |
|--------|--------|--------|--------|--------|--------|---------|
| $m_1$  | 12.19  | 12.22  | 12.23  | 12.23  | 12.22  | 12.21   |
| $m_2$  | 2.41   | 1.96   | 1.63   | 1.22   | 0.97   | 0.79    |
| *Min*  | 7.80   | 8.65   | 9.05   | 10.00  | 10.35  | 10.53   |
| *Max*  | 21.80  | 20.75  | 18.75  | 16.20  | 15.10  | 14.10   |
| $m_3$  | 1.04   | 0.96   | 0.69   | 0.54   | 0.20   | 0.00    |
| $m_4$  | 1.58   | 1.22   | 0.39   | 0.18   | -0.52  | -0.84   |

Panel A reports the sample means $(m_1)$, standard deviations $(m_2)$, skewness $(m_3)$, and kurtosis $(m_4)$, as well as the minimum and maximum values for the implied volatilities of varying maturities.

**Panel B:** Correlation matrix and test of the equality of the means, *T*=1,169

| *t/corr* | *1Mo*  | *2 Mo* | *3 Mo* | *6 Mo* | *9 Mo* | *12 Mo* |
|----------|--------|--------|--------|--------|--------|---------|
| *1Mo*    | -      | 0.99   | 0.97   | 0.94   | 0.91   | 0.88    |
| *2 Mo*   | -0.05  | -      | 0.99   | 0.97   | 0.95   | 0.92    |
| *3 Mo*   | -0.05  | -0.03  | -      | 0.99   | 0.97   | 0.95    |
| *6 Mo*   | -0.04  | -0.02  | -0.01  | -      | 0.99   | 0.97    |
| *9 Mo*   | -0.02  | 0.00   | 0.01   | 0.04   | -      | 0.99    |
| *12 Mo*  | -0.01  | 0.00   | 0.02   | 0.04   | 0.04   | -       |

The upper triangular part of Panel B contains the correlation figures $(\rho_{ij})$. The lower triangular part - the values of the test statistic for the difference in the mean:

$$t_{ij} = \frac{\hat{\mu}_i - \hat{\mu}_j}{\sqrt{\hat{\sigma}_i^2 + \hat{\sigma}_j^2 - 2\rho_{ij}\hat{\sigma}_i\hat{\sigma}_j}}$$

**Panel C:** Correlogram and sample partial autocorrelations, $T=1,169$

| | 1 Mo | 2 Mo | 3 Mo | 6 Mo | 9 Mo | 12 Mo |
|---|---|---|---|---|---|---|
| *ACF* | | | | | | |
| *1* | 0.97 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| *2* | 0.94 | 0.95 | 0.95 | 0.96 | 0.96 | 0.97 |
| *3* | 0.91 | 0.93 | 0.93 | 0.94 | 0.94 | 0.95 |
| *4* | 0.89 | 0.91 | 0.91 | 0.92 | 0.93 | 0.94 |
| *5* | 0.87 | 0.89 | 0.90 | 0.90 | 0.91 | 0.92 |
| *6* | 0.85 | 0.87 | 0.88 | 0.88 | 0.90 | 0.91 |
| *7* | 0.83 | 0.85 | 0.86 | 0.87 | 0.88 | 0.89 |
| *8* | 0.81 | 0.83 | 0.84 | 0.85 | 0.87 | 0.88 |
| *9* | 0.80 | 0.82 | 0.83 | 0.84 | 0.86 | 0.87 |
| *10* | 0.78 | 0.80 | 0.82 | 0.83 | 0.84 | 0.85 |
| *PACF* | | | | | | |
| *1* | 0.97 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| *2* | -0.09 | -0.08 | -0.08 | -0.13 | -0.05 | -0.04 |
| *3* | 0.10 | 0.11 | 0.07 | 0.09 | 0.05 | 0.03 |
| *4* | 0.07 | 0.01 | 0.04 | -0.01 | 0.05 | 0.02 |
| *5* | 0.01 | 0.07 | 0.03 | 0.03 | -0.01 | 0.04 |
| *6* | -0.04 | -0.04 | -0.02 | 0.04 | 0.01 | -0.03 |
| *7* | 0.03 | 0.00 | -0.01 | 0.02 | 0.03 | -0.03 |
| *8* | 0.00 | -0.02 | -0.01 | -0.02 | -0.02 | 0.01 |
| *9* | 0.10 | 0.07 | 0.08 | 0.08 | 0.03 | 0.03 |
| *10* | -0.04 | 0.03 | 0.00 | 0.00 | 0.02 | 0.03 |

Panel C reports the sample autocorrelations and partial autocorrelations for the implied volatility series of each maturity.

**Table 4.3. Principal component analysis of the DEM/USD volatility curve**

|      | $\lambda$ | %     | 1 Mo   | 2 Mo   | 3 Mo   | 6 Mo   | 9 Mo   | 12 Mo  |
|------|-------|-------|--------|--------|--------|--------|--------|--------|
| #1   | 0.374 | 97.3% | 0.608  | 0.502  | 0.424  | 0.313  | 0.249  | 0.198  |
| #2   | 0.009 | 2.3%  | -0.621 | -0.082 | 0.221  | 0.401  | 0.444  | 0.449  |
| #3   | 0.001 | 0.2%  | 0.447  | -0.471 | -0.444 | -0.090 | 0.311  | 0.525  |
| #4   | 0.000 | 0.1%  | 0.158  | -0.357 | -0.164 | 0.759  | 0.050  | -0.493 |
| #5   | 0.000 | 0.0%  | 0.141  | -0.626 | 0.740  | -0.181 | -0.087 | -0.035 |
| #6   | 0.000 | 0.0%  | -0.013 | 0.027  | 0.009  | -0.353 | 0.797  | -0.490 |

This table reports the eigenvectors and eigenvalues of the covariance matrix of implied volatilities of the DEM/USD exchange rate, $T$=1,169. The column labelled "%" gives the proportion of total variance explained by each individual component .

## Table 4.4. Significance of the explanatory variables in the variance equation

**Panel A:** Entire data set, $T=1,169$.

| | M0 | M1 | M2 | M3 | M4 |
|---|---|---|---|---|---|
| $\phi$ | 0.911 | 0.578 | 0.892 | 0.569 | 0.828 |
| | (0.001) | (0.008) | (0.001) | (0.009) | (0.003) |
| $\sigma^2_\eta$ | 0.155 | 0.291 | 0.145 | 0.277 | 0.121 |
| | (0.002) | (0.010) | (0.002) | (0.010) | (0.004) |
| Const | -10.634 | -2.706 | 3.868 | -5.515 | -12.746 |
| | (0.137) | (0.695) | (2.614) | (1.197) | (2.320) |
| 1MoVol | | 1.497** | | 0.965** | |
| | | (0.130) | | (0.225) | |
| 12MoVol | | | 2.772** | | |
| | | | (0.499) | | |
| Slope | | | | 158.940** | |
| | | | | (55.724) | |
| PC1 | | | | | 0.824** |
| | | | | | (0.093) |
| PC2 | | | | | 3.705** |
| | | | | | (0.584) |
| LogLik | -2633.50 | -2600.30 | -2626.30 | -2596.40 | -2604.70 |
| LR1 | - | 66.40** | 14.40** | 72.20** | 57.60** |
| LR2 | 90.80** | 10.00** | 40.60** | 30.00** | 5.60* |

This panel reports the estimation results of the extended SV model (4.9) for the entire sample period. The models are labelled M0-M4 depending on the choice of the explanatory variables. The row labelled LR1 represents the likelihood ratio test against M0. $LR1 \sim \chi^2_K$, where $K=1,2$ is the number of explanatory variables in excess of M0. The row labelled LR2 represents the likelihood ratio test of $H_0$: $\phi=0$ (so that each model was re-estimated with $\phi=0$, details of which are not reported). $LR2 \sim \chi^2_1$. The standard deviations of the hyperparameters $(\phi, \sigma^2_\eta)$ are taken from numerical approximation to the Hessian, while the standard errors of $\hat{\gamma}$ are taken from the diagonal elements of $P_T$. Significance of the coefficients at (1%) 5% level is denoted by a (double) star.

**Panel B:** Hyperparameter estimates across sub-samples

| *rolling* | *M0* | *M1* | *M2* | *M3* | *M4* |
|-----------|------|------|------|------|------|
| $\phi$ | 0.792 | 0.100 | 0.544 | -0.031 | 0.080 |
|  | (0.133) | (0.451) | (0.300) | (0.467) | (0.438) |
| $\sigma^2_\eta$ | 0.256 | 0.275 | 0.316 | 0.236 | 0.226 |
|  | (0.212) | (0.177) | (0.192) | (0.170) | (0.169) |
| *updating* | *M0* | *M1* | *M2* | *M3* | *M4* |
| $\phi$ | 0.907 | 0.223 | 0.817 | 0.230 | 0.252 |
|  | (0.043) | (0.167) | (0.146) | (0.139) | (0.227) |
| $\sigma^2_\eta$ | 0.094 | 0.208 | 0.123 | 0.207 | 0.164 |
|  | (0.063) | (0.071) | (0.099) | (0.069) | (0.083) |

This panel reports the averages and standard deviations of the hyperparameters across *rolling* and *updating* samples. With the minimum sample size of 500 observations the data permit 669 estimation of each model *M0-M4* for each sub-sample construction procedure. Estimates of other model parameters are not reported.

# Table 4.5. Summary statistics of the out-of-sample volatility forecasting errors

**Panel A:** Ex post volatility proxy (4.10)

| IMPL | L=20 | L=40 | L=60 | L=130 |
|------|------|------|------|-------|
| AVG | 1.99 | 2.03 | 2.08 | 2.11 |
| MAD | 2.82 | 2.33 | 2.31 | 2.33 |
| RMSE | 4.08 | 3.72 | 3.67 | 3.58 |
| MAX | 11.10 | 9.28 | 7.84 | 6.61 |
| MIN | -8.23 | -8.22 | -7.60 | -4.26 |
| **ROLL** | **L=20** | **L=40** | **L=60** | **L=130** |
| AVG | -0.57 | -0.76 | -0.87 | -1.09 |
| MAD | 2.82 | 2.62 | 2.42 | 2.11 |
| RMSE | 3.70 | 3.29 | 3.14 | 2.75 |
| MAX | 5.63 | 3.84 | 2.82 | 2.78 |
| MIN | -13.25 | -10.65 | -10.14 | -6.63 |
| **UPD** | **L=20** | **L=40** | **L=60** | **L=130** |
| AVG | -0.13 | -0.21 | -0.27 | -0.47 |
| MAD | 2.79 | 2.56 | 2.39 | 2.08 |
| RMSE | 3.57 | 3.18 | 3.00 | 2.55 |
| MAX | 6.33 | 4.72 | 3.98 | 3.22 |
| MIN | -12.54 | -10.78 | -9.64 | -6.16 |

This table reports the summary statistics of volatility forecast errors $u_t(L) = \hat{\sigma}_{t,t+L} - \tilde{\sigma}_{t,t+L}$. The ex post volatility, $\tilde{\sigma}_{t,t+L}$ is defined as the average of squared returns over $L$ trading days, corresponding to the number of trading days in a 1Mo, 2Mo, 3Mo and 6Mo respectively. The forecast volatility, $\hat{\sigma}_{t,t+L}$ is either the implied volatility $\sigma_{impl,t}(\tau)$ of the relevant maturity (IMPL) or the forecast from the basic SV model (4.10) constructed using rolling samples (ROLL) and updating samples (UPD). The summary statistics are the sample average (AVG), mean absolute deviation (MAD), root mean squared error (RMSE), and the minimum (MIN) and maximum (MAX) values. The statistics in each column are based upon 649,629, 609, and 539 observations.

**Panel B:** Ex post volatility proxy (4.11)

| IMPL | L=20 | L=40 | L=60 | L=130 |
|------|------|------|------|-------|
| AVG | 3.11 | 3.33 | 3.45 | 3.64 |
| MAD | 1.85 | 1.71 | 1.73 | 1.80 |
| RMSE | 3.99 | 4.14 | 4.18 | 4.16 |
| MAX | 10.13 | 8.98 | 8.04 | 6.76 |
| MIN | -5.67 | -4.92 | -3.65 | -0.42 |
| **ROLL** | **L=20** | **L=40** | **L=60** | **L=130** |
| AVG | 0.99 | 1.09 | 1.09 | 1.06 |
| MAD | 1.64 | 1.55 | 1.46 | 1.12 |
| RMSE | 2.35 | 2.36 | 2.23 | 1.74 |
| MAX | 5.10 | 3.95 | 3.83 | 3.39 |
| MIN | -8.86 | -7.48 | -5.34 | -2.39 |
| **UPD** | **L=20** | **L=40** | **L=60** | **L=130** |
| AVG | 0.56 | 0.54 | 0.49 | 0.44 |
| MAD | 1.66 | 1.59 | 1.46 | 1.08 |
| RMSE | 2.24 | 2.14 | 1.98 | 1.45 |
| MAX | 5.11 | 4.09 | 3.53 | 3.64 |
| MIN | -8.52 | -7.36 | -5.39 | -2.86 |

This table reports the summary statistics of volatility forecast errors $u_t(L) = \hat{\sigma}_{t,t+L} - \tilde{\sigma}_{t,t+L}$. The ex post volatility, $\tilde{\sigma}_{t,t+L}$ is defined as the average of in-sample volatility estimates from the basic SV model (4.9) estimated at a later time, $t+L$. The forecast volatility, $\hat{\sigma}_{t,t+L}$ is either the implied volatility $\sigma_{impl,t}(\tau)$ of the relevant maturity (*IMPL*) or the forecast from the basic SV model constructed using rolling samples (*ROLL*) and updating samples (*UPD*). The summary statistics are the sample average (*AVG*), mean absolute deviation (*MAD*), root mean squared error (*RMSE*), and the minimum (*MIN*) and maximum (*MAX*) values. The statistics in each column are based upon 649, 629, 609, and 539 observations.

# Chapter 5: Effect of futures market volume on spot market volatility

## 5.1. Introduction

The focus of this Chapter is the behaviour of volatility in two parallel markets: the equity (spot or cash) market and the market for futures on an equity index. The two markets are linked by arbitrage since the price of the derivative security, the future, depends directly on the price of the underlying cash instrument, the market index. It has been found, however, that the futures returns are more volatile than the corresponding spot returns. For example, Board and Sutcliffe (1995) report several papers on this question. The general finding is of higher volatility for the futures markets, sometimes up to seven times that for the spot market: and only rarely (notably for Japan) are spot returns found to have higher volatility than the corresponding futures returns. This has led to a considerable interest, both academic and regulatory, in the hypothesis that the higher volatility of the futures market might have distortionary effects on spot market prices. For example, following the stock market crash of October 1987, it was claimed that index futures had increased stock market volatility (NYSE, 1990).

One way of testing the significance of the futures market is to model *one-off* effects like the introduction of the futures exchange. A permanent shift in spot market volatility would signify such a causal link. The empirical evidence of the existence of such an effect is inconclusive. Edwards (1988) documents a small but significant decline in cash market volatility after the introduction of the equity futures. By contrast, Harris (1989) finds that the S&P equities are more volatile subsequent to the introduction of the futures trading.

An alternative approach - adopted here - is to model a continuous influence from futures to the spot market volatility. The hypothesis that increases in futures market

trading activity increase spot market price volatility is tested by constructing proxy variables for the relative importance of the futures market. Among previous studies of this question, Schwert (1990) mentions that, when volatility for the S&P500 index is high, stock market and futures volumes are also high, while Smith (1989) found that S&P500 futures volume had no effect on changes in the volatility of S&P500 index returns. Santoni (1987) found a negative correlation between S&P500 futures volume and the daily spot (high-low)/close, suggesting that an increase in futures trading does not lead to an increase in the volatility of the index. Darrat and Rahman (1995) concluded that futures volume did not affect S&P500 spot price volatility. Bessembinder and Seguin (1992) found that expected (i.e. informationless) S&P500 futures trading activity was negatively related to spot market volatility when spot market activity variables were included in the analysis. This result supports the notion that futures trading improves liquidity provision and depth in spot markets, and rejects the hypothesis of the destabilising effect of the futures market.[1] Brown-Hruska and Kuserk (1995) also found evidence for the S&P500 that an increase in futures volume, relative to spot volume, leads to a drop in spot volatility.

However, many of the papers in this area, including those on the effect of futures volume on futures volatility and spot volume on spot volatility, suffer from a number of problems both of model specification and in the construction of the activity variables. These problems are severe enough to bias the results of the empirical investigation, and render their conclusions open to question. The remainder of this Chapter is organised as follows: Section 5.2 contains an outline of the principal difficulties with the approaches used in most previous tests of the hypothesis; Section

---

[1] Bessembinder and Seguin (1992) also find that, although price volatility does increase close to the date of the futures contract's expiry. it is not systematically related to the futures life cycle as a whole.

5.3 describes the data and preliminary tests; Section 5.4 contains the results of tests of the principal hypothesis, while Section 5.5 concludes.

## 5.2. Overview of previous work

We begin by outlining three classes of problem common in volatility tests of spot and futures markets. The first is a problem in the interpretation of the coefficients of exogenous variables commonly added to the variance equation of a GARCH model. The second is a consequence of the simultaneity bias present in variance-volume models, which means that the volume of information-based trading should not be included in the test as an exogenous variable. The third is a difficulty with the common decomposition of aggregate volume into informationless and unpredictable components.

*(i) Implied Lag Structures.* The first difficulty associated with the conventional econometric framework is that it imposes a very rigid structure within which the explanatory variables, $z^k_t$, affects volatility. For example, Chatrath, Ranchander and Song (1996), Foster (1995), Lamoureux and Lastrapes (1990), Najand, and Yung (1991), Sultan, Hogan, and Kroner (1995), Yang and Brorsen (1993) have examined the effect of volume on volatility by adding volume as an exogenous explanatory variable to the variance equation of the GARCH model as in (1.7):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + Z_t'\gamma \tag{1.7}$$

Other studies - e.g. Bessembinder and Seguin (1992, 1993), and Brown-Hruska and Kuserk (1995) - have used the estimation procedure proposed by Schwert (1990). Here two regression equations are formulated which describe the evolution of the

146

mean and the volatility of the process in terms of exogenous and lagged endogenous variables:

$$R_t = c + \sum_{i=1}^{I} \phi_i R_{t-i} + \sum_{j=1}^{J} \pi_j \hat{\sigma}_{t-j} + u_t$$

$$\hat{\sigma}_t = \alpha + \sum_{m=1}^{M} \beta_m \hat{\sigma}_{t-m} + \sum_{l=1}^{L} \delta_l \hat{u}_{t-l} + \sum_{k=1}^{K} \gamma_k z_t^k + \varepsilon_t \qquad (5.1)$$

where $R_t$ represents the return on the asset, $\sigma_t$ is the instantaneous standard deviation of the residuals $u_t$, $z_t^k$ are volume related terms, and circumflexes indicate fitted values from a previous iteration.

While it might seem that these models capture the contemporaneous effect of $z_t^k$ on $\sigma_t$, this is not the case. In fact, as has been mentioned in Chapter 1 the effect of $z_t^k$ on the variance of the process is modelled in terms of a geometrically declining lag structure. This is also true for models (5.1): the coefficient can be shown to be an exponentially weighted average of past values of the volume measure. This is an important drawback since the real issue is the effect of the instantaneous volume in the futures market on stock market volatility.

*(ii) Simultaneity Bias.* In both, GARCH models and specification of the form (5.1) the exogenous (volume related) variables $z_t^k$ and lagged values of volatility are included simultaneously. It is well known that such specifications suffer from simultaneity (or errors in variables) bias since volume cannot be assumed to exogenous. Instead, volatility and volume are jointly determined by the same unobservable variable (i.e. information arrival). To illustrate this point, assume that the true data generation process is:

$$\sigma_t = \alpha I_t + \varepsilon_{1,t} \qquad (5.2.a)$$

$$z_t = \beta I_t + \varepsilon_{2,t} \qquad (5.2.b)$$

where $\sigma_t$ is the conditional volatility (or variance) of the price process, $z_t$ is now volume and $I_t$ is the (latent) information arrival variable. Re-expressing $I_t$ in terms of $z_t$ in (5. 2.b) and substituting into (5.2.a) gives:

$$\sigma_t = \frac{\alpha}{\beta}z_t + \left(\varepsilon_{1,t} - \frac{\varepsilon_{2,t}}{\beta}\right) = \gamma z_t + u_t$$

which yields a system in which the regressor, $z_t$ and the error term, $u_t$ are not orthogonal. As a result, the estimate of the coefficient of interest, $\gamma$, will be biased, and the estimates of any other coefficients that might have been included in the regression equation will also be inconsistent. This observation led Lamoureux and Lastrapes (1994), Andersen (1996) and Liesenfeld (1996), among others, to model volume and volatility jointly in bivariate mixture models:

$$\begin{aligned} r_t &= \varepsilon_t \sqrt{F_t} \\ V_t &= P\left(m_0 + m_1 F_t\right) \end{aligned} \qquad \varepsilon_t \sim N(0, \sigma) \qquad (5.3)$$

where $P(\times)$ denotes the Poisson distribution, $F_t$ is the measure of latent information which drives both volume and volatility. Therefore, if volume is to be included in the variance equation it needs to be separated from its information component. Bessembinder and Seguin (1992), for instance, undertake this separation, but nevertheless include the information-related component in their estimation, leading to the problems of bias and inconsistency described above.

*(iii) Construction of the Activity Variables.* Having selected the estimation model, volume is conventionally decomposed into predictable and unpredictable components which are related to informationless trading activity and news arrival, respectively. For example, Bessembinder and Seguin (1992) first remove the time trend in volume (and log-volume) for both markets. They then fit a 100 day moving average to the series. Finally they fit univariate ARIMA(10,1,0) models to the residuals and interpret the new

errors as unexpected volume, while the fitted values plus the moving average term are regarded as expected volume. The sum of the three series gives the original (detrended) volume series. This is done for the futures and the spot market volumes separately, and all six variables are included as exogenous explanatory variables in the model.

A number of problems with this approach can be identified. First, as spot and futures volumes are typically highly correlated, inclusion of both futures and spot market volume series in the ultimate regression equation for volatility may lead to multicollinearity problems (resulting in difficulties in identifying the true influence of the underlying futures market volume). Second, because the two volume series are inherently interdependent - as suggested in (5.2) - the use of univariate ARIMA models may lead to the omitted variables problem. Finally it is shown below that in our data futures and spot volume are trend-stationary, which means that both detrending and differencing the original data will lead to a loss of information.

## 5.3. Empirical study

### 5.3.1. Data description

Since the real issue underlying the research is the impact of futures trading on the equity market, we define the spot market volume, $VS_t$, as the daily total value of all stocks traded on the London Stock Exchange (in £10m). Similarly, the futures market volume, $VF_t$, is defined as the total nominal value of all FT-SE 100 contracts traded at LIFFE on a particular day (also in units of £10m).[2] To eliminate potential irregularities associated with the 1987 crash, the time period has been chosen to be 4 January 1988 to 14 December 1995, yielding a total of 2,011 daily observations. The influence of

dividend payments is ignored. Even if the uncertainty associated with dividend payments is assumed negligible, lumpy dividend payments might well affect returns. On ex-dividend days the price of a share is reduced by the dividend amount. Thus, if the individual ex-dividend days are unevenly distributed over time and the returns are not corrected for the dividend payments, the returns on the index will appear to be more volatile. In the event, Gallant, Rossi and Tauchen (1992) found that this effect was negligible for the S&P500 index.

Continuously compounded daily return series is constructed from closing values of the FTSE 100 index. Summary statistics of the returns are presented in **Table 5.1**. The returns are centred around the origin ($m_1$=-3.7·10$^{-5}$); are negatively skewed ($m_3$=-1.81); and exhibit large kurtosis ($m_4$=29.94). The moments of the unconditional density provide strong evidence against the hypothesis of the returns being drawn independently from a Gaussian density. Some time series dynamics are present in the return process itself, but the autocorrelation in the squares and log squares is very much stronger. The Box-Ljung $Q$-statistic which is significant for $r_t$ ($Q_{10}$=32.29, $Q_{40}$=66.54), and is highly significant, both for $r^2_t$ ($Q_{10}$=1192.8, $Q_{40}$=1259.1) and $\ln r^2_t$ ($Q_{10}$=185.05, $Q_{40}$=339.71) with the relevant $\chi^2_{10}$ and $\chi^2_{40}$ 5% critical values being 18.3 and 55.8.

The graphs of the two volume series as well the corresponding log-transformations are presented in **Figure 5.1**.

---

[2] The stocks in the FT-SE 100 index account for over 70% of the total capitalisation of the London Stock Exchange.
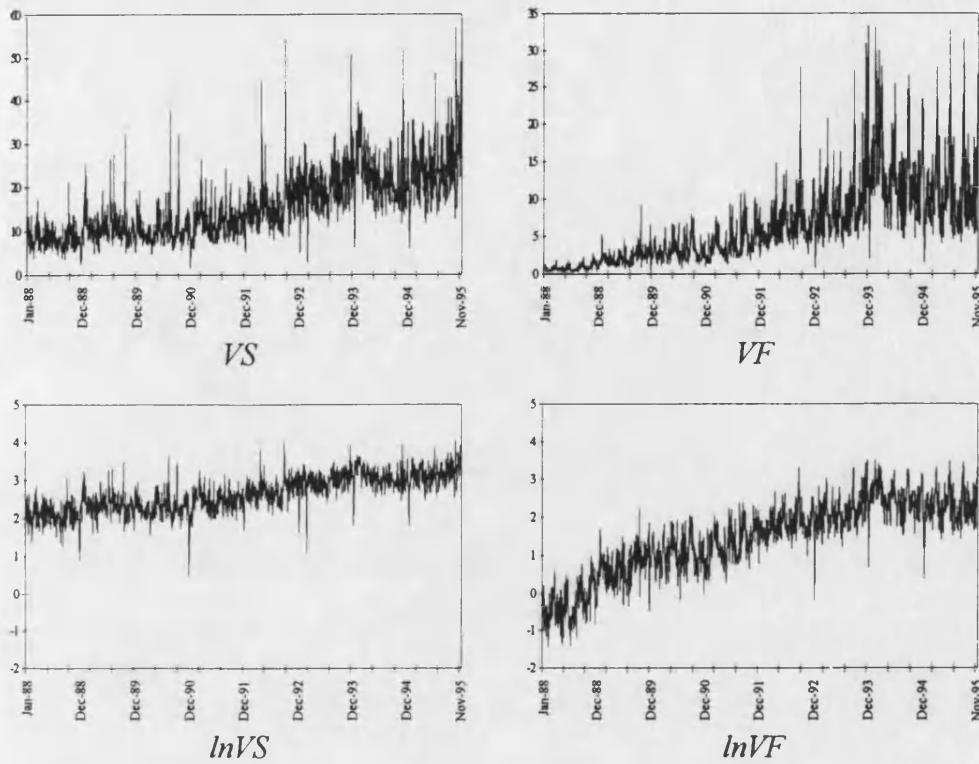
**Figure 5.1:** FTSE100: Futures and spot market volume and log-volume series.

The graphs of the log-transformation of volume lend some support to the hypothesis of a constant linear growth in volume, and this transformation also stabilises the variability of the series. It is interesting to observe that the rates of growth in volume in the two markets are not identical. Futures market volume is growing more rapidly, signifying increasing popularity of the futures market.

It is not clear *a priori* whether volume or log-volume should be used. Most of the theoretical models (Tauchen and Pitts, 1983; Glosten and Milgrom, 1985; Easley and O'Hara, 1992) concentrate on volume. This consensus is not present in the applied work. Gallant, Rossi and Tauchen (1992) and Andersen (1996) employ the log-transformation, while Lamoureux and Lastrapes (1990, 1994) and Jones, Kaul and Lipson (1994) do not. We choose to work with log-volume.

## 5.3.2. Volume and informationless trading activity

Examination of **Figure 5.1** raises the question as to the nature of the non-stationarity present in the volume series. Univariate tests for the presence of a unit root in *ln(VS)* and *ln(VF)* are reported in **Table 5.2**, Panel A. With the augmented Dickey-Fuller statistics being highly significant, the null hypothesis of a random walk with drift is rejected in all cases in favour of the alternative of a trend-stationary process.

As an alternative test, we may assume that *ln(VS)* and *ln(VF)* are integrated and formulate a Vector Error Correction Model (VECM):

$$\Delta Y_t = \alpha + \Pi Y_{t-1} + \sum_{j=1}^{p} B^j \Delta Y_{t-j} + E_t \quad , \quad E_t \sim N(0, \Sigma)$$

where $V_t = (ln(VS_t), ln(VF_t))'$ is a (2×1) vector and $\Pi$, $B^j$, and $\Sigma$ are (2×2) matrices. The rank of the long run impact matrix $\Pi$ is an indicator of the nature of non-stationarity of $V_t$. If rank($\Pi$) = 2 then each element of $V_i$ is trend stationary; when rank($\Pi$) = 1, the series are cointegrated; and when rank($\Pi$) = 0, they are individually $I(1)$ but not cointegrated. Monte Carlo evidence in Lee and Tse (1996) suggests that this test is robust under a variety of alternative assumptions about the disturbance process, $E_t$.

The results of the analysis are presented in **Table 5.2**,Panel B. The hypothesis of the reduced rank of $\Pi$ is strongly rejected, irrespective[3] of the number of lags included in the autoregressive specification and the log-transformation. All statistics are highly significant at 1% level, in particular, the $\lambda^2_{max}$ statistic which rejects the null $H_0$: rank($\Pi$)=1 in favour of $H_1$: rank($\Pi$)=2. Thus, despite the fact that in general the distinction between unit root and trend stationary data is difficult to make (Canova,

---

[3] It is a well known phenomenon that the eigenvalue statistics decrease as the number of lags rises.

1997), in our data set the detrended volume appears to be stationary, rendering additional differencing unnecessary.

In the next stage, we follow Gallant, Rossi and Tauchen (1992) and Andersen (1996) and remove the trend by regressing volume in each market on a constant and a time trend: $v_t = \alpha + \delta t + \varepsilon_t$, where $v_t$ represents $ln(VS_t)$ or $ln(VF_t)$ as appropriate.[4] The residuals from these univariate regressions, $\hat{\varepsilon}_t$ are used to construct the adjusted volume series $v_t^t = a + b\hat{\varepsilon}_t$ where $a$ and $b$ are chosen so that the sample moments of $v_t$ and $v_t^t$ are equal. This rescaling facilitates interpretation of the results by ensuring that the variables are in appropriate units.

Finally, the detrended volume is decomposed into predictable and unpredictable components through the bivariate model:

$$Y_t = \alpha + \sum_{j=1}^{p} B^j Y_{t-j} + E_t \quad , \quad E_t \sim N(0, \Sigma) \tag{5.4}$$

where $Y_t = ( v_{St}^t, v_{Ft}^t )'$. The simultaneous estimation in (5.4) allows the calculation of the level of informationless trading in each market, conditional on the observed level of past trading in the other market. Because of their role in price discovery futures prices are usually found to lead spot prices by a few minutes (Sutcliffe, 1997), but such effects are fast enough to be absent in daily data. However, there do not appear to have been any previous studies of leads and lags between spot and futures volume. **Table 5.3** shows that there are significant cross-market effects in informationless volume, with a number of significant coefficients, for both spot on futures (lags 1, 2,

---

[4] As a preliminary, following Gallant, Rossi and Tauchen (1992), we included dummy variables to identify any weekly deterministic seasonal components. Because none of these dummies was significant, the results are not presented here but are available on request.

and 5) and futures on spot (lags 3 and 5).[5] Univariate tests of Granger causality (Hamilton, 1994, p.304) reject the hypothesis of "no causality" with the test statistics[6] taking the values 12.58 and 36.15 both of which are significant at the 5% level when compared with the critical value of the $\chi_5^2$ distribution of 11.07. The size and pattern of the coefficients suggests that, in terms of daily levels of informationless trading volume, the spot market leads the futures market. In addition, both markets exhibit significant own-market autocorrelation of first and higher orders. The presence of significant cross-market effects (i.e., in the off-diagonal elements of the $B^j$ matrices) indicates that the use of univariate models to divide volume into information-driven and informationless components may be inappropriate.

The fitted values from (5.4), denoted by $V_t^*=(vs_t^*, vf_t^*)'$ are interpreted as the amount of informationless trading in the relevant market. Their graphs are presented in **Figure 5.2.**
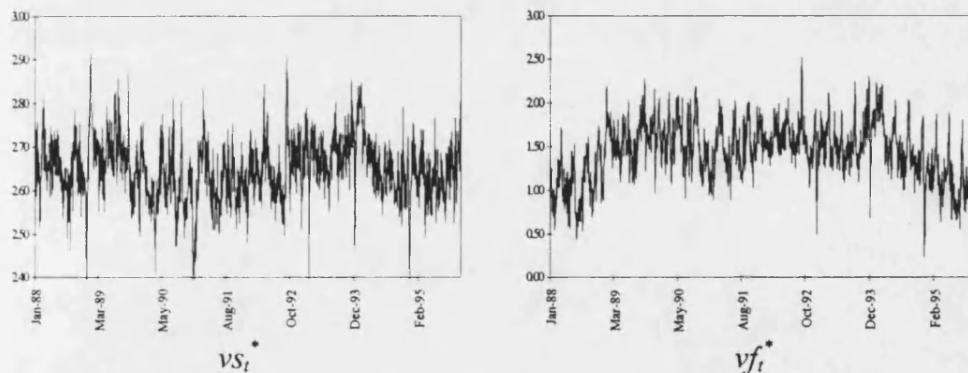


**Figure 5.2:** FTSE100: Informationless spot and futures volume.

The residuals from the model, denoted by $E_t=(e_t^{vs}, e_t^{vf})'$ are interpreted as a measure of information impact. The difference between the fitted values for the futures and spot

[5] It is interesting to observe that at $j=5$ there are significant cross effects between the spot and futures markets. While this is consistent with a lagged weekly cross-relationship, it does not imply a day-of-the-week effect.
[6] The test statistics are computed as $S=T(RSS_0-RSS_1)/RSS_1$ where $T$ is the number of observations, $RSS_0$ is the residual sum of squares from the 5th order autoregression of $x$ including a constant term, and $RSS_1$ is the residual sum of squares from the previous model augmented by 5 lags of $y$.

markets, $vf_t^* - vs_t^*$ gives a measure of informationless futures trading activity in excess of informationless spot market activity.

## 5.4. Does the level of trading in futures influence stock market volatility?

### 5.4.1. Econometric specification

It has been indicated earlier that the basic SV model, extended to include a set of explanatory variables allows for the modelling of the instantaneous impact of exogenous variables on the evolution of (log)variance:

$$\begin{cases} r_t = \sigma_t v_t & v_t \sim N(0,1) \\ \ln \sigma_t^2 = Z_t' \gamma + h_t \\ h_t = \phi h_{t-1} + \eta_t & \eta_t \sim N(0, \sigma_\eta^2) \end{cases} \tag{5.5}$$

Thus the working hypothesis of a contemporaneous effect of futures trading activity on spot market volatility can be tested in this framework by examining the significance of the coefficients $\gamma$.

However, the declining lag structure may also be modelled in the SV framework:

$$\begin{cases} r_t = \sigma_t v_t & v_t \sim N(0,1) \\ \ln \sigma_t^2 = h_t \\ h_t = Z_t' \gamma + \phi h_{t-1} + \eta_t & \eta_t \sim N(0, \sigma_\eta^2) \end{cases} \tag{5.6}$$

The change is accomplished by trivially adjusting the state system matrices (see Appendix 2 for details) without any modifications of the estimation procedure.

### 5.4.2. Estimation results

The SV model (5.5) was estimated with five alternative sets of explanatory volume variables (none, $vs^*$, $vf^*$, $vf^* - vs^*$, $vs^* - vf^*$) representing informationless trading, or

differences in informationless trading[7], and the parameter estimates are presented in **Table 5.4**. Column *M0* in shows the results of the estimation of the basic SV with no explanatory variables, for which the implied coefficient of variation, which measures the strength of the stochastic volatility process, takes the value 0.24. Volatility is found to be persistent, $\phi=0.965$, and the implied annualised unconditional volatility is about 13%. **Figure 5.3** contrasts the plot of the SV volatility estimate (solid line) and the 30 day rolling volatility often used as a crude measure of volatility. Both series are annualised by a factor $\sqrt{260}$.
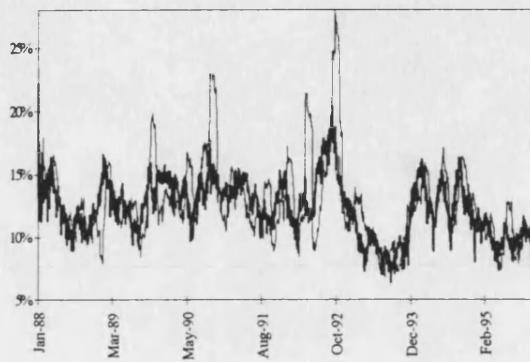


**Figure 5.3:** FTSE100: SV volatility estimate and 30 day rolling volatility.

The second and third models reported in **Table 5.4** - columns *M1* and *M2* respectively - show the effect of including the amount of informationless spot market and futures market activity. In the case of *M1*, contemporaneous spot market volume is seen to have a negative, but insignificant coefficient. The sign of the coefficient is consistent with the hypothesis of increased liquidity provision in the presence of high predictable trading activity. Model *M2* shows that informationless futures volume has no effect on spot volatility. These results suggest that spot market volatility is at best unaffected by informationless volume in either market. The hypothesis is verified by both, the *t*-ratios

---

[7] In addition. dummy variables representing: closed market. Stock Exchange account and expiration dummies were included in the models. Neither the closed market nor expiry effects were significant. Although the Stock Exchange account dummy was weakly positively significant, its definition is not straightforward. in view of the abandonment of the system towards the end of the sample period

156

and the likelihood ratio test statistics being insignificant at the 5% level (with the relevant $\chi^2_1$ 5% and 1% critical values being 3.841 and 6.635 respectively).

Because it is possible that spot volatility is affected by the way in which total volume is split between the spot and futures markets (i.e. the relative volumes), column headed M3 shows estimates of simultaneous effects of spot volume, $vs^*$ and excess futures volume, $vf^*-vs^*$ on spot volatility. The coefficient of primary interest, $vf^*-vs^*$, is insignificant, suggesting that excess futures volume has no effect on spot volatility. Similarly, in column M4 the effect of informationless futures volume is separated from any effect of excess spot market volume (measured as $vs^*-vf^*$). Again, the coefficients are insignificant, confirming the principal result that informationless futures volume has no effect on spot volatility. For completeness, column M5 shows the effect of including both $vf^*$ and $vs^*$ in the model. The results are unaffected. In all cases the likelihood ratio test statistics with M0 as the null hypothesis are insignificant at the 5% level (with the relevant $\chi^2_2$ 5% and 1% critical values being 5.991 and 9.210 respectively).

We now demonstrate that the econometric pitfalls of previous research may lead to the conclusions contrary to those presented here. First, we impose the exponentially declining lag structure on the coefficients of the explanatory variables by including them in the transition equation as in (5.6). This parallels the work of many previous studies in the area. The results of the estimation are presented in **Table 5.5**. The models are labelled M0'-M5' to signify the modification in the specification of the variance equation. The estimates of the coefficients are of the same sign as before but are now significant. For instance, in model M2' the coefficient on $vf^*$ is now significant at 1% level, with the likelihood ratio test being significant at the 5% level. Thus, the inclusion of volume, an apparently straightforward extension of the GARCH model, in

our data would incorrectly imply that futures volume is positively associated with spot market volatility.

Second, we follow Bessembinder and Seguin (1992) by including the residuals, $(e^{vs}, e^{vf})$ as well as the fitted values of (5.4) in the equation of the conditional variance. We revert to the specification of the conditional variance as in (5.5) in order to ensure that the effects of the simultaneity bias and variance specification remain well separated. This leads to the estimation of models $M6$-$M8$, results of which are presented in **Table 5.6**. It has been mentioned previously that such specifications suffer from the simultaneity bias which leads to inconsistency of the model parameters. This point is clearly illustrated in **Table 5.6**. Both information related variables, $(e^{vs}, e^{vf})$ are highly significant. Most importantly, however, the informationless futures volume, $vf$ is now significant as well, lending support to the claim regarding the distortionary effects of futures market trading activity. On the other hand, informationless spot volume, $vf$ remains insignificant and of the same sign.

The final issue is the robustness of the results to the volume decomposition procedure (5.4). To verify this hypothesis the autoregressive matrices, $B^j$ and the error covariance matrix $\Sigma$ are restricted to be diagonal which implies that (5.4) can be estimated by two univariate autoregressions. The fitted values are denoted by $vs^{**}$ and $vf^{**}$. The models $M1$-$M5$ are reestimated and denoted by $M1''$- $M5''$ to signify the fact that alternative informationless volume measures are used. The results, reported in **Table 5.7**, verify the conclusions drawn earlier: futures trading does not increase stock market volatility. The $t$-ratios associated with the coefficients on all explanatory variables, and in particular on $vf^{**}$ are insignificant and the likelihood ratio test is significant in some cases only at the 5% level. The signs of the coefficients remain unchanged.

Summing up, informationless futures trading activity is found to have no effect on stock market volatility. This finding is robust to alternative volume decomposition procedures. However, the conclusion could not be reached when information related volume variables were included in the variance specification. Similarly, opposite results were obtained when the explanatory variables were included so as to mirror models utilised in previous research.

## 5.5. Summary and conclusion

This Chapter has examined the hypothesis that futures trading destabilises the corresponding spot market, leading to an increase in price volatility. It was shown that the use of the GARCH or Schwert (1990) models to test this hypothesis leads to results that are potentially misleading. In particular, the estimated coefficients on volume represent the effect of an exponentially weighted average of past levels of volume, not the instantaneous effect. In addition, there is a need to disaggregate volume into informed and informationless trading in a way that allows for leads and lags between spot and futures volume. The resulting analysis, which used the SV model, found no evidence for the UK to support the hypothesis that futures trading destabilises the spot market. There was also no evidence that spot trading destabilises the spot market, or that an increase in volume in one market, relative to the other, destabilises the spot market. Overall, the results clearly demonstrate that, contrary to some regulatory claims (e.g. NYSE, 1990), futures trading, after adjusting for the effects of information arrival and time trends, does not destabilise the spot market.

## Table 5.1. FTSE returns: summary statistics

| | $m_1$ | $m_2$ | $min/m_2$ | $max/m_2$ | $m_3$ | $m_4$ | $Q_{10}$ | $Q_{40}$ |
|---|---|---|---|---|---|---|---|---|
| $r_t$ | $-3.7 \cdot 10^{-4}$ | $8.2 \cdot 10^{-3}$ | 13.47 | 7.84 | -1.81 | 29.94 | 32.29 | 66.54 |
| $r^2_t$ | $6.7 \cdot 10^{-5}$ | $1.1 \cdot 10^{-4}$ | -0.19 | 33.87 | 25.88 | 786.42 | 1192.8 | 1259.1 |
| $\ln r^2_t$ | -10.68 | 1.75 | 2.01 | -3.77 | -0.25 | 2.52 | 185.05 | 339.71 |

This table reports the sample means ($m_1$), standard deviations ($m_2$), skewness ($m_3$), and kurtosis ($m_4$), as well as the modulus of the minimum (and maximum) deviation from $m_1$ expressed as a multiple of $m_2$ for the returns ($r_t$), their squares ($r_t^2$), and log squares ($\ln r_t^2$). The time period is 4/1/1988-14/12/1995, $T$=2,011. The Box-Ljung statistic

$$Q_p = T(T+2)\sum_{\tau=1}^{p}(T-\tau)\hat{\rho}^2(\tau)$$

is evaluated at 10 and 40 lags, where $\hat{\rho}(\tau)$ is the sample autocorrelation at lag $\tau$.

**Table 5.2. Unit root tests in the stock and futures market volume**

**Panel A.** Augmented Dickey-Fuller tests

| | ln(VS) | ln(VF) |
|---|---|---|
| $\hat{\phi}$ | 0.453 | 0.655 |
| $s.e.(\hat{\phi})$ | (0.020) | (0.017) |
| $\hat{\alpha}$ | 1.104 | -0.011 |
| $s.e.(\hat{\alpha})$ | (0.042) | (0.017) |
| $\hat{\delta}$ | 0.0003 | 0.0005 |
| $s.e.(\hat{\delta})$ | (0.00002) | (0.00003) |
| $ADF_3$ | -1099.0 | -692.2 |
| $ADF_1$ | -27.5 | -20.4 |
| $\hat{\phi}$ | 0.677 | 0.819 |
| $s.e.(\hat{\phi})$ | (0.029) | (0.020) |
| $\hat{\alpha}$ | 0.652 | -0.002 |
| $s.e.(\hat{\alpha})$ | (0.059) | (0.016) |
| $\hat{\delta}$ | 0.0002 | 0.0003 |
| $s.e.(\hat{\delta})$ | (0.00002) | (0.00003) |
| $ADF_2$ | -338.7 | -190.2 |
| $ADF_1$ | -11.2 | -8.9 |

The top half of this panel reports the results of univariate OLS regressions of the form:

$$y_t = \phi y_{t-1} + \alpha + \delta t + \varepsilon_t$$

where $y_t$ is the volume series indicated in the column heading. The second half reports the results from the regressions of the form:

$$y_t = \phi y_{t-1} + \alpha + \delta t + \sum_{i=1}^{p} \psi_i \Delta y_{t-1} + \varepsilon_t$$

where the coefficients $\psi_i$ are not reported. The number of lags, $p=5$, was selected in accordance with the minimised value of the Schwartz information criterion. The statistics $ADF_1$, $ADF_2$, and $ADF_3$ are computed as:

$$ADF_1 = \frac{(\hat{\phi}-1)}{\hat{\sigma}_\phi} \qquad ADF_2 = \frac{T(\hat{\phi}-1)}{1-\sum_{i=1}^{p}\hat{\psi}_i} \qquad ADF_3 = T(\hat{\phi}-1)$$

and are tabulated in Fuller (1976). The 1% critical values for these statistics are, respectively, -3.96, -29.5, -29.5.

**Panel B**: Cointegration test

| Statistic | Ln(VS), Ln(VF) | | |
|---|---|---|---|
| | $\lambda_{trace}$ | $\lambda^1_{max}$ | $\lambda^2_{max}$ |
| $H_0$ | $r=0$ | $r=0$ | $r=1$ |
| Crit Value 1% | 19.310 | 17.936 | 6.936 |
| Crit Value 5% | 15.197 | 14.036 | 3.962 |
| Lag 1 | 338.47 | 295.27 | 43.21 |
| Lag 2 | 232.91 | 206.53 | 26.38 |
| Lag 3 | 160.46 | 141.22 | 19.24 |
| Lag 4 | 105.77 | 91.93 | 13.84 |
| Lag 5 | 102.57 | 90.72 | 11.85 |

This panel reports the maximal eigenvalue and trace statistics:

$$\lambda_{trace} = -T\sum_{i=1}^{2}\ln\left(1-\lambda_i\right) \qquad \lambda^i_{max} = -T\ln\left(1-\lambda_i\right)$$

where $\lambda_i$ are the eigenvalues of the canonical correlation matrix, see Johansen (1989). $\lambda_{trace}$ is a test of $H_0$:$r=0$ against $H_1$:$r=2$, while $\lambda^i_{max}$ is a test of $H_0$:$r=i$ against $H_1$:$r=i+1$ where $r$ is the rank of the long run impact matrix $\Pi$ in the VECM:

$$\Delta Y_t = \alpha + \Pi Y_{t-1} + \sum_{j=1}^{p} B^j \Delta Y_{t-j} + E_t \qquad , \quad E_t \sim N(0,\Sigma)$$

$Y_t = (lnVS_t, lnVF_t)'$, and $\Pi$, $B^j$, and $\Sigma$ are (2×2) matrices. $VS$ and $VF$ respectively represent the spot and futures market volumes.

## Table 5.3. Cross effects between spot and futures volume

| | $vs^\dagger$ | | |
|---|---|---|---|
| | $\hat{\psi}$ | $s.e.(\hat{\psi})$ | $\hat{\psi}/s.e.(\hat{\psi})$ |
| **Const** | 0.8037 | 0.0750 | 10.7090 |
| $vs^\dagger_{t-1}$ | 0.3516 | 0.0248 | 14.1650 |
| $vs^\dagger_{t-2}$ | 0.0487 | 0.0266 | 1.8299 |
| $vs^\dagger_{t-3}$ | 0.0003 | 0.0267 | 0.0140 |
| $vs^\dagger_{t-4}$ | 0.0668 | 0.0266 | 2.5037 |
| $vs^\dagger_{t-5}$ | 0.2300 | 0.0249 | 9.2303 |
| $vf^\dagger_{t-1}$ | -0.0014 | 0.0088 | -0.1547 |
| $vf^\dagger_{t-2}$ | 0.0041 | 0.0097 | 0.4258 |
| $vf^\dagger_{t-3}$ | 0.0195 | 0.0097 | 1.9926 |
| $vf^\dagger_{t-4}$ | 0.0029 | 0.0097 | 0.3018 |
| $vf^\dagger_{t-5}$ | -0.0270 | 0.0088 | -3.0580 |
| | $vf^\dagger$ | | |
| **Const** | 0.8442 | 0.2148 | 3.9297 |
| $vs^\dagger_{t-1}$ | -0.1851 | 0.0710 | -2.6052 |
| $vs^\dagger_{t-2}$ | -0.1669 | 0.0762 | -2.1892 |
| $vs^\dagger_{t-3}$ | -0.1385 | 0.0764 | -1.811 |
| $vs^\dagger_{t-4}$ | 0.0395 | 0.0764 | 0.5178 |
| $vs^\dagger_{t-5}$ | 0.2246 | 0.0713 | 3.1493 |
| $vf^\dagger_{t-1}$ | 0.4755 | 0.0252 | 18.8550 |
| $vf^\dagger_{t-2}$ | 0.1418 | 0.0279 | 5.0785 |
| $vf^\dagger_{t-3}$ | 0.1259 | 0.0280 | 4.4946 |
| $vf^\dagger_{t-4}$ | 0.0505 | 0.0279 | 1.8087 |
| $vf^\dagger_{t-5}$ | 0.0365 | 0.0252 | 1.4480 |

This table reports the estimated coefficients of two VARs of the form:

$$Y_t = \alpha + \sum_{j=1}^{p} B^j Y_{t-j} + E_t, \quad E_t \sim N(0, \Sigma)$$

where $Y_t = (vs^\dagger_t, vf^\dagger_t)'$ is a (2×1) vector. The symbol $\dagger$ symbolises the fact that detrended volume figures are used. The coefficients are grouped to facilitate comparison. The first panel shows the equation for spot market volume, $vs^\dagger$ while the second shows the equation of futures volume, $vf^\dagger$.

# Table 5.4. Effect of volume on spot volatility

| | M0 | M1 | M2 | M3 | M4 | M5 |
|---|---|---|---|---|---|---|
| $\phi$ | 0.965 | 0.961 | 0.965 | 0.959 | 0.959 | 0.959 |
| | (0.0004) | (0.0005) | (0.0005) | (0.0005) | (0.0005) | (0.0005) |
| $\sigma^2_\eta$ | 0.015 | 0.017 | 0.015 | 0.017 | 0.017 | 0.017 |
| | (0.0002) | (0.0003) | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| $\ln \bar{\sigma}^2$ | -9.779 | -7.843 | -9.850 | -7.476 | -7.475 | -7.475 |
| | (0.084) | (1.565) | (0.213) | (1.600) | (1.600) | (1.600) |
| $vs^*$ | | -0.732 | | -0.810 | | -0.942 |
| | | (0.590) | | (0.591) | | (0.626) |
| $vf^*$ | | | 0.051 | | -0.810 | 0.132 |
| | | | (0.137) | | (0.591) | (0.146) |
| $vf^*-vs^*$ | | | | 0.132 | | |
| | | | | (0.146) | | |
| $vs^*-vf^*$ | | | | | -0.942 | |
| | | | | | (0.626) | |
| LogLik | -4136.6 | -4136.0 | -4135.6 | -4133.5 | -4133.5 | -4133.5 |
| LR | | 1.28 | 2.08 | 6.15 | 6.15 | 6.15 |

This table reports the estimation results of the extended SV model (5.5). The explanatory variables, $z^i$ represent informationless trading in spot and futures markets and are measured as the fitted values from (5.4) in the text. The models are labelled M0-M5 depending on the choice of $z^i$. The standard errors are reported below each parameter estimate. These are obtained from the numerical approximation to the Hessian for the hyperparameters $(\phi, \sigma^2_\eta)$ and from the diagonal elements of the state covariance matrix $P_T$ for the remaining coefficients. The $t$-ratio is asymptotically Gaussian and the likelihood ratio test, $LR \sim \chi^2_K$, where $K=1,2$ is the number of explanatory variables in excess of M0.

**Table 5.5. Effect of the model specification on the volume-volatility relationship**

|  | *M0'* | *M1'* | *M2'* | *M3'* | *M4'* | *M5'* |
|---|---|---|---|---|---|---|
| $\phi$ | 0.959 | 0.959 | 0.944 | 0.943 | 0.943 | 0.943 |
|  | (0.0005) | (0.0005) | (0.0009) | (0.0009) | (0.0009) | (0.0009) |
| $\sigma^2_\eta$ | 0.017 | 0.017 | 0.019 | 0.017 | 0.017 | 0.017 |
|  | (0.0003) | (0.0003) | (0.0004) | (0.0004) | (0.0004) | (0.0004) |
| $\ln\bar{\sigma}^2$ | -9.797 | 0.230 | -10.821 | -1.050 | -1.051 | -1.051 |
|  | (0.079) | (0.177) | (0.020) | (0.191) | (0.191) | (0.191) |
| $vs^\bullet$ |  | -0.155* |  | -0.165* |  | -0.213** |
|  |  | (0.067) |  | (0.072) |  | (0.073) |
| $vf^\bullet$ |  |  | 0.040** |  | -0.165* | 0.048** |
|  |  |  | (0.014) |  | (0.072) | (0.014) |
| $vf^\bullet-vs^\bullet$ |  |  |  | 0.048** |  |  |
|  |  |  |  | (0.014) |  |  |
| $vs^\bullet-vf^\bullet$ |  |  |  |  | -0.213** |  |
|  |  |  |  |  | (0.073) |  |
| *LogLik* | -4136.6 | -4136.2 | -4134.3 | -4131.2 | -4131.2 | -4131.2 |
| *LR* |  | 0.88 | 4.68* | 10.75** | 10.75** | 10.75** |

This table reports the estimation results of the extended SV model (5.6) where the explanatory variables enter the transition equation. The standard errors are reported below each parameter estimate. These are obtained from the numerical approximation to the Hessian for the hyperparameters $(\phi, \sigma^2_\eta)$ and from the diagonal elements of the state covariance matrix $P_T$ for the remaining coefficients. Significance of the coefficients at (1%) 5% level is denoted by a (double) star. The $t$-ratio is asymptotically Gaussian and the likelihood ratio test, $LR \sim \chi^2_K$ where $K=1,2$ is the number of explanatory variables in the model specification.

# Table 5.6. Effect of the simultaneity bias on the volume-volatility relationship

| | M6 | M7 | M8 |
|---|---|---|---|
| $\phi$ | 0.985 | 0.993 | 0.996 |
| | (0.0002) | (0.0001) | (0.0001) |
| $\sigma^2_\eta$ | 0.005 | 0.002 | 0.001 |
| | (0.0001) | (0.0000) | (0.0000) |
| $\ln \bar{\sigma}^2$ | -8.845 | -10.765 | -8.317 |
| | (1.478) | (0.224) | (1.459) |
| $vs^*$ | -0.368 | | -0.952 |
| | (0.556) | | (0.568) |
| $vf^*$ | | 0.636** | 0.698** |
| | | (0.131) | (0.141) |
| $e^{vs}$ | 2.829** | | 0.838** |
| | (0.256) | | (0.305) |
| $e^{vf}$ | | 1.489** | 1.327** |
| | | (0.092) | (0.109) |
| **LogLik** | -4059.0 | -3975.5 | -3967.2 |
| **LR** | 155.15** | 322.15** | 338.70** |

This table reports the estimation results of the extended SV model (5.5). The explanatory variables, $z^i$ are $vs^*$, $vf^*$, $e^{vs}$, $e^{vf}$ which represent informationless and information based trading in spot and futures markets and are measured as the fitted values and the residuals from (5.4) in the text. The standard errors are reported below each parameter estimate. These are obtained from the numerical approximation to the Hessian for the hyperparameters $(\phi, \sigma^2_\eta)$ and from the diagonal elements of the state covariance matrix $P_T$ for the remaining coefficients. Significance of the coefficients at (1%) 5% level is denoted by a (double) star. The likelihood ratio test, $LR \sim \chi^2_K$ where $K=2,4$ is the number of explanatory variables in the model specification.

**Table 5.7. Effect of volume decomposition on the volume-volatility relationship**

| | $M1''$ | $M2''$ | $M3''$ | $M4''$ | $M5''$ |
|---|---|---|---|---|---|
| $\phi$ | 0.961 | 0.965 | 0.959 | 0.959 | 0.959 |
| | (0.0005) | (0.0004) | (0.0005) | (0.0005) | (0.0005) |
| $\sigma^2_\eta$ | 0.017 | 0.015 | 0.017 | 0.017 | 0.017 |
| | (0.0003) | (0.0002) | (0.0003) | (0.0003) | (0.0003) |
| $\ln\bar{\sigma}^2$ | -7.531 | -9.808 | -7.145 | -7.144 | -7.144 |
| | (1.585) | (0.215) | (1.640) | (1.640) | (1.640) |
| $vs^{\bullet}$ | -0.850 | | -0.940 | | -1.061 |
| | (0.597) | | (0.603) | | (0.645) |
| $vf^{\bullet}$ | | 0.021 | | -0.940 | 0.121 |
| | | (0.138) | | (0.603) | (0.150) |
| $vf^{\bullet}\text{-}vs^{\bullet}$ | | | 0.121 | | |
| | | | (0.150) | | |
| $vs^{\bullet}\text{-}vf^{\bullet}$ | | | | -1.061 | |
| | | | | (0.645) | |
| **LogLik** | -4135.7 | -4135.5 | -4132.9 | -4132.9 | -4132.9 |
| **LR** | 1.71 | 2.28 | 7.35* | 7.35* | 7.35* |

This table reports the estimation results of the extended SV model (5.5). The explanatory variables, $z^i$: $vs^{**}$, $vf^{**}$, $vf^{**}\text{-}vs^{**}$, $vs^{**}\text{-}vf^{**}$ represent informationless trading in spot and futures markets and are measured as the fitted values from restricted (5.4) in the text. The standard errors are reported below each parameter estimate. These are obtained from the numerical approximation to the Hessian for the hyperparameters $(\phi, \sigma^2_\eta)$ and from the diagonal elements of the state covariance matrix $P_T$ for the remaining coefficients. The $t$-ratio is asymptotically Gaussian and the likelihood ratio test, $LR \sim \chi^2_K$, where $K=1,2$ is the number of explanatory variables in excess of $M0$.

# Conclusion

It is widely documented that the time series dynamics of many financial time series is well described by models which account for the time varying nature of return volatility. While ARCH models predominated the empirical analysis in the past decade, stochastic volatility models have been recently proposed as an alternative way of modelling volatility. Here the volatility is assumed to be driven by a stochastic process of its own, rather than some function of past realisations of the return process.

However, the estimation of SV models is a formidable task because the presence of the latent volatility variable makes the likelihood function difficult to construct. In consequence, comparatively few empirical applications of the SV model can be found in the literature.

The aim of this Thesis is therefore twofold. First a new and efficient estimation method is proposed. Second, the empirical validity of the SV model is verified in two applications, demonstrating that the model is a viable alternative to the ARCH methodology often used in applied empirical research.

The first part of the research agenda is covered in Chapters 2 and 3. Here, the new estimation method is developed. At the core of the procedure is the representation of the model in a linear state space form with non-Gaussian disturbances. It is well known that the Kalman filter can be used to construct the likelihood function but - due to the non-Gaussianity of the measurement equation errors - the estimates thus obtained (QML) will be inefficient. Durbin and Koopman (1997) have shown that the likelihood function of the general non-Gaussian state space model can be approximated arbitrarily accurately by decomposing it into a Gaussian part (constructed by the Kalman filter)

and a remainder function (whose expectation is evaluated by simulation). This general methodology is specialised here to the estimation of SV models.

The final sample performance of the resulting Monte Carlo Maximum Likelihood (MCL) estimator is examined in a simulation study. The results indicate that the performance of the method is comparable to (and often better than) that of the fully efficient Bayesian MCMC estimator. The MCL method owns its computational efficiency and flexibility to the linear state space form which allows powerful algorithms of filtering and smoothing to be utilised. The gain is also due to the fact that the Monte Carlo simulation is only employed to construct that residual part of the likelihood function, which is not already captured by the QML component. In the event, it is shown that only five simulations (draws) are required to achieve finite sample efficiency.

Apart from reducing the computational effort (while attaining full finite sample efficiency), the algorithm has several distinct advantages. First, the sampling variation can be reduced giving arbitrarily close approximations to the true likelihood function. Second, estimation of the SV model with stationary AR(1) dynamics as well as estimation with a nonstationary volatility component is equally feasible. Taken together, the two aspects enable tests for the presence of the unit root in the volatility process to be constructed. This issue is addressed in Chapter 3. It is shown that the augmented Dickey-Fuller (ADF) tests are unreliable. The tests are ether oversized (when the lag of the autoregressive polynomial in is chosen too small) or have low power (when the lag is chosen so as to approximate the correct size). The power declines as the true value of the autoregressive coefficient approaches unity, which is, arguably, the most interesting case from the point of view of applied empirical analysis. It is shown that the likelihood ratio test based upon the estimation of the SV model by

MCL is more powerful. However, the distribution of the likelihood ratio test statistic is unknown. The Monte Carlo evidence presented here suggest that it can be well approximated by the weighted $\chi^2$ density, critical values of which are readily available.

The second part of the Thesis, Chapters 4 and 5, consists of two empirical applications of the SV model employing the new algorithm. First, the hypothesis that implied volatility of options contain relevant information about the evolution of the latent return volatility process is examined. A unique data-set of contemporaneously recorded quotes of the DEM/USD exchange rate and a term structure of implied volatility of over-the-counter (OTC) options is used. We find that *in-sample* the implied volatility captures most of the time series dynamics of the conditional volatility of the return process. In particular, implied volatility of short maturity options are more informative. However, the *out-of-sample* forecasting experiment suggests that the predictions from the basic SV model across all forecasting horizons are at least as accurate as the forecasts obtained from the implied volatility data. This finding augments the conclusions reached in some studies in that implied volatility of options may not be the best forecast of the average realised volatility. A possible explanation for this regularity is the existence of non-negligible volatility risk premia. In this case the inversion of the Black and Scholes formula in the presence of stochastic volatility does not permit the interpretation of implied volatility as an expectation of future historical volatility.

In the second empirical application, Chapter 5, the hypothesis that futures trading destabilises the corresponding spot market, leading to an increase in price volatility is critically examined. It is shown that the way in which ARCH models have been used in

the literature of volume-volatility relationship, leads to results that are potentially misleading. Using the SV model, no evidence for the UK in support of the hypothesis (that futures trading destabilises the spot market) is found. Overall, the results clearly demonstrate that, contrary to some regulatory claims, futures trading, after adjusting for the effects of information arrival and time trends, does not destabilise the spot market.

# Appendix 1: Nonparametric density estimation

A sample, $\{X_1, ..., X_n\}$, is drawn from a continuous univariate distribution with the (unknown) probability density function $f(x) \in C^n$. The kernel estimate is given by:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \qquad \text{(A.1.1)}$$

where the kernel $K(t)$ is a symmetric, non-negative function satisfying

$$\int K(t)dt = 1 \quad \int tK(t)dt = 0 \quad \int t^2 K(t)dt = k_2 \neq 0$$

and $h$ is the smoothing parameter, or bandwidth (Silverman, 1986). The kernel density estimator is thus an average over probability density functions centred at each observation point and converges to $f(x)$ under suitable regularity conditions. Despite the availability of alternative kernels, in this work $K(t)$ is set to the Gaussian density. In our experience the estimated densities are robust with respect to the choice of the kernel.

In some cases the support of the density to be estimated is bounded on one side, e.g. the distribution of $\xi_{LR}$ in Chapter 3 is defined on the positive half line. Because of the observations on (or close to) the boundary, $b$, the kernel method induces some mass to outside of the region on which the "true" density is defined. This may lead to the density estimate not integrating to unity. In these circumstances the density is estimated in the following way. First, values $X_j = b$, $j = 1, .., n_b$ positioned exactly on the boundary are subtracted from the data set. The remaining observations are reflected giving a new data set, $\{X_i, -X_i\}$, $i = 1, .., n$, $i \neq j$, of size $2(n - n_b)$. Finally:

$$\hat{f}(x) = \begin{cases} 2\left(\dfrac{n-n_b}{n}\right)f*(x) & \text{for } x > b \\ \dfrac{n_b}{n} & \text{for } x = b \\ 0 & \text{for } x < b \end{cases}$$

where $f*(x)$ is the density estimate based on the augmented data set.

The bandwidth choice is important but nontrivial. As $h$ tends to zero $\hat{f}(x)$ becomes a sum of Dirac delta functions at the observations, while as $h$ becomes large, all detail in the shape of the estimated density is obscured. Available cross-validation procedures (Bowman *et al*, 1984) are computationally intensive, sensitive to outliers, and sometimes inconsistent (Silverman, 1986). This naturally leads to a subjective choice of $h$. Whenever density estimation is performed for illustrative purposes, e.g. Chapter 2, this option is preferred and values of $h$ are not reported.

When accurate estimates of the density are required, a reference bandwidth, $h_{opt}$, is calculated such that the mean integrated square error is minimised (Parzen, 1962):

$$h_{opt} = k_2^{-2/5}\left\{\int K(t)^2 dt\right\}^{1/5}\left\{\int f''(x)^2 dx\right\}^{-1/5}n^{-1/5}$$

For the Gaussian kernel the expression is reduced, but still depends on the second derivative of the "true" density:

$$h_{opt} = \{4\pi\}^{-1/10}\left\{\int f''(x)^2 dx\right\}^{-1/5}n^{-1/5} \qquad\qquad \text{(A.1.2)}$$

One way obtaining $h_{opt}$ is by reference to a standard distribution (Silverman, 1986). For the likelihood ratio test, Chapter 3, it is natural to take a $\chi_v^2$ density in which case the integrand in (A.1.2) becomes:

$$f_{\chi_\nu^2}''(x)^2 = \frac{x^{\nu-6}\left(8 + 4x + x^2 - 6\nu - 2x\nu + \nu^2\right)^2}{16 \cdot 2^\nu e^x \Gamma\left(\frac{\nu}{2}\right)^2}$$

For $\nu=2$ this dramatically reduces to $f_{\chi_2^2}''(x)^2 = \left(64e^x\right)^{-1}$ leading to $h_{opt}=0.45$ for

$n=1000$. When $\nu=1$, however, the integral

$$\int_z^\infty f_{\chi_1^2}''(x)^2 dx = \int_z^\infty \frac{\left(3 + 2x + x^2\right)^2}{32\pi e^x x^5} dx \qquad (A.1.3)$$

diverges as $z \to 0$, due to the discontinuity at the origin. Numerical approximations

translate into the following values for $h_{opt}$ when the number of observations is $n=1000$:
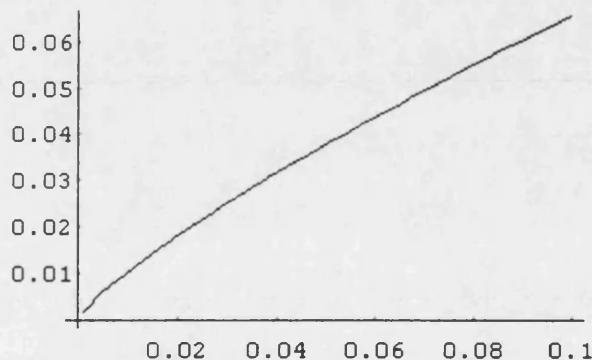


**Figure A1:** Bandwidth, $h_{opt}$ as a function of the lower bound, $z$ in (A.1.3).

# Appendix 2: State space form, filtering and smoothing

*(i)* The general multivariate state space form with $N$ series is:

$$y_t = Z_t \alpha_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, H_t)$$
$$\alpha_t = T_t \alpha_{t-1} + \eta_t \qquad \eta_t \sim N(0, Q_t) \qquad t = 1, \dots, T \qquad \text{(A.2.1)}$$

where $y_t$ is an $(N \times 1)$ vector of observations at time $t$, $\alpha_t$ is the $(m \times 1)$ state vector, and

the covariance matrices $H_t$ $(N \times N)$ and $Q_t$ $(m \times m)$ are non-singular (Harvey, 1989).

The univariate $(N=1)$ SV model with $k$ explanatory variables, $z_t^k$ grouped into

$z_t = (z_{1t}, \dots, z_{kt})'$ as in (1.15) is put into the state space form by defining the matrices

$$\alpha_t = \begin{bmatrix} h_t \\ \gamma_t \end{bmatrix}, \quad T_t = \begin{bmatrix} \phi & e_k' \\ e_k & I_k \end{bmatrix}, \quad Q_t = \begin{bmatrix} \sigma_\eta^2 & e_k' \\ e_k & e_k e_k' \end{bmatrix}, \quad Z_t = \begin{bmatrix} 1 & z_t' \end{bmatrix}, \quad H_t = \frac{\pi^2}{2}$$

where $\gamma_t$ is $(k \times 1)$, $e_k$ is a $(k \times 1)$ vector of zeros, and $I_k$ is the $(k \times k)$ identity matrix. The

basic SV model (1.12′) is obtained a special case by setting $k=1$, $z_t'=1$, $\forall t$ so that the

constant, $\ln \bar{\sigma}^2$ becomes an element of the state vector. When the explanatory

variables are included in the transition equation as in (1.17) only the $T_t$ and $Z_t$

matrices are modified to:

$$T_t = \begin{bmatrix} \phi & z_t' \\ e_k & I_k \end{bmatrix}, \quad Z_t = \begin{bmatrix} 1 & e_k' \end{bmatrix}$$

When AR(2) dynamics are present as in (3.7), $H$ is unchanged, and the remaining

matrices are

$$\alpha_t = \begin{bmatrix} h_t \\ h_{t-1} \\ \gamma_t \end{bmatrix}, \quad T = \begin{bmatrix} \phi_1 & \phi_2 & e_k' \\ 1 & 0 & e_k' \\ e_k & e_k & I_k \end{bmatrix}, \quad Q = \begin{bmatrix} \sigma_\eta^2 & 0 & e_k' \\ 0 & 0 & e_k' \\ e_k & e_k & e_k e_k' \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 0 & z_t' \end{bmatrix}$$

Similarly, the multivariate model (1.19) with explanatory variables is implemented by setting

$$
\alpha_t = \begin{bmatrix} h_t \\ \gamma_t \end{bmatrix}, \quad
T = \begin{bmatrix} \Phi & e_N e_{Nk}{}' \\ e_{Nk} e_N{}' & I_{Nk} \end{bmatrix}, \quad
Q = \begin{bmatrix} \Sigma_\eta & e_N e_{Nk}{}' \\ e_{Nk} e_N{}' & e_{Nk} e_{Nk}{}' \end{bmatrix}, \quad
Z = \begin{bmatrix} 1 & z_t{}' \end{bmatrix} \otimes I_N
$$

where $h_t$ is now $(N \times 1)$, $\gamma_t$ is $(Nk \times 1)$, while $\Phi$ and $\Sigma_\eta$ are $(N \times N)$. The off-diagonal elements of $H$ are given in the text, equation (2.32):

$$
\rho_{i,k}{}^* = \frac{2}{\pi^2} \sum_{j=1}^{\infty} \frac{\Gamma\left(\frac{1}{2}\right)(j-1)!}{\Gamma\left(\frac{1}{2}+j\right)j} \rho_{i,k}^{2j}, \qquad i,j = 1,\ldots,N, \quad j \neq i
$$

where $\rho_{i,k}$ are the off-diagonal elements of the untransformed correlation matrix $\Omega_v$. The diagonal elements remain $\pi^2/2$.

*(ii)* The Kalman filter is given by:

$$
\begin{aligned}
\upsilon_t &= y_t - Z_t a_t & F_t &= Z_t P_t Z_t{}' + H_t \\
K_t &= T_{t+1} P_t Z_t{}' F_t^{-1} & L_t &= T_{t+1} - K_t Z_t \\
a_{t+1} &= T_{t+1} a_t + K_t \upsilon_t & P_{t+1} &= T_{t+1} P_t L_t{}' + Q_{t+1}
\end{aligned}
\qquad \text{(A.2.2)}
$$

for $t = 1, .., T$. The recursions are initialised with the unconditional distribution of the state vector, $\alpha_0 = N(a_0, P_0)$ where $a_0$ is a zero vector and

$$
P_0 = \begin{bmatrix} \tilde{P}_0 & e_{m-d} e_d{}' \\ e_d e_{m-d}{}' & 10^5 I_d \end{bmatrix}
$$

where $d$ is the number of elements in the state vector not related to $h_t$. Letting $\tilde{T}$ and $\tilde{Q}$ to correspond to the state vector elements related to $h_t$, $\tilde{P}_0$ is obtained from

$$
vec(\tilde{P}_0) = \left( I_{2(m-d)} - \tilde{T} \otimes \tilde{T} \right)^{-1} vec(\tilde{Q})
$$

For instance, for the for the basic SV model with AR(1) dynamics and $k$ explanatory variables as in (1.15) this implies

$$P_0 = \begin{bmatrix} \dfrac{\sigma_\eta^2}{(1-\phi^2)} & e_k{}' \\ e_k & 10^5 I_k \end{bmatrix}$$

When the explanatory variables are included as in (1.17) the filter is initialised with the diffuse prior, $\widetilde{P}_0 = 10^5$. Koopman's (1997) method may used for the first $d$ iterations to make the initialisation exact.

*(iii)* The disturbance smoother (de Jong, 1988; Koopman, 1993) is used to construct:

$$\hat{\varepsilon}_t = E(\varepsilon_t|Y) = H_t e_t \qquad C_t = Var(\varepsilon_t|Y) = H_t - H_t D_t H_t$$

where $Y$ is the matrix of all observations, and the quantities $e_t$ and $D_t$ are obtained from the backwards recursions:

$$\begin{aligned} e_t &= F_t^{-1} v_t - K_t{}' r_t & D_t &= F_t^{-1} + K_t{}' N_t K_t \\ r_{t-1} &= Z_t{}' F_t^{-1} v_t + L_t{}' r_t & N_{t-1} &= Z_t{}' F_t^{-1} Z_t + L_t{}' N_t L_t \end{aligned} \qquad \text{(A.2.3)}$$

for $t=T, \ldots, 1$ with $r_T=0$ and $N_T=0$. The prediction errors, $v_t$, their variances, $F_t$, and the Kalman gain matrix, $K_t$ are outputs of the Kalman filter (A.2.2).

*(iv)* The smoothed estimate of the state vector

$$\hat{\alpha}_{t|T} = E(\alpha_t|Y)$$

is constructed by the fixed interval state smoother of de Jong (1988):

$$\hat{\alpha}_{t|T} = \alpha_t + P_t r_{t-1} \qquad \text{(A.2.4)}$$

for $t=T$, ..., 1 where $\alpha_t$ and $P_t$ are taken from the Kalman filter (A.2.2), and $r_t$ is taken from the disturbance smoother (A.2.3).

(v) A special version of de Jong and Shephard's (1995) simulation smoother is used to obtain draws of $\varepsilon^{(i)}$ from $p_G(\varepsilon|\psi)$:

$$\varepsilon_t^{(i)} = H_t \tilde{\varepsilon}_t + u_t^{(i)} \qquad u_t^{(i)} \sim N(0, \tilde{C}_t)$$

where the quantities $\tilde{\varepsilon}_t$ and $\tilde{C}_t$ are obtained from the backwards recursions:

$$\tilde{\varepsilon}_{t_t} = F_t^{-1} v_t - K_t' \tilde{r} \qquad\qquad \tilde{D}_t = F_t^{-1} + K_t' \tilde{N}_t K_t$$

$$M_t = H_t \left( \tilde{D}_t Z_t - K_t' \tilde{N}_t T_{t+1} \right) \qquad \tilde{C}_t = H_t - H_t \tilde{D}_t H_t \qquad\text{(A.2.5)}$$

$$\tilde{r}_{t-1} = Z_t' F_t^{-1} v_t - M_t' \tilde{C}_t^{-1} u_t^{(i)} + L_t' \tilde{r} \quad \tilde{N}_{t-1} = Z_t' F_t^{-1} Z_t + M_t' \tilde{C}_t^{-1} M_t + L_t' \tilde{N}_t L_t$$

for $t=T$, ..., 1 with $\tilde{r}_T = 0$ and $\tilde{N}_T = 0$. Note that when a set of samples is required, the Kalman filter and the recursions for $\tilde{D}_t, M_t, \tilde{C}_t, \tilde{N}_{t-1}$ need only be applied once since these quantities remain the same for each sample.

## Appendix 3: Numerical implementation

*(i)* The Gaussian likelihood function is based on the prediction error decomposition obtained form the Kalman filter (A.2.2):

$$LogL_G(\psi) = -\frac{1}{2}\sum_{t=d}^{T}\left(\ln 2\pi + \ln|F_t| + \upsilon_t{'}F_t^{-1}\upsilon_t\right)$$

(A.3.1)

The hyperparameters are placed into the system matrices $Q_t$ and $T_t$ in (A.2.1). The estimates of the hyperparameters are obtained by numerical maximisation of (A.3.1) while the estimates of the coefficients on the explanatory variables, $\gamma$ are given by the relevant elements of the final state vector, $\alpha_T$. The standard errors of $\phi$ and $\sigma_\eta$ are taken from the numerical approximation to the Hessian.. The estimated variances of $\hat{\gamma}$ are given by the relevant diagonal elements of $P_T$.

*(ii)* The likelihood function is maximised numerically by means of a variable metric optimisation routine like BFGS, details of which can be found in Press *et al.* (1992, pp.425) and Koopman *et al.* (1995, p. 211).

*(iii)* Since the parameters are estimated by numerical methods, restrictions need to be imposed to ensure the stability of the algorithm. This can be accomplished by either invoking some constrained optimisation method (Schoenberg, 1995) or transforming the parameters to a new parameter space so that unconstrained optimisation *(ii)* can be performed. This method is discussed in e.g. Hamilton (1994, p.146).

Let the hyperparameters of the basic SV model (2.1) with stationary AR(1) dynamics be represented by a (2×1) vector $\psi=(\psi_\phi, \psi_\sigma)'$, where in an obvious notation $\psi_\phi=\phi$ and

179

$\psi_\sigma = \sigma_\eta$. Let $\theta$ be another (2×1) vector, partitioned as $\theta = (\theta_\phi, \theta_\sigma)'$. Then the vector function $h(\cdot)$: $\Re \times \Re \to (-1,1) \times \Re^+$ ensures that the restrictions on the parameters are satisfied and is given by:

$$\psi_\phi = \theta_\phi \left(1 + \theta_\phi^2\right)^{-0.5} \qquad \text{(A.3.2a)}$$

$$\psi_\sigma = \exp\left(\theta_\sigma\right) \qquad \text{(A.3.2b)}$$

Alternatively, (A.3.2a) may be replaced by: $\psi_\phi = 2\pi^{-1} \tan\left(\theta_\phi\right)$. The graph of the inverse transformation, $h^{-1}(\cdot)$ is presented in **Figure 3.1** in the text. In the stationary AR(2) case the parameter vector is $\psi = (\phi_1, \phi_2, \sigma_\eta)'$, which may be replaced by $\psi^* = (\lambda_1, \lambda_2, \sigma_\eta)'$ where the roots of the autoregressive polynomial are related to the coefficients via:

$$\phi_1 = \lambda_1 + \lambda_2 \qquad \phi_2 = -\lambda_1 \lambda_2$$

A vector function which ensures that the roots lie within the unit circle and the variance parameter is positive consists of a combination of (A.3.2a) and (A.3.2b).

Finally, a stationary VAR(1) model with $N$ series requires that the covariance matrix of the disturbances, $\Sigma_\eta$ is positive definite, and the roots of the matrix polynomial $I_N - \Phi L$ lie outside the unit circle. The constraints may be imposed by means of an algorithm proposed by Ansley and Kohn (1986).

# Bibliography

Amin, K.I. (1993), "Jump Diffusion Option Valuation in Discrete Time," *Journal of Finance, 48*, 1833-63.

Andersen, T.G. (1994a), "Stochastic Autoregressive Volatility: A Framework for Volatility Modeling," *Mathematical Finance, 4*, 75-102.

Andersen, T. G. (1994b), "Comment on Jacquier, Polson and Rossi," *Journal of Business and Economic Statistics*, 12, 389-392.

Andersen, T.G. (1996), "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility," *Journal of Finance, 51*, 169-204.

Andersen, T. G. and B. Sørensen (1996), "GMM Estimation of a Stochastic Volatility Model: a Monte Carlo Study," *Journal of Business and Economic Statistics, 14*, 328-52.

Ansley, C.F. and R. Kohn (1986), "A Note on Reparameterising a Vector Autoregressive Moving Average Model to Enforce Stationarity," *Journal of Statistical Computation and Simulation, 24*, 99-106.

Bachelier, L.J.B.A (1900), "Theorie de la Speculation," reprinted in P.H. Cootner (ed.), *The Random Character of Stock Market*. Cambridge, Mass.: MIT Press.

Bailey, R.W. (1994), "Polar generation of random variates with the t-distribution," *Mathematics of Computation, 62*, 779-81.

Baillie, R.T. and T. Bollerslev (1989), "The Message in Daily Exchange Rates: A Conditional Variance Tale," *Journal of Business and Economic Statistics, 7*, 297-305.

Baillie, R.T. and T. Bollerslev (1990), "A Multivariate Generalised ARCH Approach to Modelling Risk Premia in Forward Foreign Exchange Rate Markets," *Journal of International Money and Finance, 16,* 109-124.

Ball, C. A. and A. Roma (1994), "Stochastic Volatility Option Pricing," *Journal of Financial and Quantitative Analysis, 29,* 589-607.

Ball, C.A. and W. Torous (1983), "A Simplified Jump Process for Common Stock Returns," *Journal of Financial and Quantitative Analysis, 18,* 53-65.

Bartholomew, D.J. (1987), *Latent Variable Models and Factor Analysis,* London: Griffin.

Basilevski, A. (1994) *Statistical Factor Analysis and Related Methods,* New York: John Wiley & Sons.

Bates, D.S. (1996), "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," *Journal of Financial Studies, 9,* 69-107.

Bennett, B.M. (1955), "Note on the Moments of the Logarithmic Noncentral $\chi^2$ and z Distributions," *Annals of the Institute of Statistical Mathematics, Tokyo, 7,* 57-61.

Bera, A.K. and M.L. Higgins (1993), "ARCH Models: Properties, Estimation and Testing," *Journal of Economic Surveys, 7,* 305-66.

Berndt, E.K, B.H. Hall, R.E. Hall and J. Haussman (1974), "Estimation and Inference in Nonlinear Structural Models, *Annals of Economic and Social Measurement, 4,* 653-65.

Bessembinder, H. (1994), "Bid-Ask Spreads in the Interbank Foreign Exchange Markets," *Journal of Financial Economics, 35,* 317-48.

Bessembinder, H. and P.J. Seguin (1992), "Futures Trading Activity and Stock Price Volatility," *Journal of Finance, 47,* 2015-34.

Bessembinder, H. and P.J. Seguin (1993), "Price Volatility, Trading Volume and Market Depth: Evidence from Futures Markets," *Journal of Financial and Quantitative Analysis, 28,* 21-39.

Billio, M. and A. Monfort (1995), "Switching State Models: Likelihood Function, Filtering and Smoothing," mimeo, CREST-INSEE.

Black, F. (1976), "Studies in Stock Volatility Changes," *Proceedings of the 1976 Meetings of the Business and Economic Statistics Section, American Statistical Association,* 177-81.

Black, F. and M.S. Scholes (1973), "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy, 81,* 637-54.

Blattberg, R.C. and N. Godenes (1974), "A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices," *Journal of Business, 47,* 244-280.

Board, J.L. and C.M.S. Sutcliffe (1995), "The Relative Volatility of the Markets in Equities and Index Futures," *Journal of Business Finance and Accounting, 22,* 201-23.

Bollerslev, T. (1986), "Generalised Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics, 31,* 307-27.

Bollerslev, T. (1987), "A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics, 69,* 542-547.

Bollerslev, T. (1990), "Modelling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *Review of Economics and Statistics, 72,* 498-505.

Bollerslev, T., R.Y. Chow and K.F. Kroner (1992), "ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics, 52*, 5-59.

Bollerslev, T., R.F. Engle and D.B. Nelson (1994), "ARCH Models" in R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics, Vol. IV*, Amsterdam: Noth-Hollnd.

Bowman, A.W., P.Hall and D.M. Titterington (1984), "Cross-validation in Nonparametric Estimation of Probabilities and Probability Densities," *Biometrika, 71*, 341-51.

Breidt, F.J. and A.L. Carriquiry (1996), "Improved Quasi-Maximum Likelihood Estimation for Stochastic Volatility Models," *Modelling and Prediction: Honouring Seymour Geisel*, A. Zellner and J.S. Lee (eds). New York: Springer Verlag.

Brown-Hruska, S. and G. Kuserk (1995),"Volatility, Volume and the Notion of Balance in the S&P500 Cash and Futures Markets," *Journal of Futures Markets, 15*, 677-689.

Campa, J.M. and P.H.K. Chang (1995), "Testing the Expectations Hypothesis on the Term Structure of Volatilities in Foreign Exchange Market," *Journal of Finance, 50*, 529-47.

Canina, L. and S. Figlewski (1993), "The Informational Content of Implied Volatility," *Review of Financial Studies, 6*, 659-81.

Canova, F. (1997), "Detrending and Business Cycle Facts," *Journal of Monetary Economics*, forthcoming.

Chatrath, A., S. Ranchander and F. Song (1996), "The Role of Futures Trading Activity in Exchange Volatility," *Journal of Futures Markets, 16*, 561-584.

Chen. R.R. and L. Scott (1993), "Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests from the Kalman Filter Model," mimeo, University of Georgia.

Chesney, M. and L. Scott (1989), "Pricing European Currency Options: A Comparison of the Modified Black-Scholes Model with a Random Variance Model," *Journal of Financial and quantitative Analysis,* 24, 267-84.

Christie, A.A. (1982), "The Stochastic Behaviour of Common Stock Variances: Value, Leverage and Interest Rate Effects," *Journal of Financial Economics,* 10, 407-432.

Clark, P.K. (1973), " A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices," *Econometrica, 41,* 135-55.

Cox, J.C. and M. Rubinstein (1985), *Options Markets,* New Jersey: Prentice-Hall.

Danielsson, J. (1994a), "Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood," *Journal of Econometrics,* 61, 375-400.

Danielsson, J. (1994b), "Comment on Jacquier, Polson and Rossi," *Journal of Business and Economic Statistics,* 12, 389-392.

Danielsson, J. and J.F. Richard (1993), "Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models," *Journal of Applied Econometrics, 8,* S153-S174.

Darrat, A.F. and S. Rahman (1995), "Has Futures Trading Activity Caused Stock Price Volatility?", *Journal of Futures Markets, 15,* 537-557.

Dassios, A. (1992), "Asymptotic Approximations to Stochastic Variance Models," mimeo, London School of Economics.

Day, T.E. and C.M. Lewis (1991), "Stock Market Volatility and the Information Content of Stock Index Options," *Journal of Econometrics, 52,* 265-87.

de Jong, F. (1996), "Time Series and Cross-Section Information in the Term Structure Models," mimeo, Tilburg University.

de Jong, P. (1988), "A Cross-Validation Filter for Time Series Models," *Biometrika, 75,* 594-600.

de Jong, P. and N. Shephard (1995), "The Simulation Smoother for Time Series Models," *Biometrika, 82,* 339-50.

Derman, E. and I. Kani (1994), "Riding on a Smile," *Risk, 7,* 32-39.

Dickey, D.A. and W.A. Fuller (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica, 49,* 1057-72.

Diebold, F.X. and J. Nason (1990), "Nonparametric Exchange Rate Prediction?," *Journal of International Economics, 28,* 315-32.

Dumas, B., J. Fleming and R.E. Whaley (1995), "Implied Volatility Functions: Empirical Tests," mimeo, Duke University.

Dunsmuir, W. (1979), "A Central Limit Theorem for Parameter Estimation in Stationary Time Series and Its Applications to Models for a Signal Observed with Noise," *Annals of Statistics, 7,* 490-506.

Dupire, B. (1994), "Pricing with a Smile," *Risk, 7,* 18-20.

Durbin, J. and S.J. Koopman (1997a), "Monte Carlo Maximum Likelihood Estimation for non-Gaussian State Space Models," *Biometrika,* forthcoming.

Durbin, J. and S.J. Koopman (1997b), "Time Series Analysis for Non-Gaussian Observations Based on State Space Models," mimeo, London School of Economics.

Easley, D. and M. O'Hara (1992), "Time and the process of security price adjustment," *Journal of Finance, 47*, 577-605.

Edwards, F.R. (1988), "Futures Trading and Cash Market Volatility: Stock Index and Interest Rate Futures," *Journal of Futures Markets, 8*, 421-439.

Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the UK Inflation," *Econometrica, 50*, 987-1007.

Engle, R.F. (1984), "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," in: Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics, Vol. 2*. Amsterdam: North-Holland.

Engle, R.F. (1990), "Discussion [of "Stock Volatility and the Crash of '87", by G.W. Schwert]", *Review of Financial Studies, 3*, 103-6.

Engle, R.F. and T. Bollerslev (1986), "Modelling the Persistence of Conditional Variances," *Econometric Reviews, 5*, 1-50.

Engle, R.F and K.F. Kroner (1995), "Multivariate Simultaneous Generalised ARCH," *Econometric Theory, 11*, 122-150.

Engle, R.F. and G.G.J. Lee (1992), "A Permanent and Transitory Component Model of Stock Return Volatility," mimeo, University of California.

Engle, R.F. and G.G.J. Lee (1994), "Estimating Diffusion Models of Stochastic Volatility," mimeo, University of California.

Engle, R.F., D.M. Lilien and R.P. Robins (1987) "Estimating Time Varying Risk Premium in the Term Structure: the ARCH-M Model," *Econometrica, 55*, 391-407.

Engle, R.F., V.K. Ng and M. Rothschild (1990) "Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills," *Journal of Econometrics*, *45*, 213-238.

Engle, R.F. and M. Watson (1981), "A One-Factor Multivariate Model of Metropolitan Wage Rates," *Journal of the American Statistical Association*, *76*, 774-81.

Everitt, B.S. (1994), *An Introduction to Latent Variable Models*, New York: Chapman and Hall.

Fair, R.C. and R.J. Shiller (1990), "Comparing Information in Forecasts from Econometric Models," *American Economic Review*, *80*, 375-89.

Fama, E.F. (1965), "The Behaviour of Stock Market Prices," *Journal of Business*, *38*, 34-105.

Fama, E.F. (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance*, *25*, 383-417.

Fama, E.F. (1991), "Efficient Capital Markets II," *Journal of Finance*, *46*, 1575-1617.

Fama, E.F. and R. Roll (1971), "Parameter Estimates for Symmetric Stable Distributions," *Journal of the American Statistical Association*, *66*, 331-338.

Feller, W. (1971), *An Introduction to Probability Theory and its Applications*, New York: Wiley.

Foster, A.J. (1995), "Volume-Volatility Relationships for Crude Oil Futures Markets," *Journal of Futures Markets*, *15*, 929-52.

Friedman, M. and L. Harris (1996), "A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models," mimeo, University of Southern California.

Fuller, W.A. (1976), *Introduction to Statistical Time Series*. New York: Wiley.

Gallant, A.R., D.A. Hsieh and G. Tauchen (1991), "On Fitting a Recalcitrant Series: the Pound/Dollar Exchange Rate, 1974-1983," in W.A. Barnett, J. Powell and G. Tauchen (eds.) *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge: Cambridge University Press.

Gallant, A.R., P.E. Rossi and G. Tauchen (1992), "Stock Prices and Volume," *Review of Financial Studies*, 5, 199-242.

Garbade, K. (1977), "Two Methods for Examining the Stability of Regression Coefficients," *Journal of American Statistical Association*, 72, 54-63.

Ghysels, E. and J. Jasiak (1996), "Stochastic Volatility and Time Deformation: An Application to Trading Volume and Leverage Effects," mimeo, Univerité de Montréal.

Ghysels, E., A.C. Harvey and E. Renault (1996), "Stochastic Volatility," in C.R. Rao and G.S. Maddala (eds.), *Statistical Methods in Finance*. Amsterdam: North-Holland.

Glosten, L.R. and P.R. Milgrom (1985), "Bid-Ask, Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, *14*, 71-100.

Godfrey, L. G. (1988), *Misspecification Tests in Econometrics*, Econometric Society monographs, 16, Cambridge: Cambridge University Press.

Gouriéroux, C., A. Holly and A. Monfort (1982), "Likelihood Ratio Test, Wald Test and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters," *Econometrica*, *50*, 63-80.

Gouriéroux, C. and A. Monfort (1995), *Statistics and Econometric Models, Vol. 2*, Cambridge: Cambridge University Press.

Gouriéroux, C. and A. Monfort (1995), *Simulation Based Econometric Methods*, Cambridge: Cambridge University Press.

Hamilton, J.D. (1994), *Time Series Analysis*, Princeton: Princeton University Press.

Hamilton, J.D. and R. Susmel (1994), "Autoregressive Conditional Heteroscedasticity and Changes in Regime," *Journal of Econometrics, 64*, 307-33.

Hansson, B. and P. Hordahl (1996), "Forecasting Variance Using Stochastic Volatility and GARCH," mimeo, Lund University.

Harris, L. (1989), "S&P 500 Cash Stock Price Volatilities," *Journal of Finance, 44*, 1155-75.

Harvey, A.C. (1976), "Estimating Regression Models with Multiplicative Heteroskedasticity," *Econometrica, 44*, 461-5.

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

Harvey, A.C. (1993), *Time Series Models.* Cambridge Mass.: MIT Press, 2$^{nd}$ Ed.

Harvey, A.C. and N. Shephard (1993), "Estimation and Testing of Stochastic Variance Models," mimeo, London School of Economics.

Harvey, A.C. and N. Shephard (1996), "Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns," *Journal of Business and Economic Statistics, 14*, 429-34.

Harvey, A.C. and M. Streibel (1997), "Testing for a Slowly Changing Level with Special Reference to Stochastic Volatility," *Journal of Econometrics*, forthcoming.

Harvey, A.C., E. Ruiz and N. Shephard (1994), "Multivariate Stochastic Variance Models," *Review of Economic Studies, 61*, 247-64.

Heston, S. (1993), "A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies, 6*, 327-43.

Heynen, R. (1994), "An Empirical Investigation of Observed Smile Patterns," *Review of Futures Markets, 13*, 317-54.

Heynen, R., A.G.Z. Kemma and T. Vorst (1994),"Analysis of the Term Structure of Implied Volatility," *Journal of Financial and Quantitative Analysis, 29*, 31-56.

Hull, J. (1993), *Options, Futures and Other Derivative Securities*, New Jersey: Prentice-Hall, 2$^{nd}$ Ed.

Hull, J. and A. White (1987), "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance, 42*, 281-300.

Jackwerth, J.C. and M. Rubinstein (1996), "Recovering Probability Distributions from Option Prices," *Journal of Finance, 51*, 1611-31.

Jacquier, E., N.G. Polson and P.E. Rossi (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics, 12*, 371-417 (with discussion)

Jacquier, E., N.G. Polson and P.E. Rossi (1995), "Stochastic Volatility: Univariate and Multivariate Extensions," R.L. White Centre for Financial Research, DP #19-95. Wharton School, University of Pennsylvania.

Johansen, S. (1988), "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control, 12*, 231-54.

Johnson, N.L. and S. Kotz (1970), *Continuos univariate distributions-2*, Chichester: John Wiley.

Johnson, N.L. and S. Kotz (1972), *Continuos multivariate distributions*, Chichester: John Wiley.

Johnson, H. and D. Shanno (1987), "Option Pricing when the Variance is Changing," *Journal of Financial and Quantitative Analysis, 22*, 143-51.

Jones, C.M., G. Kaul and M.L. Lipson (1994), "Transactions Volume and Volatility," *Review of Financial Studies, 7*, 631-651.

Jorion, P. (1995), "Predicting Volatility in the Foreign Exchange Market," *Journal of Finance, 50*, 507-28.

Karlin, S. and H.M. Taylor (1981), *A Second Course in Stochastic Processes*, London: Academic Press.

Kim, C.J. (1993), "Unobserved Component Time Series Models with Markov Switching Heteroskedasticity: Changes in Regime and the Link Between Inflation Rates and Inflation Uncertainty," *Journal of Business and Economic Statistics, 11*, 341-349.

Kim, D. and S.J. Kon (1994), "Alternative Models for the Conditional Heteroskedasticity of Stock Returns," *Journal of Business, 67*, 563-98.

Kim, K. and P. Schmidt (1993), "Unit Root Tests with Conditional Heteroskedasticity," *Journal of Econometrics, 59*, 287-300.

Kim, S., N. Shephard and S. Chib (1996), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH models," mimeo, Nuffield College, Oxford.

Kitagawa, G. (1987), "Non-Gaussian State Space Modelling of Nonstationary Time Series," *Journal of the American Statistical Association, 82*, 1032-63 (with discussion)

Kon, S.J. (1984), "Models of Stock Returns: A Comparison," *Journal of Finance, 39*, 147-65.

Koopman, S.J. (1993), "Disturbance Smoother for State Space Models," *Biometrika, 80*, 117-26.

Koopman, S.J. (1997), "Exact Initial Kalman Filter and Smoother for Non-Stationary Time Series Models," *Journal of American Statistical Association*, forthcoming.

Koopman, S.J., A.C. Harvey, J.A. Doornik and N, Shephard (1995), *STAMP 5.0, Structural Time Series Analyser, Modeller and Predictor*. London: Chapman and Hall.

Kremers, J.J., N.R. Ericsson and J.J. Dolado (1992), "The Power of Cointegration Tests," *Oxford Bulletin of Economics and Statistics, 54*, 325-48.

Kwiatowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992), "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We that Economic Time Series Have a Unit Root," *Journal of Econometrics, 44*, 159-78.

Lamoureux, C.G. and W.D. Lastrapes (1990), "Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects," *Journal of Finance, 45*, 221-29.

Lamoureux, C.G. and W.D. Lastrapes (1993), "Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities," *Review of Financial Studies, 6*, 293-326.

Lamoureux, C.G. and W.D. Lastrapes (1994), "Endogenous Trading Volume and Momentum in Stock-Return Volatility," *Journal of Business and Economic Statistics*, *12*, 253-60.

Latane. H.A. and R.J. Rendelman (1976), "Standard Deviations of Stock Price Ratios Implied from Option Prices," *Journal of Finance, 31*, 369-81.

Lee, S.W. and B.E. Hansen (1994), "Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator," *Econometric Theory, 10*, 29-52.

Lee, T.-H. and Y. Tse (1996), "Cointegration Tests with Conditional Heteroskedasticity," *Journal of Econometrics, 73*, 401-10.

Liesenfeld, R. (1996), "Dynamic Bivariate Mixture Models: Modelling the Behaviour of Prices and Trading Volume," discussion paper N#78, University of Tübingen.

Lin, W.L. (1992), "Alternative Estimators for Factor GARCH models - a Monte Carlo Comparison," *Journal of Applied Econometrics, 7*, 259-279.

Lo, A.W. and A.C. MacKinley (1988), "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test," *Review of Financial Studies*, 1, 41-66.

Madan, D.B. and E. Seneta (1990), "The Variance Gamma (V.G.) Model for Share Market Returns," *Journal of Business, 63*, 511-524.

Malz, A.M. (1996), "Option-Based Estimates of the Probability Distribution of Exchange Rates and Currency Excess Returns," mimeo, Federal Reserve Bank of New York.

Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices," *Journal of Business, 36*, 394-419.

Mandelbrot, B. (1967), "The Variation of Some Other Speculative Prices," *Journal of Business, 40*, 393-413.

Mandelbrot, B. and H. Taylor (1967), "In the Distribution of Stock Prices Differences," *Operations Research, 15*, 1057-1062.

Mankiw, N.G. and M.D. Shapiro (1986), "Do We Reject too Often?," *Economics Letters, 20*, 139-145.

Melino, A. and S.M. Turnbull (1990), "Pricing Foreign Currency Options with Stochastic Volatility," *Journal of Econometrics, 45*, 239-65.

Merton, R.C. (1976), "Option Pricing when Underlying Stock Returns are Discontinuous," *Journal of Financial Economics, 3*, 125-144.

Naik, V. (1993), "Option Valuation and Hedging Strategies with Jumps in the Volatility of the Asset Returns," *Journal of Finance, 48*, 1969-84.

Najand, M, and K. Yung (1991), "A GARCH Examination of the Relationship Between Volume and Price Variability in Futures Markets, *Journal of Futures Markets, 11*, 613-621.

Nelson, D.B. (1990a), "ARCH Models as Diffusion Approximations," *Journal of Econometrics, 45*, 7-38.

Nelson, D.B. (1990b), "Stationarity and persistence in the GARCH(1,1) Model," *Econometric Theory, 6*, 318-334.

Nelson, D.B. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica, 59*, 347-70.

Nelson, D.B. and D.P. Foster (1994), "Asymptotic Filtering Theory for Univariate ARCH Models," *Econometrica, 62*, 1-41.

New York Stock Exchange (1990), "Market Volatility and Investor Confidence, Report to the Board of Directors of the NYSE," *New York Stock Exchange*, June.

Noh, J., R.F. Engle and A. Kane (1994), " Forecasting Volatility and Option Prices of the S&P 500 Index," *Journal of Derivatives, 1*, 17-30.

Nyblom, J. and T. Makelainen (1983), "Comparison of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model," *Journal of the American Statistical Association, 78*, 856-64.

Pagan, A.R. (1996), "The Econometrics of Financial Markets," Journal of Empirical Finance, 3, 15-102.

Pagan, A.R. and G.W. Schwert (1990), "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics, 45*, 267-90.

Pantula, S.G. (1991), "Asymptotic Distributions of Unit Root Tests When the Process is Nearly Stationary," *Journal of Business and Economic Statistics, 9*, 63-71.

Parzen, E. (1962), "On Estimation of a Probability Density Function and Mode," *Ann. Math. Statist., 33*, 1065-76.

Phillips, P.C.B. and P. Perron (1988), "Testing for a Unit Root in Time Series Regression," *Biometrika, 75*, 335-46.

Press, S.J. (1967), "A Compound Events Model for Security Prices, " *Journal of Business, 40*, 317-35.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling (1992). *Numerical Recipes in C.* Cambridge: Cambridge University Press, 2$^{nd}$ Ed.

Renault, E. and N. Touzi (1996), "Option hedging and Implied Volatilities in a Stochastic Volatility Model," *Mathematical Finance, 6*, 279-302.

Ripley, B.D. (1987), *Stochastic Simulation*, New York: Wiley.

Rockinger, M. (1994), "Regime Switching: Evidence for the French Stock Market," mimeo, HEC.

Rubinstein, M. (1985), "Nonparametric Tests of Alternative Option Pricing Models Using All Reported Trades and Quotes on the 30 Most Active CBOE Option Classes from August 23, 1976 through August 3, 1978," *Journal of Finance, 40*, 455-80.

Rubinstein, M. (1994), "Implied Binomial Trees," *Journal of Finance, 49*, 771-818.

Ruiz, E. (1994), "Quasi-maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Econometrics, 63*, 289-306.

Said, S.E. and D.A. Dickey (1984), "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika, 71*, 599-607.

Santoni, G.J. (1987), "Has Programmed Trading Made Stock Prices More Volatile?," *Federal Reserve Bank of St. Louis Review*, May, 18-29.

Schmalensee, R. and R.R. Trippi (1978), "Common Stock Volatility Expectations Implied by Option Premia, "*Journal of Finance, 33*, 129-47.

Schoenberg, R. (1995), *The GAUSS Constrained Optimisation Application*, Aptech Systems, available on: ftp://ftp.u.washington.edu/public/rons.

Schwert, G.W. (1989), "Tests for Unit Roots, A Monte Carlo Investigation," *Journal of Business and Economic Statistics, 7*, 147-60.

Schwert, G.W. (1990), "Stock Volatility and the Crash of '87," *Review of Financial Studies, 3*, 77-102.

Scott, L. (1987), "Option Pricing when the Variance Changes Randomly: Theory, Estimation and an Application," *Journal of Financial and Quantitative Analysis*, 22, 419-38.

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*, London, Chapman and Hall.

Shephard, N. and M.K. Pitt (1997), "Likelihood Analysis of Non-Gaussian Parameter-Driven Models," *Biometrika*, forthcoming.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Silvey, S.D. (1975), *Statistical Inference*, London: Chapman and Hall.

Smith, C.W. (1989), "Market Volatility: Causes and Consequences, *Cornell Law Review*, 74, 953-962.

Stein, E.M. and J.C. Stein (1991), "Stock Price Distributions with Stochastic Volatility: An Analytic Approach," *Review of Financial Studies*, 4, 727-52.

Stein, J.C. (1989), "Overreactions in the Options Market," *Journal of Finance*, 44, 1011-23.

Stock, J.H. (1988), "Estimating Continuos Time Processes Subject to Time Deformation," *Journal of the Statistical Association*, 83, 77-84.

Stock, J.H. (1993), "Unit Roots and Trend Breaks," in: R.F. Engle and D. McFadden, eds., *Handbook of Econometrics, Vol. 4*, Amsterdam: North-Holland.

Stoll, H.R. and R.E. Whaley (1990), "The Dynamics of Stock Index and Stock Index Futures Returns," *Journal of Financial and Quantitative Analysis*, 25, 441-68.

Sultan, J., K. Hogan and K.F. Kroner (1995), "The Effects of Programme Trading on Market Volatility: New Evidence," in: *New Directions in Finance* edited by D.K. Ghosh and S. Khaksari, Routledge, London.

Sutcliffe, C.M.S. (1997), *Stock Index Futures: Theories and International Evidence*, London: International Thomson Business Press, 2<sup>nd</sup> Ed.

Tanaka, K. (1983), "Non-Normality of the Lagrange Multiplier Statistic for Testing the Constancy of Regression Coefficients," *Econometrica, 51*, 1577-83.

Tauchen, G.E. and M. Pitts (1983), "The Price Variability Volume Relationship on Speculative Markets," *Econometrica, 51*, 485-505.

Taylor, S.J. (1986), *Modelling Financial Time Series*, Chichester: John Wiley.

Taylor, S.J. (1994), "Modelling Stochastic Volatility: a Review and Comparative Study," *Mathematical Finance, 4*, 183-204.

Teräsvirta, T. (1996), "Two Stylized Facts and the GARCH(1,1) model," mimeo, Stockholm School of Economics.

Tucker, A.L., J. Madula and J.F. Marshall (1994), "Pricing Currency Futures Options with Lognormally Distributed Jumps," *Journal of Business Finance and Accounting, 21*, 857-873.

Watson, M.W. and R.F. Engle (1985), "Testing for Regression Coefficient Stability with a Stationary AR(1) Alternative," *Review of Economics and Statistics*, 341-346

Weiss, A.A. (1986), "Asymptotic Theory for ARCH models: Estimation and Testing," *Econometric Theory, 2,* 107-131.

West, K.D. and D. Cho (1995), "The Predictive Ability of Several Models of Exchange Rate Volatility," *Journal of Econometrics, 69*, 367-91.

Willard, G.A. (1996), "Calculating Prices and Sensitivities for Path-Dependent Derivative Securities in Multifactor Models," mimeo, Washington University in Saint Louis.

Wiggins, J.B. (1987), "Option Values under Stochastic Volatilities," *Journal of Financial Economics, 19,* 351-72.

Xu, X. and S.J. Taylor (1994), "The Term Structure of Volatility Implied by Foreign Exchange Options", *Journal of Financial and Quantitative Analysis, 29,* 57-74.

Xu, X. and S.J. Taylor (1995), "Conditional Volatility and the Informational Efficiency of the PHLX options market", *Journal of Financial and Quantitative Analysis, 29,* 57-74.

Yang, S.R. and B.W. Brorsen (1993), "Nonlinear Dynamics of Daily Futures Prices: Conditional Heteroskedasticity or Chaos?," *Journal of Futures Markets,* 13, 175-191.