Estimation and Inference with the Efficient Method of Moments: With Applications to Stochastic Volatility Models and Option Pricing

van der Sluis, P.J.

Citation for published version (APA):
van der Sluis, P. J. (1999). Estimation and Inference with the Efficient Method of Moments: With Applications to Stochastic Volatility Models and Option Pricing Amsterdam: Thela Thesis. TI Research Series nr. 204
Chapter 2

Analysis of Financial Time Series

Successively, this chapter provides a brief review of the characteristics of financial time series, an introduction of models for financial time series, a review of recent (simulation-based) estimation techniques for these models, and a concise overview of option pricing. Recent surveys that cover more or less the same issues are Shephard (1996a) and Ghysels, Harvey and Renault (1996). For a broader introduction to the study of financial time series and empirical finance in general see Campbell, Lo and MacKinlay (1997). For a broad review of simulation-based estimation see Gouriéroux and Monfort (1996). A good introduction to option pricing is Hull (1997).

The outline of this chapter is as follows. Section 2.1 briefly discusses the main characteristics of financial time series. Section 2.2 introduces models for volatility with emphasis on SV models. Section 2.3 positions the efficient method of moments estimation technique in the (simulation-based) estimation literature. This introductory chapter concludes with a short review of option pricing theory in Section 2.4.

2.1 Characteristics of Financial Time Series

Various types of financial data, such as time series of daily stock returns and daily exchange-rates movements, display similar features. This section successively discusses which particular type of time-series data are considered in this thesis, why such historical data are relevant in the light of modern investment theory and which are the most prominent features of such data that our models should capture.
2.1.1 Nature of the Data

Financial data are often analysed at the daily frequency, although recently also higher frequencies are considered. Usually closing prices are considered in the time-series literature. There is an extensive literature on issues of market microstructure such as closing prices versus opening prices and the frequency of the data. This literature is reviewed in Goodhart and O'Hara (1997). Some ideas on the choice of the sample interval can be found in Campbell et al. (1997, pp. 364-366). It is well beyond the scope of this thesis to deal with these issues. In this thesis we adhere to the standard practice in the econometrics literature on the estimation of SV models, i.e. to consider the daily frequency. Closing prices are considered throughout this thesis except for Chapter 6 where in order to avoid the problem of non-synchronous trading data sets are considered that have been recorded in a different manner. We postpone discussion of this issue to Chapter 6.

In the literature price changes have been analysed in different forms like percentage changes and compounded returns; see Campbell et al. (1997, Section 1.4). In this thesis we will work with continuously compounded percentage returns, i.e.

\[ y_t = 100[\ln(x_t + d_t) - \ln x_{t-1}] \]  \hspace{1cm} (2.1)

where \( x_t \) is the price of a some asset at time \( t \), \( d_t \) is the dividend (if any) paid during time period \( t \), and \( y_t \) is the return series. Throughout this thesis we work with non-dividend paying assets, so \( d_t = 0 \). We work with compounded returns for reasons that are mentioned in Campbell et al. (1997, Section 1.4): continuously compounded multi-period returns are the sum of continuously compounded single-period returns. Usually in this thesis we will work with variables in discrete time, denoted \( y_t \). Occasionally we switch to processes in continuous time. Variables of such processes will be denoted \( y(t) \).

2.1.2 Risk versus Return

It was common believe in the 1970s that financial data like stock prices are unpredictable. We have to be cautious about what is exactly meant by unpredictable. Today's level of the S&P500 index\(^1\) will provide a rather good estimate of tomorrow's level of the S&P500 index. Obviously we are not interested in such predictions, but we are interested in a prediction of tomorrow's return, as defined in (2.1), from the S&P500. It is usually believed within the walls of academia that a shift in the return cannot be predicted from the past of the series alone. This is formulated in the efficient market hypothesis (EMH), which goes back to Bachelier (1900), and, bluntly stated, says:

\(^1\)The S&P500 index is a value-weighted index of primarily high-capitalization firms.
2.1. CHARACTERISTICS OF FINANCIAL TIME SERIES

(i) The past history is fully reflected in the present price, which does not hold any further information;

(ii) Markets respond immediately to any new information.

It is beyond the scope of this thesis to give a review of the EMH. Excellent reviews are provided in LeRoy (1989) and Lo (1996). The view that is taken in this thesis is that a model for future values of asset prices must incorporate a certain degree of randomness. Note that randomness is associated with risk\(^2\). Risk is modelled by assigning probabilities to possible outcomes. Investment theory is founded on the concepts of risk and expected return; see Markowitz (1952) and Roy (1952). This theory states that there is a trade-off between risk and expected return. In investment theory utility functions are used to model the preferences of the investor regarding risk and expected return. As argued in Rothschild and Stiglitz (1970), risk is often associated with standard deviation or variance, which by itself is a measure of the variability of a series. This brings up the notion of volatility. Volatility is the process driving the variability. Conditioning on different information sets gives rise to different volatility concepts, as we shall see in Section 2.2. In this thesis we will focus on the time-dependence of the volatility, and in particular the modelling of this time-dependent volatility through SV models. Accurate models for volatility provide an accurate quantification of risk.

Not only stock returns are volatile: because of several institutional changes in the 1970s volatility has appeared also in foreign exchange and interest rates. One type of institutional change is labelled globalization: in the past two decades we have witnessed both a growth of world trade and an unprecedented liberalization such as freeing of exchange and capital controls. This process has introduced volatility in the exchange-rate markets in the 1970s, prompting a search for hedging instruments for the elimination of currency risk. Another institutional change is the elimination of interest-rate controls. Together with large new issues of government debt due to budget deficits in many countries, this has prompted a search for financial instruments to eliminate interest-rate risk.

Opposed to predicting returns, the EMH says little about predicting tomorrow’s volatility from the past of the series. As distinct from returns, there exists strong evidence that volatility is highly predictable as we shall see in Chapter 7 in particular. Option markets are sometimes labelled as markets where volatilities are

\(^2\)In the literature an important distinction is made between systematic risk (or market risk) and unsystematic risk. Total risk of a security is the sum of systematic risk and unsystematic risk. Systematic risk is the part of the risk that is due to the variability of the general market, whereas unsystematic risk is attributed to factors specific to that particular security. This insight is due to the CAPM model of Sharpe (1964) and Lintner (1965). For a lively account of the history and current manifestations of risk see Bernstein (1996).
traded. Therefore it may be possible to use models for volatility for both specula-
tion and hedging or, in other words, for taking risk and for the elimination of risk.

A simple example of speculation is as follows. If we believe an asset will be highly
volatile in the future, it is more likely that a large price movement will occur than
a small. By buying both a call and a put option both with the same strike price
and time to expiration, we create a straddle. If we believe an asset will have low
volatility in future and consequently small price movements will occur, we sell the
straddle. Hedging of risk will be improved because more accurate estimates of the
actual option prices or of the parameters of a possibly more advanced option pric-
ing formula, can be obtained as shown in Chapter 6. Depending on a person’s risk
profile such models may also be used to improve speculation.

2.1.3 Empirical Regularities

A listing of empirical regularities or “stylized facts” that are present in financial
time series can be found in e.g. Taylor (1986), Karpoff (1987), Dimson (1988)
and Bollerslev, Engle and Nelson (1994), and in the references therein. Tests for
these empirical regularities are mainly $t$-tests for significance of some coefficient
in a certain statistical model. Empirical regularities can be divided into two sub-
classes: (i) regularities due to imperfections in the trading process itself; e.g. day-
of-the-week effects, half-of-the-month effects and non-trading periods, and (ii) regulari-
ties in more economic terms; e.g. the small-firm effect, the turn-of-the-year effect.

Below we will briefly discuss some of the empirical regularities in financial data
that are relevant for this thesis.

In financial returns we observe periods of high volatility followed by periods
of low volatility. This phenomenon is referred to as volatility clustering and was
coined by Mandelbrot (1963). Volatility clustering is clearly present in the series
from the lower panel of Figure 2.1, which displays levels and returns from the
S&P500 1963–1993. A simple statistical method that reveals this feature is to first
fit some regression model to the returns and then to regress the squared residuals
on a constant and several of its own lagged values; see the upper right panel of
Figure 2.2 which displays a correlogram of the squared residuals of the S&P500
series. These tests parallel tests for AR errors in ordinary Box-Jenkins time-series

---

3 See e.g. Hull (1997, p.187).
4 Note that one should be aware of the dangers of data-mining or data-snooping in such practice;
see White (1998).
5 Day-of-the-week effect: Mondays tend to have a statistically significant negative mean return;
1977. The negative mean of the mondays is highly significant in case a $t$-test is employed.
6 In Chapter 5 we discuss how the returns were pre-whitened. This was done to remove some
small autocorrelation which is often present in daily (index) returns series. As can be seen from
2.1. CHARACTERISTICS OF FINANCIAL TIME SERIES

Figure 2.1: Daily levels and returns of S&P500, 1963-1993

analysis, except that for AR errors the first moments are considered and for testing for volatility clustering the second moments are considered; see Engle (1982). Volatility clustering motivates models that include some sort of autocorrelation of the time-dependent volatility. Explanations why volatility is not constant over time are given in Clark (1973) where price changes are modelled as the results of random information arrivals. This idea has later been refined in Tauchen and Pitts (1983). Furthermore there seems to be some price-volume relationship causing this volatility. High trading volume seems to indicate more information flowing to the market and seems to cause changes in the price volatility; see Karpoff (1987) for a review on these price-volume relationships. Other more recent explanations for volatility clustering refer to the heterogeneity of the market participants; see Grossman and Zhou (1996): the dynamic interaction between groups of market participants who have different risk and reward profiles and different time frames, sc. some people trade at short time intervals with high risk for profit, others trade infrequently at low risk for hedging purposes.

Two other features that can be calculated by statistical measures are skewness and excess kurtosis (leptokurtosis). Compared to the normal density, the empirical density of financial returns has in general thick tails and seems to be somewhat skewed to the left. The excess kurtosis feature is clearly visible from a plot of the unconditional empirical density of the S&P500 series as in Figure 2.2 bot-

Figure 2.2 (upper left panel), where a correlogram of the pre-whitened returns is displayed, there is not much autocorrelation present.
Figure 2.2: Salient features of pre-whitened daily returns of S&P500, 1963-1993. Top left displays correlogram of the residuals; Top right displays correlogram of the squared pre-whitened returns; Bottom left displays a QQ-plot of the pre-whitened returns versus the Normal distribution; Bottom right displays the empirical density of the pre-whitened residuals and a Normal approximation. Here $s$ denotes the estimated standard deviation.
tom right. There exist two competing hypotheses that explain the excess kurtosis: (i) The stable-Paretian hypothesis: rates of return stem from distributions with infinite variances\(^7\). (ii) The mixture of distributions hypothesis: rates of return stem from a mixture of distributions with different conditional variances\(^8\). The skewness is not very obvious from the unconditional density estimate of Figure 2.2.

Furthermore, there seems to be considerable asymmetry in the way volatility responds to changes; negative returns tend to increase the investors' expectation about future volatility more than positive returns; see French, Schwert and Stambaugh (1987). We shall see below that asymmetry does not necessarily lead to skewness of the empirical density. There exist at least two competing hypotheses that explain the asymmetry: (i) The leverage-effect hypothesis, see Black (1976) and Christie (1982): firms fail to adjust their debt-equity ratio. A negative return in the stock price increases this debt-equity ratio and this in turn increases the risk of the investor. (ii) The volatility feedback hypothesis, see Campbell and Hentschel (1992): positive shocks to volatility drive down returns. Below we shall introduce stochastic volatility models that accommodate both the asymmetry and the lepto-kurtosis.

The last feature that should be dealt with here, as it partly motivates multivariate models, is co-movements in volatilities: markets tend to move together. This is a trivial observation of the simultaneous aspect of economic data that is present in all branches of empirical economics. We will introduce multivariate SV models in Section 2.2.2.

### 2.2 Models for Volatility

Starting off with an imaginative and remarkable doctoral dissertation by Bachelier (1900), which is both a remarkable study of speculative prices and an imaginative empirical investigation, the analysis of financial data has regained interest only decades later through Working (1934) and Kendall (1953). In these papers the first serious quantitative attempts have been made to investigate financial data empirically. The academic world got only fully interested in financial data through the

---

\(^7\)Since data from financial data sets often display very large outliers, there is some evidence supporting the stable-Paretian hypothesis. In Cootner (1964, pp. 333-337) it is argued that the infinite variance property of these distributions causes most of our statistical tools which are based on finite-moment assumptions to be worthless, even the expectation of the arithmetic price change does not exist. On the other hand it is observed that the parameter representing the stable distribution does not remain constant when looking at different frequencies. This is contradicting the stable-Paretian hypothesis. For this reason the stable-Paretian hypothesis is not widely accepted at present.

\(^8\)Carlin and Polson (1991) show that a mixture of normals can account for a double-exponential, exponential power, logistic or \(t\)-distribution.
papers by Mandelbrot (1963) and Fama (1965) in the 1960s. In these papers it is assumed that the log price changes for cotton and common stock prices stem from a non-Gaussian distribution, or more precisely, a stable-Paretian distribution with infinite variance. Also it was found that these series display pronounced volatility clustering. Still, it took until the 1980s for this to be accepted, mainly due to the introduction of ARCH models in Engle (1982). A landmark in the early empirical-finance literature is Cootner (1964), in which a bundle of major articles have been put together, including most of the above mentioned. By the beginning of the seventies it was still generally believed that stock prices followed a random walk, or more precisely a martingale process, where the returns were thought to be log-normally distributed. The mid 1970s and the 1980s brought a variety of articles where new statistical models, like regime-switching (STAR) models, the ARCH class of models, models from chaos theory and cointegration models were introduced. Furthermore new data sets, longer data sets, data sets based on different time periods, and causalities between series were investigated. At present, while computer speed is accelerating, the focus seems to be on developing sophisticated estimation techniques in order to employ these complicated large data sets and to estimate these intricate models. Among these intricate models are artificial intelligence models such as neural networks, and the stochastic volatility models that will be considered in this thesis. The development of stochastic volatility models — models that cannot be estimated in general by ordinary direct Maximum Likelihood techniques — is boosted by the recent massive increase in computing power.

After recognising that volatility is changing over time, researchers attempted to model it. To capture the serial correlations in volatility one can model the conditional variance as a function of the previous returns and past variances. This has led to the AutoRegressive Conditional Heteroskedasticity (ARCH) models, which were developed by Engle (1982) and Bollerslev (1986). An alternative approach is to model the conditional variance as a latent variable as a function of previous returns and variances. For example, we may assume that the logarithm of conditional volatility follows an autoregressive time-series model with an idiosyncratic error term. The models that arise from this approach are SV models. In an SV model,

---

9 For an account of the perception of quantitative techniques on the trading floor see Bernstein (1992, Part 5).
10 See e.g. Tong (1990) for a review.
11 See e.g. Bollerslev, Chou and Kroner (1992) and Bollerslev et al. (1994) for a review.
13 See e.g. Mills (1993).
14 See e.g. Goodhart and O'Hara (1997).
15 See e.g. Hutchinson, Lo and Poggio (1994) for an application in finance.
16 Harvey, Ruiz and Shephard (1994) refers to these models as stochastic variance models. Taylor (1994) refers to these models as autoregressive random variance (ARV) models.
there is an extra idiosyncratic error term in the volatility process, causing the volatility to be latent. The model where the volatility follows an AR(1) scheme is used as a benchmark model in the early literature on the estimation of these models. For this reason this model is often referred to as the stochastic-volatility model. The first reference to the stochastic volatility class of models is Clark (1973).

It should be noted that under ARCH and SV models the martingale property of the returns can still be preserved, so ARCH and SV models are not contradicting random-walk theory necessarily. An all-encompassing theoretical model replacing the EMH has yet to emerge, however.

In this thesis we will adhere to the categorisation into observation-driven and parameter-driven or state-space models as suggested in Cox (1981) and Shephard (1996a). Observation-driven methods, like ARCH models, can in principle be estimated by standard likelihood techniques. This is because the one-step prediction density has a closed form. Parameter-driven models, like SV models do not have this property. Below we will discuss these latter models. We will restrict ourselves to parameter-driven models for volatility, but since in this thesis observation-driven models also play a role as auxiliary models, we will discuss these first.

### 2.2.1 Observation-driven Models

The nomenclature for the observation-driven class of models in the time-varying-volatility literature seems to have been evolved from comics books: ARCH, EGARCH, GARCH and so on; see Bollerslev et al. (1992) and Bollerslev et al. (1994) for a review and Engle (1995) for a collection of reprints of some important papers in this area. In the following we will illustrate the mechanics of these models.

Consider the following model for $y_t$ in (2.1) for $t \in \{1, \ldots, T\}$

\[ y_t = m_t + h_t z_t \quad (2.2) \]

\[ z_t \sim \text{IIN}(0, 1) \quad (2.3) \]

Here and throughout this thesis IIN denotes identically and independently normally distributed. In this section we are not interested in modelling the mean of the process defined by $m_t$ so we set $m_t = 0$ here, but in principle both the mean $m_t$ and the volatility $h_t$ should be modelled simultaneously. The main feature of the observation-driven models is that the variance $h_t^2$ is a function of past observations alone, as in e.g. the GARCH($p$, $q$) model\footnote{Generalised AutoRegressive Conditional Heteroskedasticity} of Bollerslev (1986):

\[ h_t^2 = \xi + \sum_{i=1}^{p} \rho_i L^i h_t^2 + \sum_{j=1}^{q} \alpha_j L^j h_t^2 z_t^2 = \xi + \rho(L)h_{t-1}^2 + \alpha(L)h_{t-1}^2 z_{t-1}^2 \quad (2.4) \]
where $L$ is the lag operator. Model (2.2) to (2.4) is an extension of the ARCH($p$) model of Engle (1982), which is obtained by setting $q = 0$. For the conditional variance $h_t^2$ in the GARCH($p$, $q$) to be well defined all the coefficients in the corresponding infinite-order ARCH model must be positive. Provided that $\rho(L)$ and $\alpha(L)$ have no common roots and that the roots of $\rho(z) = 1$ lie outside the unit circle, this positivity constraint is satisfied if and only if all the coefficients in the infinite power-series expansion for $(1 - \rho(z))^{-1} \alpha(z)$ are non-negative. See Nelson and Cao (1992) for necessary and sufficient conditions. The model is covariance stationary if and only if the roots of $\alpha(z) + \rho(z) = 1$ lie outside the unit circle. It is beyond the subject of this thesis to discuss all the different stationarity concepts associated with GARCH models. The interested reader is referred to Drost and Nijman (1993), Kleibergen and van Dijk (1993) and Nelson and Cao (1992).

As an alternative observation-driven model to (2.4) Nelson (1991) proposes the EGARCH($p$, $q$) model:

$$\ln h_t^2 = \xi + \sum_{i=1}^{p} \rho_i L^i \ln h_{t-i}^2 + (1 + \sum_{j=1}^{q} \alpha_j L^j)(\kappa_1 z_{t-1} + \kappa_2 || z_{t-1} | | - E | | z_t | |) \quad (2.5)$$

Stationarity conditions for this model follow from the usual stationarity conditions for ARMA models. This model has the important feature that it measures asymmetry through the parameter $\kappa_1$. This asymmetry parameter could capture the leverage effect mentioned in Section 2.1.3. The EGARCH model plays a major role in this thesis for the estimation of SV models by EMM as we will see in Chapter 3.

Although for certain parameter values of the GARCH and EGARCH models the unconditional distribution for $h_t z_t$ is leptokurtic, it is not sufficient to explain the fat tails usually found in financial data. For this reason Bollerslev (1987) proposes a GARCH model with Student-$t$ errors and Nelson (1991) proposes an EGARCH model with the Generalized Error Distribution. In Section 3.2.1 we will introduce EGARCH models with Semi-NonParametric (SNP) errors and with Student-$t$ errors. There we will also introduce multivariate EGARCH models.

Section 2.3.1 briefly discusses how to estimate observation-driven models. Estimation can in principle be tackled by straightforward maximum-likelihood methods, since for $t \in \{1, \ldots, T\}$ the explicit conditional densities

$$y_t | \sigma(\mathcal{Y}_{t-1}) \sim N(0, h_t^2) \quad (2.6)$$

are the components of the prediction-error decomposition of the likelihood. Here $\sigma(\mathcal{Y}_t)$ denotes the $\sigma$-algebra generated by the set $\mathcal{Y}_t = \{y_{-L}, y_{-1}, y_0, \ldots, y_t\}$, where $L$ denotes the maximum lag length of the endogenous variables. When $h_t$ is contained in this information set, as is the case in ARCH, GARCH and EGARCH models where $h_t^2 = \text{Var}(y_t | \sigma(\mathcal{Y}_{t-1}))$, this density has a closed-form expression. Similar expressions can be derived for error structures other than the Gaussian error structure.
2.2. MODELS FOR VOLATILITY

2.2.2 Parameter-driven Models

The mechanics of the parameter-driven models in the time-varying-volatility literature can be illustrated as follows. Let us assume the following model for \( y_t \) in (2.1) for \( t \in \{1, \ldots, T\} \):

\[
\begin{align*}
y_t &= \mu_t + \sigma_t z_t \\
z_t &\sim \text{IIN}(0,1)
\end{align*}
\]

As in the previous section we set the terms corresponding to the mean \( \mu_t \) equal to 0. In the SV models the \( \sigma_t \) are a function of some unobserved or latent variables, as in for example the following equation

\[
\ln \sigma_{t+1}^2 = \omega + \gamma \ln \sigma_t^2 + \sigma_\eta \eta_{t+1}
\]

Here \( \omega, \gamma, \) and \( \sigma_\eta \) are parameters and \( \eta_t \sim \text{IIN}(0,1) \). This is a stochastic volatility model in which \( \ln \sigma_t^2 \) follows an AR(1) process. In this case \( \sigma_t = \text{Var}(y_t | \sigma(Y_{t-1}), \sigma(S_t)) \) where \( \sigma(S_t) \) denotes the \( \sigma \)-algebra generated by the set \( S_t = \{ \sigma_{-1}, \ldots, \sigma_0, \sigma_1, \ldots, \sigma_t \} \).

Since \( z_t \) in (2.7) is always strictly stationary, for \(-1 < \gamma < 1 \) and \( \sigma_\eta \geq 0 \), \( y_t \) is strictly stationary and ergodic, and unconditional moments of any order exist, as \( y_t \) is the product of two strictly stationary processes in this case. In empirical work employing this model, it has been reported that \( \gamma \) is smaller than but close to unity.

In e.g. Harvey et al. (1994), Mahieu and Schotman (1998), Jacquier, Polson and Rossi (1994), Ruiz (1994), Danielsson and Richard (1993, 1994), Andersen and Sørensen (1996), Andersen, Chung and Sørensen (1999), Fridman and Harris (1998) and Sandmann and Koopman (1998) the process defined by (2.7) and (2.8) is used as a benchmark for their estimation procedures. Taylor (1994) and Andersen (1994) employ an AR(1) process for \( \ln \sigma_t \) instead of \( \ln \sigma_t^2 \).

Model (2.7) and (2.8) with \( \mu_t = 0 \) represents the Euler discretisation of the following continuous-time model (diffusion or Stochastic Differential Equation (SDE)) for the log asset price \( y^*(t) \) of Hull and White (1987):

\[
\begin{align*}
dy^*(t) &= c \sigma(t) dW_1(t) \\
d\ln \sigma(t)^2 &= -a \ln \sigma(t)^2 dt + b dW_2(t)
\end{align*}
\]

where \( W_1 \) and \( W_2 \) represent independent Brownian motions. Very often in this thesis we will work with discrete-time models without stating their continuous-time counterparts. Section 3.2.2 discusses the estimation of continuous-time SV models.

Estimation of stochastic volatility models is far from straightforward. Consider again the basic SV model (2.7) and (2.8) with \( \mu_t = 0 \). Let \( \theta = (\omega, \gamma, \sigma_\eta)' \) be the parameter vector of interest. Define two separate stochastic processes \( \Sigma_T = \{ \sigma_t^2 \}_{t=1}^T \).
CHAPTER 2. ANALYSIS OF FINANCIAL TIME SERIES

and \( Y_T = \{y_t\}_{t=1}^T \). The likelihood of the process is \( p(Y_T, \Sigma_T \mid \theta) \). Since the process \( \Sigma_T \) is unobservable or latent, we must integrate this variable out in order to obtain

\[
p(Y_T \mid \theta) = \int p(Y_T, \Sigma_T \mid \theta) d\Sigma_T
\]

(2.11)

This integral will be of dimension \( T \), the number of observations. In financial time series this number will in general be large, say \( 1,000 < T < 10,000 \). Standard numerical or analytical methods are not useful for this problem. It can also be seen that the explicit forecast densities

\[
p(y_t \mid Y_{t-1})
\]

(2.12)

are very difficult to compute for \( t \in \{1, \ldots, T\} \). The problem for SV models is that \( \sigma_t \) is not contained in the information set \( Y_t \), whereas for ARCH models this is the case, cf. (2.6). A way to work around this problem is to note that the equations in the stochastic volatility model resemble the state-space equations of the Kalman filter. Equation (2.7) with \( \mu_t = 0 \) and taking logs, together with (2.8) seem to tie in with the Gaussian state-space models of Harvey (1989) for parameter-driven models for the mean. However, parameter-driven volatility models do not exactly fit in this framework, because of the lack of explicit forecast densities. In Section 2.3 we will deal with several proposed solutions to this problem.

Though the SV class of models has, unlike the ARCH class of models, the unappealing property that its likelihood function is in general analytically intractable, SV models have other appealing properties. First, in Jacquier et al. (1994) the autocorrelations of the squared returns are compared with the implied theoretical autocorrelations of a stochastic volatility model and a GARCH model. The stochastic volatility model is in closer correspondence to the data than the GARCH model. Second, as we shall see in Section 2.2 these models are easier to formulate, understand, manipulate and generalize to the multivariate case. With respect to the latter we also mention that multivariate versions of ARCH models induce a proliferation of parameters, whereas stochastic volatility models allow for a more natural extension to higher dimensions. Third, SV models also have simpler continuous-time analogues or reversely, discrete time SV models are natural approximations to the diffusions from theoretical finance; see Melino and Turnbull (1990) and Wiggins (1987). From a different perspective we may add that stochastic volatility models match the theoretical models for the generation of asset returns that have been built by using unobservable or latent factors, e.g. arbitrage pricing theory (APT) and the mixture of disturbances hypothesis. The last remark that can be made is on the correspondence between the discrete-time ARCH and stochastic volatility models and the continuous-time models (diffusion processes) from financial theory as given in, among others, Duffie (1996). Nelson (1990) shows how ARCH models approximate these diffusion processes. Dassios (1992) shows that a stochastic
2.2. MODELS FOR VOLATILITY

volatility, with volatility following an AR(1) process, is a “better” discrete time approximation than an EGARCH model to the continuous-time model of Hull and White (1987), in the sense that the density of the variance process converges to the density of the continuous time process at a rate $\delta$ in the SV case and at a rate $\sqrt{\delta}$ in the case of an EGARCH model, where $\delta$ denotes the distance between the observations.

Univariate (Asymmetric) Gaussian SV Models

Many variations on the model defined by (2.7) and (2.8) are possible. Departures from the basic model affect inter alia the measured persistence and hence the prediction of volatility. This has policy implications on decisions and models for e.g. asset allocation and option pricing. First we generalize the dynamics of the model, but stay within the univariate Gaussian class of models. Later we will leave the Gaussian class and the univariate class.

We propose the following SV model allowing for more general dynamics

$$y_t = \mu_t + \sigma_t z_t$$

$$\ln \sigma_t^2 = \omega + \sum_{i=1}^{p} \gamma_i L^i \ln \sigma_t^2 + \sigma_\eta (1 + \sum_{j=1}^{q} \zeta_j L^j) \eta_t$$

$$\begin{bmatrix} z_t \\ \eta_{t+1} \end{bmatrix} \sim \text{IN}(0, \begin{bmatrix} 1 & \lambda \\ \lambda & 1 \end{bmatrix}), -1 < \lambda < 1$$

This class of models will be referred to as ASARMAV($p, q$) models\textsuperscript{18}, as the role of the $\omega$, $\gamma$ and $\zeta$ parameters is similar to their role in ARMA models. The parameter $\sigma_\eta$ is the volatility-of-volatility parameter and governs the idiosyncratic volatility in the model. The parameter $\lambda$ governs the correlation between $z_{t-1}$ and $\eta_t$. This allows for asymmetric behaviour which is often present in financial time series, due to the leverage effect: an increase in predicted volatility tends to be associated with a decrease in the stock price, suggesting $\lambda < 0$. This asymmetric generalization is due to Harvey and Shephard (1996). For $\lambda = 0$, we will, for obvious reasons, refer to the SARMAV($p, q$) class of models. Using the idea that volatility follows a unit-root process, Engle and Lee (1999) find that the leverage effect is more a temporary than a permanent feature of the volatility process.

Note that if $z_t$ is a martingale difference, as is the case in (2.13) where $z_t|Y_{t-1}, \Sigma_t \sim N(0, 1)$, then even if $z_t$ and $\eta_{t+1}$ are dependent, $y_t - \mu_t$ is also a martingale difference sequence. The martingale property is an important property that is shared with the EGARCH class of models. This is not true if we would model $z_t$ and $\eta_t$ to be dependent as in Taylor (1994).

\textsuperscript{18}Asymmetric Stochastic AutoRegressive Moving Average Volatility models
CHAPTER 2. ANALYSIS OF FINANCIAL TIME SERIES

The statistical properties of model (2.13) for \( \mu_t = 0 \) are as follows. We find that \( y_t \) is stationary if \( \ln \sigma_t^2 \) is stationary. Therefore for strict stationarity we need the roots of \( 1 - \sum_{i=1}^{p} \gamma_i z^i \) to lie outside the unit circle. Furthermore using moment generating functions as found in e.g. Abramowitz and Stegun (1972, Ch. 26) we find the following properties. Let \( \phi = \omega/(1 - \sum_{i=1}^{p} \gamma_i) \), and \( \tau^2 = \sigma_n^2(1 + \sum_{j=1}^{q} \zeta_j^2)/(1 - \sum_{i=1}^{p} \gamma_i^2) \). Then

\[
E y_t^i = 0, \text{ i odd}
\]

\[
E y_t^i = E(\sigma_t^i z_t^i) = E(\sigma_t^i)E(z_t^i) = \frac{i!}{2^{i/2}(i/2)!} \exp\left\{ \frac{i}{2} \phi + \frac{i^2}{8} \tau^2 \right\}, \text{ i even}
\]

Therefore for the ASARMAV(p, q) model the kurtosis equals \( 3e^{\tau^2} \geq 3 \). This is generic for models with changing volatility, though, unlike the ARCH-type models, for \( \gamma = 0 \) we still have excess kurtosis for \( \sigma_n > 0 \). Note that from (2.16) we find that the distribution of \( y_t \) is symmetric even if \( \lambda \neq 0 \). From Taylor (1986, pp. 74–75) we have that for \( p = 1 \) and \( q = 0 \), but allowing \( \lambda \neq 0 \), the autocorrelation between squared observations is

\[
\text{Cor}(y_t^2, y_{t-1}^2) = \frac{\exp(\tau^2 \gamma_1^2) - 1}{3 \exp(\tau^2) - 1}
\]

which means exponential decay for \( |\gamma_1| < 1 \). Since \( \ln y_t^2 \) is the sum of an AR(1) component and white noise, its autocorrelation function (ACF) is identical to that of an ARMA(1, 1) process. The ACF of \( y_t^2 \) in a GARCH(p, q) process also looks like that of an ARMA(1, 1) process. Taylor (1986) shows that when the variance of \( \ln \sigma_t^2 \) is small and/or \( \gamma \) is close to unity, \( y_t^2 \) is similar to an ARMA(1, 1) process. Expressions for \( E[y_t^2] \) and \( E[y_t^2 y_{t-1}^2] \) can be obtained in a similar fashion, see Harvey (1993) and Jacquier et al. (1994), but do not provide interesting new insights into the behaviour of the model.

It should be noted that high-order ASARMAV models do not easily tie in with the continuous-time literature, though recently Renault\(^19\) has shown that higher-order dynamics in discrete time can be reproduced by marginalization of multivariate continuous-time processes of underlying factors. In this thesis these high-order models are more empirically motivated, which is similar to the role of high-order ARMA models in econometric theory. The link between the continuous-time models and the asymmetry parameter \( \lambda \) is well-known and goes back to Hull and White (1987).

\(^{19}\)Personal communication, Eric Renault, Ecole Nationale de la Statistique et de l’Analyse de l’Information, France
2.2. MODELS FOR VOLATILITY

Univariate (Asymmetric) Non-Gaussian SV Models

A modification of the Gaussian class of models that allows for even more excess kurtosis is the SV model with a scaled Student-\( t \_\nu \) distribution. Such a generalization is motivated by empirical observations that Gaussian SV processes do not display the degree of leptokurtosis that is present in financial time series. Again this generalization has policy implications.

This class of models reads

\[
\begin{align*}
\ln \sigma_t^2 &= \omega + \sum_{i=1}^{p} \gamma_i \ln \sigma_{i-1}^2 + \sigma_{\eta}(1 + \sum_{j=1}^{q} \zeta_j \eta_j) \eta_t \\
\xi_t &\sim \sqrt{\frac{\chi^2_{\nu} \nu}{\nu - 2}}, \nu > 2 \\
\begin{bmatrix} z_t \\ \eta_{t+1} \end{bmatrix} &\sim \text{IN}(0, \begin{bmatrix} 1 & \lambda \\ \lambda & 1 \end{bmatrix}), -1 \leq \lambda \leq 1 \\
\xi_t &\text{is independent of} \begin{bmatrix} z_t \\ \eta_{t+1} \end{bmatrix}
\end{align*}
\]

where \( \nu \) is treated as a parameter to be estimated. Again for strict stationarity we need the roots of \( 1 - \sum_{i=1}^{p} \gamma_i z^2_i \) to lie outside the unit circle. We will refer to this model as the ASARMAV\((p, q, -t\_\nu\) model. The extension \("-t\_\nu\) is motivated by the fact that \( z_t/\xi_t \sim N(0, 1)/\sqrt{\nu - 2} \) follows a standardized Student-\( t \_\nu \) distribution. Note that taking \( \nu \to \infty \) in (2.19), yields model (2.13). The ASARMAV-\( t \) model will be able to capture both asymmetry and leptokurtosis, beyond the leptokurtosis already captured by the ASARMAV model. The properties of Student-\( t \) errors are well known, and we only mention that the Student-\( t \) errors are normalized in order for the parameters of model (2.19) to be comparable to those of model (2.13). Kim, Shephard and Chib (1998) were the first who proposed model (2.19), however they put \( \lambda = 0 \). We mention that this model has finite variance for \( \nu > 2 \). The statistical properties for the ASARMAV\((p, q, -t\_\nu\) model are

\[
\begin{align*}
\mathbb{E} y_t^i &= 0 \text{ for } i \text{ odd} \\
\mathbb{E} y_t^i &= \frac{1 \cdot 3 \cdots (i-1)(\nu - 2)^{i/2}}{(\nu - 2)(\nu - 4) \cdots (\nu - i)} \exp\{\frac{i}{2} \phi + \frac{i^2}{8 \tau^2}\}, \nu > i \text{ for } i \text{ even}
\end{align*}
\]

Therefore for the ASARMAV\((p, q, -t\_\nu\) model the kurtosis equals \( 3 \frac{\nu - 2}{\nu - 4} \) for \( \nu > 4 \). The continuous-time counterparts of these non-Gaussian models are not known to the author; recently Barndorff-Nielsen and Shephard (1999) have some results on non-Gaussian continuous-time SV models other than the one studied here.
Multivariate (Asymmetric) Gaussian SV Models

Until recently multivariate generalizations of the SV model have not been extensively studied in the literature. Multivariate models are important because they enable to identify co-movements or common persistence in volatility. In the ARCH literature Bollerslev (1990) proposes a multivariate variant of the GARCH model. Other studies on multivariate GARCH models include Engle and Kroner (1995) and Bollerslev, Engle and Wooldridge (1988). From that literature it is clear that the number of parameters becomes very large, so restrictions should be imposed.

The first multivariate generalizations of the univariate SV model were proposed in Harvey et al. (1994). Recently Danielsson (1998) looks at estimating this model using SML. In this thesis we expand the SV model of Harvey et al. (1994) allowing for asymmetry.

In this thesis we use the following representation of an n-variate asymmetric stochastic autoregressive volatility model of order $p$ for the possibly detrended and pre-whitened asset return process, for $i \in \{1, \ldots, n\}$

\[
\begin{align*}
 y_t & = \mathbf{N}_t z_t \\
 \ln[\text{diag}(\mathbf{N}_t)^2] & = \omega + \sum_{i=1}^{p} \Gamma_i L^i \ln[\text{diag}(\mathbf{N}_t)^2] + \Sigma_\eta \eta_t \\
 \begin{bmatrix} z_t \\ \eta_{t+1} \end{bmatrix} & \sim \text{INN}(0, \begin{bmatrix} C & Q \\ Q & V \end{bmatrix})
\end{align*}
\]

(2.26)

(2.27)

(2.28)

where $y_t$ is an $n$-vector of observations, $\mathbf{N}_t$ is an $n \times n$ diagonal matrix with the latent volatility $\sigma_{it}$ on the diagonal, $\Gamma_i$ is an $n \times n$ matrix with elements $\gamma_{ij}$ and $C$ and $V$ are $n \times n$ symmetric matrices with elements denoted $c_{ij}$ and $v_{ij}$ respectively. For identification the diagonal elements of $C$ equal 1. Furthermore, $\omega$ is an $n$-vector with elements $\omega_i$ and $Q$ is a diagonal matrix with diagonal elements $q_i$. Here $\text{diag}(A)$ denotes $(a_{11}, a_{22}, \ldots, a_{nn})'$, where $a_{ii}$ is the $i$th diagonal element of the $n \times n$ matrix $A$ and $\ln[\text{diag}(A)]^2$ denotes a vector of $\ln a_{ii}^2$ for $i = 1, \ldots, n$.\(^{20}\)

For identification we also need $\Sigma_\eta$ to be a diagonal matrix, with elements $\sigma_{ii} > 0$ on the diagonal, $i = 1, \ldots, n$. We will refer to this model as AMSV($p$)\(^{21}\) and when $Q = I_n$ we will refer to it as MSV($p$)\(^{22}\).

The above model implies

\[
\eta_{t+1} = q_t z_t + \sqrt{1 - q_t^2} u_{it}
\]

(2.29)

---

\(^{20}\)Throughout this thesis we employ the following notation for $\text{diag}$: (i) if $A$ is an $n \times n$ matrix with elements $a_{ij}$ then $\text{diag}(A) := (a_{11}, \ldots, a_{nn})'$. (ii) if $a$ is a vector of elements $a_{ii}$, i.e. $a = (a_{11}, \ldots, a_{nn})'$ then $\text{diag}(a)$ denotes an $n \times n$ diagonal matrix with elements $a_{ii}$ on the diagonal.

\(^{21}\)Asymmetric Multivariate Stochastic Volatility

\(^{22}\)Multivariate Stochastic Volatility
where the $u_{it}$ are assumed $INN(0, 1)$. Since $z_{it}$ and $\eta_{it+1}$ are random shocks to the return and volatility of a specific stock respectively and, more importantly, both are subject to the same information set, it is reasonable to assume that $u_{it}$ is purely idiosyncratic or, in other words, it is independent of other random noises, including $u_{jt}$. This leads to the following restriction on the elements of the matrix $V$,

$$
u_{ij} = \text{Cor}(\eta_{it+1}, \eta_{jt+1}) = \text{Cov}(\eta_{it+1}, \eta_{jt+1}) = \text{Cov}(z_{it}, z_{jt})$$

$$= q_i q_j \text{Cov}(z_{it}, z_{jt})$$

Co-movements in volatility which are ascribed to correlation in the volatility shocks are modelled by the off-diagonal elements of $V$. Parallel to VAR models, co-movements in volatility are dynamically modelled by the off-diagonal elements of $\Gamma$. The returns are correlated through the off-diagonal elements of $C$. The matrix $Q$ governs the asymmetry or leverage effect.

The above model is stationary if the roots of $|I_n - \sum_{i=1}^{p} \Gamma_i z_i^2| = 0$ lie outside the unit circle. One may think of cointegration in the elements of $K_t$. It may be tempting to apply cointegration tests from Johansen (1988) using $\ln y_{it}^2$ in model (2.26) to (2.28). In principle this can be done but most likely the power of such tests will be very low. This can be seen from the basic symmetric univariate model (2.7) and (2.8) with $\mu_t = 0$. Rewriting this model in its reduced form yields

$$\ln y_{it}^2 = \omega/(1 - \gamma) + (1 - \gamma L) \ln z_{it}^2 + (1 - \gamma L) \ln(\sigma_{\eta}^2)$$

Since for financial data $\gamma$ is close to 1, from Pantula (1991) and Schwert (1989) we infer that in these cases it is difficult to distinguish the reduced-form model (2.31) from white noise, let alone to determine the cointegration rank in multivariate models. However the estimated roots of $|I_n - \sum_{i=1}^{p} \Gamma_i z_i^2| = 0$ in model (2.26) to (2.28) will give us an indication of the dynamics of the volatility process. In order to identify common sources of volatility we could also apply principal-components analysis to the elements of $V$ as was done in Harvey et al. (1994).

Some final remarks on the (A)MSV model are. In continuous time the (A)MSV(1) model corresponds to a system of SDEs. The generalization of equation (2.27) to include lagged $\eta_t$, as was done in the univariate case of (2.13), is straightforward but will not be pursued here.

Other Extensions of SV Models

In the literature other extensions of the SV model have been proposed. More sophisticated dynamics could also be introduced by factor models; see Kim et al. (1998). Factor structures have also been developed in Jacquier, Polson and Rossi (1998), Gallant and Long (1997), Mahieu and Schotman (1998) and Shephard and
Pitt (1998). Long memory stochastic volatility models, mimicking fractionally integrated ARCH-type models have been introduced by Harvey (1993), Comte and Renault (1998) and Breidt, Crato and de Lima (1998). These extensions will not be considered in this thesis.

2.3 Estimation Methods for SV Models

In the previous section we distinguished two classes of volatility models: observation-driven and parameter-driven models. Roughly we can divide estimation methods also into two classes: likelihood-based estimation and moment-based estimation. All observation-driven models can in principle be tackled by likelihood-based methods. There are however reasons why one may use moment-based estimation: first, consistency of moment-based estimation is easily proved with less compelling assumptions. Second, the model does not need to be fully specified which means that there is a whole class of models for which a specific moment-based estimation technique remains valid. Though for moment-based estimation in conjunction with simulation methods, as in this thesis, we need a fully specified model for our simulations, moment-based techniques still have the advantage that we can use the same implementation of the estimation technique for different models and we only need to change the generator of the simulated data.

Often parameter-driven models cannot be tackled by standard maximum likelihood methods. Exceptions to this are models that permit a Gaussian state-space representation, which in turn provides exact likelihood functions. In Section 2.3.1 we will see how this works out and why in general SV models do not fit into this state-space framework.

Simulation-based estimation and inference is one of the recent developments in both moment-based and likelihood-based econometric theory. Simulation provides an estimation technique and a specification-testing procedure for structural models for which no closed form for the likelihood exists or for which this closed form consists of high-dimensional integrals. Simulation methods may be subdivided in indirect-inference techniques and direct-inference techniques. Indirect inference techniques are based on an idea of Smith (1993) and refined into Indirect Inference (II) by Gouriéroux, Monfort and Renault (1993) (see Section 2.3.2) and EMM of Gallant and Tauchen (1996b) (see Section 2.3.2 and Chapter 3) respectively. Schematically, these simulation techniques may be described as follows. Let \( p(y_T, \ldots, y_t | \theta) \) be the density associated with the structural model. Let \( \theta \in \Theta \subset \mathbb{R}^d \).

\[23\]It can be shown that likelihood estimation is in fact moment-based estimation for a very specific choice of the moments, namely the scores of the model that we want to estimate, so the subdivision is somewhat arbitrary.
(i) Choose an auxiliary model \( f(y_T, \ldots, y_1 | \beta) \) with auxiliary parameters \( \beta \in B \subset \mathbb{R}_{\beta}, l_\beta \geq l_\theta \)
(ii) The (dynamic) properties of the observed sample \( \{y_t, 1 \leq t \leq T\} \)
are investigated under the auxiliary model.
(iii) Given a parameter value \( \theta \), \( S \) sample paths of length \( N \)
\( \{y_{t,s}(\theta), 1 \leq t \leq N, 1 \leq s \leq S\} \)
are generated from the structural model \( p(y_N, \ldots, y_1 | \theta) \).
(iv) An estimator \( \hat{\theta}_{S,N,T} \) is determined by the \( \theta \) that makes the (dynamic)
properties of the auxiliary model and the structural model as similar as possible.

Differences in these techniques are mainly determined by the choice of how
the dynamic properties are measured. These dynamic properties can be viewed
as moments and the differences are measured through some minimum chi-square
criterion. Indirect-inference techniques are typically moment-based.

Direct-inference techniques such as Simulated Maximum Likelihood (SML)
and Monte Carlo Markov Chain (MCMC) methods attempt to approximate the
analytically intractable transition density (2.12) using simulation techniques.
These methods will be discussed below. The SML techniques for stochastic vola­
tility models are mainly due to Danielsson and Richard (1993) and Danielsson
(1994), and are discussed in Section 2.3.1. Direct-inference techniques are typi­
cally likelihood-based.

Another distinction which is present in direct-inference techniques, is between
Bayesian and classical estimation. Both techniques assume an a priori density for
\( \sigma_t, \pi(\Sigma | \theta) \), which can be seen by writing (2.11) as

\[
p(Y | \theta) = \int p(Y | \Sigma, \theta) \pi(\Sigma | \theta) d\Sigma
\]  

(2.32)

Here \( \pi(\Sigma | \theta) \) can be viewed as a prior for \( \Sigma \), as given by e.g. (2.8). The controversy
is that Bayesian techniques also require a prior \( \pi(\theta) \) for the parameters. Recent
Bayesian techniques that can estimate SV models are the MCMC techniques that
will be discussed briefly in Section 2.3.1.

Estimation on the basis of simulation techniques is computationally very de­
manding. That is the reason why it is interesting to have — at least for some
models — methods available that do not impose such a computational burden.
Roughly speaking there are two such analytical methods available: GMM and
Quasi-Maximum Likelihood (QML). Sections 2.3.1 and 2.3.2 discuss these tech­
niques. Recently the Kalman filter techniques used for QML have been improved
by Fridman and Harris (1998) and Sandmann and Koopman (1998). Section 2.3.1
briefly discusses these techniques. Finally, the Method of Simulated Moments,
II and EMM can be viewed as variants of GMM. These techniques are discussed briefly in Section 2.3.2. Chapter 3 is completely devoted to EMM.

2.3.1 Likelihood-based Techniques

Likelihood-based techniques are often based on the prediction-error decomposition. Below we will give the general set-up that can be used for observation-driven models, such as the ARCH-class of models. The ARCH-class of models will serve as an auxiliary model throughout this thesis. Since ARCH models are obviously non-linear models, numerical optimization techniques must be employed. The generic optimization procedure can already be found in the original paper on ARCH models of Engle (1982), but Bollerslev et al. (1994) have put this procedure in a much broader context.

The model for the conditional mean is \( m_t(\beta) = E_t[y_t | t - 1] \), where the subscript \( t - 1 \) denotes the information set up to time \( t \). Next the zero mean process is defined as \( \epsilon_t(\beta) = y_t - m_t(\beta) \). The model for the conditional variance is \( h_t^2(\beta) = \text{Var}_t[\epsilon_t(\beta)] = E_t[\epsilon_t^2(\beta)] \), where \( \beta \) denotes the full parameter vector. This leads to the standardized process: \( z_t(\beta) = \epsilon_t(\beta)[h_t^2(\beta)]^{-1/2} \). Let \( f(z_t; \eta) \) be the density for \( z_t(\beta) \), where \( \eta \) denotes the vector of nuisance parameters \( \eta \in H \subset \mathbb{R}^k \).

Let \( \psi = (\beta', \eta')' \). The log-likelihood of \( y_t \) equals

\[
\ln_\ell(y_t; \psi) = \ln[f(z_t(\beta); \eta)] - 0.5 \ln[h_t^2(\beta)], t = 1, 2, ..., T \tag{2.33}
\]

By the prediction-error decomposition we get the following expression for the likelihood \( L_T \) of the full sample

\[
L_T(y_1, ..., y_T; \psi) = \sum_{t=1}^{T} \ln_\ell(y_t; \psi) \tag{2.34}
\]

This expression can be maximized using numerical optimization techniques, see e.g. Fletcher (1988). Such techniques could be speeded up by using analytical expressions for the gradients. In Appendix A explicit formulae for the gradients of several ARCH-type models can be found for reasons that will become clear in Chapter 3.

State-Space Techniques (Kalman Filter)

Harvey et al. (1994) and Nelson (1988) argued independently that the benchmark SV model given by (2.7) and (2.8) can be approximated by standard Kalman filter

\[24\) The notation \( \beta \) is the same as for parameters of an auxiliary model. This is done on purpose since these models will later be used as auxiliary models.\]
2.3. ESTIMATION METHODS FOR SV MODELS

techniques. The approximation turns out to be a bad approximation, giving rise to
The original standard Kalman filter approach is as follows. Consider the following
model (2.7) and (2.8). We may transform this model into

\[
\ln y_t^2 = \ln \sigma_t^2 + \ln z_t^2
\]  

(2.35)

where for simplicity \( \mu_t = 0 \). As Harvey et al. (1994) point out \( E(\ln z_t^2) = -1.2704 \)
and \( \text{Var}(\ln z_t^2) = \pi^2/2 \), see Abramowitz and Stegun (1972, p. 943). For the uni-
variate case their estimation technique basically comes down to the following state-
space model

\[
\begin{align*}
\ln y_t^2 &= -1.2704 + \ln \sigma_t^2 + \xi_t, \quad \text{measurement equation} \\
\ln \sigma_t^2 &= \omega + \gamma \ln \sigma_{t-1}^2 + \eta_t, \quad \text{transition equation}
\end{align*}
\]  

(2.36)

where \( \xi_t = \ln z_t^2 + 1.2704 \) and \( \text{Var}(\xi_t) = \pi^2/2 \). To this state-space model the
Kalman filter may be applied as follows:

- Prediction equations:

\[
\begin{align*}
\ln \sigma_{t|t-1}^2 &= \omega + \gamma \ln \sigma_{t-1|t-1}^2 \\
p_{t|t-1} &= \gamma^2 p_{t-1|t-1} + \sigma_t^2
\end{align*}
\]  

(2.37)

- Update equations:

\[
\begin{align*}
\ln \sigma_{t|t}^2 &= \ln \sigma_{t|t-1}^2 + p_{t|t-1} \frac{v_t}{f_t} \\
p_{t|t} &= p_{t|t-1} - \frac{p_{t|t-1} v_t}{f_t}
\end{align*}
\]  

(2.38)

where \( v_t = \ln(y_t^2) + 1.2704 - \ln \sigma_{t|t-1}^2 \), \( f_t = p_{t|t-1} + \frac{\pi^2}{2} \), \( \ln \sigma_{0|0}^2 = \frac{\omega}{1-\gamma} \) and
\( p_{0|0} = \frac{\sigma_t^2}{1-\gamma^2} \). The normal quasi-likelihood reads

\[
\ln L = -\frac{T \ln 2\pi}{2} - \frac{1}{2} \sum_{t=1}^{T} \ln f_t - \frac{1}{2} \sum_{t=1}^{T} v_t^2
\]  

(2.39)

Some remarks on the Kalman filter approach. First, because of the approxi-
mation only a minimum mean square linear estimator instead of a minimum mean
square estimator is obtained. Second, because of the approximation the technique
is a quasi maximum likelihood (QML) technique, and asymptotic standard errors
which take the non-normality of \( \xi_t \) into account may be obtained by using results
from Dunsmuir (1979, p. 502). Third, Ruiz (1994) presents interesting Monte
Carlo evidence on the properties of this technique. Finally, within the QML framework Harvey and Shephard (1996) modify the Kalman filter to deal with the asymmetry of the ASARMAV(1,0) model. In their seminal paper Harvey et al. (1994) already consider some multivariate and non-Gaussian extensions considered and estimate them using QML.

In a classical framework Sandmann and Koopman (1998) and Fridman and Harris (1998) improve the standard Kalman filter technique using simulation techniques. Sandmann and Koopman (1998) use results from Durbin and Koopman (1997) who show that the log-likelihood for state-space models with non-Gaussian measurement disturbances can be written as (2.39) plus a correction term given by

$$E_G \frac{f_{true}(\epsilon|\theta)}{f_G(\epsilon|\theta)}$$

where $f_{true}$ is the density function of the measurement disturbances, which is in the case of the above SV model a $\chi^2_1$ density, see (2.31) and $f_G$ is the Gaussian density of the measurements disturbances of the approximating model. The expectation is taken with respect to the density $f_G(\epsilon|y, \theta)$. The correction term in (2.40) must be determined by simulation. Details regarding implementation of this method can be found in Sandmann and Koopman (1998). A related method was developed by Fridman and Harris (1998). In this method the forecast density (2.12) is evaluated for each $t$ using an iterated numeric integration procedure for non-Gaussian state-space models of Kitagawa (1987) and by integrating $\Sigma_T$ out as in (2.11). A tractable solution is obtained by discretizing the domain of $\Sigma_T$.

A dramatic improvement of the QML approach has also been made by Kim et al. (1998) and Mahieu and Schotman (1998) using Bayesian methods which we will discuss in Section 2.3.1.

Simulated Maximum Likelihood

Hendry and Richard (1992), Danielsson and Richard (1993) and Danielsson (1994) propose importance sampling, see Ross (1990, Ch. 8), to perform the high-dimensional integration in (2.11). In brief the following is done. Define a density function $g(\Sigma|Y, \theta)$ such that the expected value of $p(Y, \Sigma|\theta)$ equals the marginal density of the observable variable. Sample from this density in order to obtain a Monte Carlo estimate of the expected value. That is

$$p(Y_T \mid \theta) = \int \frac{p(Y_T, \Sigma_T \mid \theta)}{g(\Sigma_T|Y_T, \theta)} g(\Sigma_T|Y_T, \theta) d\Sigma_T$$

2.3. **ESTIMATION METHODS FOR SV MODELS**

where it is easy to draw from \( g(\Sigma_T | Y_T) \). It can be shown that straightforward application of these techniques is often inefficient and may thus require an extremely large number of drawings \( N \). The use of standard variance reduction techniques\(^{26}\) such as control variates usually does not reduce \( N \) sufficiently, so one should resort to more clever techniques. One such technique is the *Accelerated Gaussian Importance Sampler (AGIS)*, introduced by Danielsson and Richard (1993). This is a clever implementation of importance sampling. The functions \( g \) are designed to recursively improve their performance converging to an optimal importance function. The interested reader is referred to the original papers for an exact description of this method.

**Monte Carlo Markov Chain Methods**

References on Monte Carlo Markov Chain (MCMC) methods in the context of SV models include Jacquier *et al.* (1994) and Kim *et al.* (1998). The MCMC method enables us to draw from \( p(\Sigma_T | \theta, Y_T) \) and — for Bayesians only — \( p(\theta | Y_T) \). Within the classical framework where one does not want to put a prior on the parameters, the \( p(\Sigma_T | \theta, Y_T) \) can be used inside the EM algorithm, as suggested in Kim *et al.* (1998). Otherwise \( p(\theta | Y_T) \) can be used to determine a posterior mode.

### 2.3.2 Moment-based Techniques

**Generalized Method of Moments**

In its full form this method is originally due to Hansen (1982). The asymptotic properties of this technique are well understood, see e.g. Newey and McFadden (1994). Andersen and Sørensen (1996) apply GMM to the stochastic volatility model defined by (2.7) and (2.8) with \( \mu_t = 0 \). Monte Carlo studies in Andersen and Sørensen (1996) suggest that GMM has poor small-sample performance. This is a generic problem of GMM: its small-sample problems are serious, in particular with high-dimensional weight matrices and strongly dependent moment conditions. References on this topic include an issue of *Journal of Business and Economic Statistics* (1996), Vol. 14, No. 3. In Andersen and Sørensen (1996) it is argued that the performance of the GMM approach largely depends on the choices that need to be made in the implementation stage.

GMM boils down to the choice of several sample moments; at least as many as the number of parameters to be estimated. The convergence of these sample moments to their unconditional expected values is used to determine the estimators.

Let \( m_t(\theta) = (m_{1t}(\theta), \ldots, m_{qt}(\theta)) \) denote a \( q \)-vector of selected functions of \( \theta \) and the data at time \( t \). The sample moments of these functions are \( M_T(\theta) = \)

\(^{26}\)Also known as acceleration techniques; see Ross (1990, Ch. 8).
(M_{1T}(0), \ldots, M_{qT}(0)), \text{ where}

\begin{equation}
M_{iT}(\theta) = \frac{1}{T} \sum_{t=1}^{T} m_{it}(\theta), \quad i = 1, \ldots, q,
\end{equation}

(2.42)

The vector of population moments is given by \( E[m_t(\theta)] \). The GMM-estimator is defined as

\begin{equation}
\hat{\theta}_T = \underset{\theta \in \Theta}{\arg \min} (E[m_T(\theta)] - M_T(\theta))' \Lambda_T^{-1}(E[m_T(\theta)] - M_T(\theta))
\end{equation}

(2.43)

where the metric is determined by a p.d. weighting matrix \( \Lambda_T \). Hansen (1982) shows that under some conditions

\begin{equation}
T^{1/2}(\hat{\theta}_T - \theta_0) \overset{d}{\to} N(0, \Omega)
\end{equation}

(2.44)

where \( \Omega \) depends on \( \Lambda_T \). In order to minimise the asymptotic covariance one can choose \( \Lambda_T^{-1} \) as a sequence converging to the inverse of the covariance matrix of the appropriately standardised sample moments:

\begin{equation}
\Lambda = \lim_{T \to \infty} E \left[ \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{T} (m_{it}(\theta) - E[m_t(\theta)])(m_{is}(\theta) - E[m_s(\theta)])' \right]
\end{equation}

(2.45)

This matrix may be estimated by a kernel estimator of the spectral density of the vector of sample moments at frequency zero given an initial estimate \( \hat{\theta} \). The class of kernel estimators of the spectral density matrix is of the general form

\begin{equation}
\hat{\Lambda}_T = \sum_{j=-T+1}^{T-1} k(j) \hat{\Gamma}_T(j)
\end{equation}

(2.46)

where \( k(j) \) are weights that may become zero for \( |j| > L_T \), a lag truncation parameter which grows toward infinity at a lower rate than \( T \) and

\begin{equation}
\hat{\Gamma}_T(j) = \frac{1}{T-j} \sum_{t=j+1}^{T} (E[m_t(\hat{\theta})] - m_t(\hat{\theta}))(E[m_{t-j}(\hat{\theta})] - m_{t-j}(\hat{\theta}))'
\end{equation}

(2.47)

Typically, this estimator depends on \( \hat{\theta} \). A common solution for finding \( \hat{\Lambda}_T \) in the literature on GMM estimation is to start with the identity matrix and to employ an iterative procedure. In the context of SV models Andersen and Sørensen (1996) report that they used three sets of iterations. In the first step they employed a simple estimate of the weighting matrix. In the second and third step they used a kernel weighting matrix. No more steps should be employed because Andersen and Sørensen (1996) also report that in all cases there is only a minor difference between the estimates of the weighting matrix obtained in the second and third step.
Several other choices have to be made by the researcher. He or she must specify the weights $k(j)$, the bandwidth $L_T$ and he or she must address the issue of pre-whitening, see Andrews and Monahan (1992). Andersen and Sørensen (1996) address all these issues for SV models. Their first suggestion is to employ a bandwidth $L_T = 10$, and weights $k(j) = 1 - \frac{j}{L_T}$, for $j \leq L_T$ and $k(j) = 0$ elsewhere. This is the well-known Newey-West estimator, see Newey and West (1987b). For the basic SV model (2.7) and (2.8) with $\mu = 0$, Andersen and Sørensen (1996) find that the choice of the moments is important; we may wonder which moments, $E[y_t^2 y_{t-i}^2]$ (quadratic forms), $E | y_t y_{t-i} |$ (absolute moments) or even $E | y_t y_{t-i}^2 |$ (third-order moments) contain most information. Andersen and Sørensen (1996) report that including the third-order moments seems not a fruitful avenue for improving estimation performance. They also report that absolute moments have a slightly better performance than quadratic forms. A mix of the two seems to perform even better. For the SV model Andersen and Sørensen (1996) also experiment with different numbers of moments, 3, 5, 9, 14 and 24 and conclude that 14 is the best choice.

It should be mentioned that Hansen and Scheinkman (1996) have developed a theory for generating moment conditions for the continuous-time diffusions such as (2.9) and (2.10). This field of research is still in development and promising, but is beyond the scope of this thesis.

**Method of Simulated Moments**

The GMM method has the advantage that it can easily be generalised to the case where the theoretical moments $E[m_t(\theta)]$ are unknown. This will occur for a more substantial class of SV models than the basic SV model. Wiggins (1987), Chesney and Scott (1989), Duffie and Singleton (1993) and Melino and Turnbull (1990) calculate the unknown moments by simulation from the model and then basically use GMM. This is called the Method of Simulated Moments (MSM). Under some conditions one can show that

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, [1 + \frac{1}{S}]\Omega) \tag{2.48}$$

where the simulated moments are based on $S$ simulations of sample paths of the process of sample size $T$, see Gourieroux and Monfort (1996, Chapter 2) for an exposition.

The error that is made by simulation can be made arbitrary small by increasing the number of simulations. The lack of efficiency relative to Maximum Likelihood (ML) remains.

27 Also sometimes called Simulated Method of Moments (SMM).
Efficient Methods of Moments

Gallant and Tauchen (1996b) propose a general method to solve models in which the structural model can be easily used to generate data in order to compute the expectation of a non-linear function given values of the structural parameters. These expectations are used in a GMM-type estimation procedure. Examples of such structural models are general-equilibrium models, auction models and SV models. In Gallant and Tauchen (1996b) and Gallant, Hsieh and Tauchen (1997) a stochastic volatility model is estimated using this technique. Their technique resembles GMM estimation, but here the arbitrary moments are replaced by the scores of some auxiliary model. This auxiliary model is chosen in such a way that one can obtain or at least approach first-order asymptotic efficiency for the resulting estimator. Recent Monte Carlo studies in the context of stochastic volatility models show that the small-sample properties have been very much improved. In this thesis we will use the efficient method of moments technique extensively. Chapter 3 provides a detailed description, Monte Carlo results and applications of EMM.

Indirect Inference

Gouriéroux et al. (1993) use the parameters $\beta$ of the auxiliary model as the dynamic properties of the auxiliary model in their moment-based estimation technique, i.e.

$$\hat{\theta}_T = \arg\min_{\beta \in \Theta} [\hat{\beta}_T - \frac{1}{S} \sum_{i=1}^{S} \hat{\beta}_{iT}(\theta)]^\prime \Omega [\hat{\beta}_T - \frac{1}{S} \sum_{i=1}^{S} \hat{\beta}_{iT}(\theta)]$$

(2.49)

where $\hat{\beta}_T$ is a consistent estimator of a certain auxiliary model using the sample data of length $T$, $\hat{\beta}_{iT}(\theta)$ is a consistent estimator of the same auxiliary model for the $i$th simulated path ($i = 1, \ldots, S$) of length $T$ and $\Omega$ is some symmetric nonnegative matrix. The matrix $\Omega$ can be chosen in an optimal way. Again the efficiency loss from simulation is summarized by a factor $1 + \frac{1}{S}$, as in (2.48), which can be made arbitrary close to zero. The loss of efficiency from choosing arbitrary moment conditions remains.

In a certain way this method resembles the EMM approach, where instead of parameters, the score of the auxiliary model is used for calibration. For given auxiliary model Gouriéroux et al. (1993) prove that EMM and II are asymptotically equivalent. However, unlike the EMM approach, in the Indirect Inference approach the choice of the auxiliary method for a particular structural model is not well motivated. Although, in case a sensible auxiliary model is coupled to the structural model, Indirect Inference may also yield virtually efficient estimates.

An application to the basic SV model (2.7) and (2.8) can be found in Monfaradini (1998). Here the auxiliary models used are a high-order AR models and an
ARMA(1,1) model for ln \( y_t^2 \). The method performs rather well, i.e. somewhat better than the QML approach but less than the recent MCMC methods and EMM. The Indirect Inference approach has the appeal that it is very easy to implement and has a greater generality than likelihood-based methods.

### 2.4 Option Pricing

The theory of option pricing has been developed by Black and Scholes (1973), henceforth BS, and Merton (1973). Bernstein (1992, Ch.11) and Duffie (1998) provide highly readable accounts of the (r)evolution of option pricing theory in the early 1970s. The standard reference on option pricing is Hull (1997). In this section we will, for explanatory purposes, only consider European call options. A European call option is the right to purchase an asset on a specific time, the maturity date or time of expiration \( T \), for a specific exercise price \( K \). The value of this call at time of expiration \( T \), \( C(T) \) is determined by three variables: the maturity date \( T \), the exercise price \( K \) and the value of the stock at time \( T \), \( S(T) \). We may deduce the following expression for \( C(T) \):

\[
C(T) = \max(S(T) - K, 0) \quad (2.50)
\]

At time \( t < T \), \( S(T) \) is unknown. We should therefore proceed to assume \( S(t) \) to follow some sort of stochastic process. In their seminal work BS assume a geometric Brownian motion for the stock price \( S(t) \)

\[
dS(t) = \alpha S(t)dt + \sigma S(t)dB(t) \quad (2.51)
\]

where the volatility \( \sigma \) and the expected rate of return or "mean", \( \alpha \), are assumed constant, and where \( B(t) \) is a standard Brownian motion. BS also assume the instantaneous nominal discount rate, \( r \), to be constant. The solution of the above SDE is

\[
S(t) = S(0) \exp\{(\alpha - \frac{1}{2} \sigma^2) t + \sigma B(t)\} \quad (2.52)
\]

From the properties of the Brownian motion we may deduce that \( S(t) \) follows a log-normal distribution with

\[
E[\ln(S(t)/S(0))] = (\alpha - \frac{1}{2} \sigma^2) t \quad (2.53)
\]

\[
\text{Var}[\ln(S(t)/S(0))] = \sigma^2 t. \quad (2.54)
\]

Using an economic dynamic-equilibrium model BS derive the BS option pricing formula. The distributional assumptions of this model admit an equilibrium-based
derivation of an option pricing formula from the underlying continuous-time capital asset-pricing model. This is the BS option-pricing formula. Without using an equilibrium model Merton (1973) uses an arbitrage argument to derive the same formula. For the arbitrage argument to hold one needs the possibility of continuously changing portfolios and consequently the continuous-time formulation. A similar derivation holds for discrete-time equilibrium-based models, see Rubinstein (1976) and Brennan (1979).

An important concept is risk-neutrality. It turns out that the resulting BS option pricing formula does not depend on the risk preferences of the investors: only non-saturation of the economic agents is required. For this reason the BS formula is said to be preference-free. This admits a simpler derivation of the BS formula than the equilibrium-based derivation, namely a derivation in a risk-neutral world where the probability measure of the asset-price process is modified to a risk-neutral probability measure. Under this risk-neutral probability measure one can show that the resulting diffusion does not involve variables that are affected by the risk preference of investors. The resulting option-pricing formula for European options looks as if investors priced the options at their expected discounted pay-offs under an equivalent risk-neutral representation,

\[ C(0) = E_0^* \exp \left\{ - \int_0^T r(t) dt \right\} \max(S(T) - K, 0) \]  

(2.55)

Here \( E_0^* \) denotes the expectation under the risk-neutral probability at time 0 and \( r(t) \) is the risk-free rate. This insight is due to Cox and Ross (1976) and later formalized by Harrison and Kreps (1979). Note that in the BS formula \( r(t) \) is assumed constant, but for more generality (2.55) includes a time-dependent \( r(t) \). Using the properties of the log-normal distribution of \( S(t) \) and setting \( r(t) = r \) we obtain the famous textbook BS formula, which will not be repeated here; see Merton (1977).

In the risk-neutral process corresponding to the BS formula the “mean” of the stock return equals the risk free rate, i.e. \( \alpha \) is replaced by \( r \).

In case we let some aspect of the model vary as in the Hull-White and Merton model, we loose most of the advantages of the continuous-time processes in the derivation of the pricing formula. Therefore we might work in a discrete-time dynamic-equilibrium setting as well. In a discrete-time equilibrium setting Brennan (1979) shows that Formula (2.55) still holds. Amin and Ng (1993) work out the discrete-time model with stochastic volatility for both \( r(t) \) and \( S(t) \). In the next subsection we will describe our option-pricing formula in discrete time building on the Amin and Ng (1993) model.
2.4.1 Option Pricing in Discrete Time

The uncertainty in the economy presented in Amin and Ng (1993) is driven by a set of random variables at each discrete date. Among them are a random shock to the consumption process, a random shock to the individual stock-price process, a set of systematic state variables that determine the time-varying mean, variance, and covariance of the consumption process and stock returns, and finally a set of stock-specific state variables that determine the idiosyncratic part of the stock-return volatility. The investors' information set at time $t$ is represented by the $\sigma$-algebra $\mathcal{F}_t$ which consists of all available information up to $t$. Thus, in addition to a random noise, the stochastic consumption process is driven by its mean rate of growth and its variance, which are determined by the systematic state variables. In addition to a random noise the stochastic stock-price process is driven by its mean rate of return and its variance, which are determined by both the systematic state variables and idiosyncratic state variables. In other words, the stock-return variance can have a systematic component that is correlated and changes with the consumption variance.

An important key relationship derived under the equilibrium condition is that the variance of consumption growth is negatively related to the interest rate, or the interest rate is a proxy of the systematic volatility factor in the economy. Therefore a larger proportion of systematic volatility implies a stronger negative relationship between the individual stock-return variance and interest rate. Given that the variance and the interest rate are two important inputs in the determination of option prices and that they have the opposite effects on call option values, the correlation between volatility and interest rate will therefore be important in determining the net effect of these two inputs. In Chapter 6 of this thesis, we specify and implement a multivariate SV model of interest rate and stock returns for the purpose of pricing individual stock options.

For this multivariate SV model, that will be made explicit in Chapter 6, Amin and Ng (1993) derive that

$$C_0 = E^*_{t} [S_0 \cdot \Phi(d_1) - K \exp(-\sum_{t=0}^{T-1} r_t) \Phi(d_2)] \quad (2.56)$$

where

$$d_1 = \frac{\ln(S_0/(K \exp(\sum_{t=0}^{T-1} r_t)) + \frac{1}{2} \sum_{i=1}^{T} \sigma_{st})^{1/2}}{\sum_{i=1}^{T} \sigma_{st}}, \quad d_2 = d_1 - \sum_{i=1}^{T} \sigma_{rt};$$

$\Phi(\cdot)$ is the CDF of the standard normal distribution and $\sigma_{st}$ and $\sigma_{rt}$ are the time-varying volatilities of the stock returns and of the interest-rate movements respectively. The expectation is taken with respect to the risk-neutral measure and can be calculated from simulations.
As Amin and Ng (1993) point out, several option-pricing formulae in the available literature are special cases of the above option formula. These include the BS formula with both constant conditional volatility and interest rate, the Hull and White (1987) stochastic-volatility option-valuation formula with constant interest rate, the Baily and Stulz (1989) stochastic-volatility index-option-pricing formula, and the Merton (1973), Amin and Jarrow (1991), and Turnbull and Milne (1991) stochastic-interest-rate option-valuation formula with constant conditional volatility.

2.4.2 Testing Option Pricing Models

As Bates (1996b) points out, fundamental to testing option-pricing models against time-series data is the issue of identifying the relationship between the true process followed by the underlying state variables in the objective measure and the risk-neutral processes implied through option prices in an artificial measure. Representative-agent equilibrium models such as Rubinstein (1976), Brennan (1979), Bates (1988, 1991), and Amin and Ng (1993) among others, indicate that European options that pay off only at maturity are priced as if investors priced options at their expected discounted pay-offs under an equivalent “risk-neutral” representation that incorporates the appropriate compensation for systematic asset, volatility, interest rate, or jump risk. Thus, the corresponding risk-neutral specification of the general model specified by the multivariate SV model involves compensation for various factor risks. Namely, the mean of stock return in the risk-neutral specification will be equal to the risk-free rate and the mean of the interest-rate process as well as the means of the stochastic conditional volatilities for both interest rate movements and stock returns will be adjusted for the interest-rate risk and systematic volatility risk. Standard approaches for pricing systematic volatility risk, interest-rate risk, and jump risk have typically involved either assuming that risk is nonsystematic and therefore has zero premium, or imposing a tractable functional form on the risk premium (e.g. the factor-risk premiums are proportional to the respective factors) with extra (free) parameters to be estimated from observed options prices or bond prices (for interest rate risk).

Note that every option-pricing model has to make at least two fundamental assumptions: efficiency of the markets and the functional forms of the stochastic processes of the underlying asset prices. While the latter assumption identifies the risk factors associated with the underlying asset returns, the former ensures the existence of a market price of risk for each factor that leads to a risk-neutral specification. The joint hypothesis that is tested in Chapter 6 is the underlying model specification is correct and option markets are efficient. If the joint hypothesis holds, the option pricing formula derived from the underlying model under equilibrium should be able to correctly predict option prices. Obviously such a joint hypothesis
is testable by comparing the model-predicted option prices with market-observed option prices. The advantage of the framework that is used in this thesis is that we estimate the underlying model specified in its objective measure, and more importantly, EMM lends us the ability to test whether the model specification is acceptable or not. Test of such a hypothesis, combined with the test of the above joint hypothesis, can lead us to infer whether the option markets are efficient or not, which is one of the most interesting issues in analysing options to both practitioners and academics.
2.4.2 Testing Option Pricing Models

As Bates (1996b) points out, fundamental testing option-pricing models against time-series data is the issue of identifying the relationship between the true process followed by the underlying state variables in the objective measure and the risk-neutral processes implied through option prices in an artificial measure. Representative-agent equilibrium models such as Rubinstein (1976), Duffie (1979), Bates (1988, 1991), and Merton and Ng (1993) among others, indicate that European options have payoffs at maturity are priced as if investors price options as their expected discounted payoffs under an equivalent "risk-neutral" representation that incorporates the appropriate compensation for systematic asset volatility, interest-rate, or other risk. That is, all corresponding risk-neutral specification of the general model reflected by the multidimensional SV model involves compensation for various forms of risk.

Second, the mean of stock return in the risk-neutral specification will be equal to the risk-free rate and the mean of the interest rate process as well as the means of the stochastic conditional volatilities for both interest rate movements and stock return will be adjusted for the interest-rate and systematic volatility risk. Standard approaches for pricing stochastic volatility, interest-rate risk, and jumps have typically involved either assuming that risk is nonsystematic and therefore has zero premium, or imposing an untraceable functional form on the risk premium (e.g., the factor risk premiums are proportional to the respective factors) with extra (true) parameters to be estimated from observed option prices or bond prices for interest rate risk).

Note that every option-pricing model has to make at least two fundamental assumptions: efficiency of the markets and the functional forms of the stochastic processes of the underlying asset prices. While the latter assumption identifies the risk factors associated with the underlying asset returns, the former ensures the existence of a market price of risk for each factor that leads to a risk-neutral specification. The joint hypothesis that is tested in Chapter 6 is the underlying model specification of correlated option markets and efficiency. If the joint hypothesis holds, the option pricing formula derived from the underlying model under consideration should be able to correctly predict option prices. Obviously, such a joint hypothesis...