# Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns

**Andrew C. HARVEY**
Department of Statistics, London School of Economics, London WC2A 2AE, United Kingdom

**Neil SHEPHARD**
Nuffield College, Oxford, OX1 1NF, United Kingdom

A stochastic volatility model may be estimated by a quasi-maximum likelihood procedure by transforming to a linear state-space form. The method is extended to handle correlation between the two disturbances in the model and applied to data on stock returns.

KEY WORDS: Ancillarity; Kalman filter; Leverage; Quasi-maximum likelihood; Stock returns.

The volatility of returns on assets tends to change over time. In the finance literature, such movements are usually modeled by an unobserved stochastic process representing the underlying variance. For theoretical work, especially with derivatives, it is convenient to formulate the models in continuous time. An example is the geometric diffusion models used by Hull and White (1987) in their work generalizing the Black–Scholes option-pricing formula to allow for stochastic volatility. For empirical work, however, it is convenient to use discrete-time data. A simple discrete-time model, due to Taylor (1986), for returns for regularly spaced data is

$$\log p_t = \log p_{t-1} + \beta + \exp(h_t/2)\varepsilon_t,$$
$$E(\varepsilon_t) = 0, \qquad \mathrm{var}(\varepsilon_t) = 1, \qquad t = 1, \ldots, T$$
$$h_{t+1} = \alpha + \phi h_t + \eta_t, \qquad \eta_t \sim \mathrm{ID}(0, \sigma_\eta^2), \quad (1)$$

where $\mathrm{ID}(\cdot,\cdot)$ denotes independently distributed with stated mean and variance.

If in (1) $\varepsilon_t$ is a martingale difference (MD), then even if $\varepsilon_t$ and $\eta_t$ are contemporaneously dependent, $y_t = \log(p_t/p_{t-1}) - \beta$ is also an MD. Notice then that this stochastic volatility (SV) model shares this MD property for $y_t$ with the exponential generalized autoregressive conditional heteroscedasticity (EGARCH) model of Nelson (1991) and the threshold autoregressive conditional heteroscedasticity (ARCH) model of Glosten, Jagannathan, and Runkle (1993). This would not be true, however, if today's price innovation and volatility innovation, $\varepsilon_t$ and $\eta_{t-1}$, are correlated, for then $y_t$ is not an MD sequence. The moments of $y_t$ (when they exist) are the same as when $\varepsilon_t$ and $\eta_t$ are independent. Thus the odd moments are 0, and there is excess kurtosis when $\varepsilon_t$ is normal.

A possible dependence between $\varepsilon_t$ and $\eta_t$ allows the model to pick up the kind of asymmetric behavior that is often found in stock prices (see Campbell and Kyle 1993; Engle and Ng 1993; Nelson 1991; Schwert 1989). In particular, an increase in predicted volatility tends to be associated with falls in the stock price, suggesting a negative correlation between $\varepsilon_t$ and $\eta_t$. Hull and White (1987) emphasized the role of this correlation, suggesting that it can cause quite significant biases in the Black–Scholes pricing equation.

Although stochastic volatility models have considerable attractions, they suffer from the disadvantage that it is not possible to analytically solve the likelihood function

$$f(y;\rho,\beta,\sigma_\eta^2,\phi,\alpha) = \int f(y|h;\rho,\beta) f(h;\sigma_\eta^2,\phi,\alpha)\, dh,$$

a $T$-dimensional integral, nor allow $h_t$ to be easily estimated. This has prompted an explosion of work to tackle this problem. This was surveyed at some length by Shephard (1996). So far there is no fast and simple approach for dealing with estimation and filtering when the disturbances are contemporaneously correlated. This article looks at a modified quasi-maximum likelihood (QML) approach. Even if likelihood-based estimation is eventually carried out, this method still provides the basis for preliminary estimation and testing.

Section 1 reviews the QML approach proposed by Harvey, Ruiz, and Shephard (1994). Section 2 presents our method for incorporating possible dependence between $\varepsilon_t$ and $\eta_t$ into the QML procedure. Section 3 looks at the sampling properties of the proposed estimator, and Section 4 presents some applications to stock-return data.

## 1. QUASI-MAXIMUM LIKELIHOOD ESTIMATION

Model (1) may be written as

$$y_t = \sigma\varepsilon_t e^{h_t/2}, \qquad t = 1, \ldots, T, \qquad (2)$$

by letting $y_t$ denote $\Delta \log p_t - \beta$. It will be assumed that $\varepsilon_t$ is a series of independent, identically distributed random disturbances, symmetrically distributed around 0. The $\sigma$ parameter is a scale factor, which subsumes the effect of the constant term in the autoregression of $h_t$. The parameter $\beta$ is treated as known; issues concerned with its estimation are


Journal of Business & Economic Statistics
October 1996, Vol. 14, No. 4

addressed as a special case of the results given by Harvey and Shephard (1993).

The $h_t$ process is a first-order autoregression

$$h_{t+1} = \phi h_t + \eta_t, \qquad \eta_t \sim \mathrm{ID}(0, \sigma_\eta^2). \tag{3}$$

When $\phi$ is less than 1 in absolute value, $h_t$, and therefore $y_t$, is stationary. Most of our analysis will be concerned with this case, although the methods presented can easily be adapted to handle the situation in which $h_t$ is assumed to follow a random walk (see Harvey et al. 1994). Allowing the $h_t$ to follow a more complex linear process does not affect the basic statistical issues.

Squaring the observations in (2) and taking logarithms gives

$$\log y_t^2 = \log \sigma^2 + h_t + \log \varepsilon_t^2. \tag{4}$$

Alternatively,

$$\log y_t^2 = \omega + h_t + \xi_t, \tag{5}$$

where

$$\omega = \log \sigma^2 + E \log \varepsilon_t^2.$$

This equation is the measurement equation in a linear state-space model in which (3) is the transition equation. The disturbance term $\log \varepsilon_t^2$ has a mean and variance that depend on the distribution of $\varepsilon_t$. Thus, if $\varepsilon_t$ is standard normal, the mean is minus 1.270 and the variance is 4.934 (see Abramowitz and Stegun 1970, p. 943). The disturbance $\xi_t$ in (5) is defined so as to have zero mean. Even if $\eta_t$ and $\varepsilon_t$ are correlated, the disturbances in the linear state-space form are uncorrelated if the joint distribution of $\varepsilon_t$ and $\eta_t$ is symmetric; that is, $f(\varepsilon_t, \eta_t) = f(-\varepsilon_t, -\eta_t)$ (see Harvey et al. 1994).

The "unrestricted" QML estimators of the parameters $\phi, \sigma_\eta^2, \omega$ and the variance of $\xi_t, \sigma_\xi^2$, are obtained by treating $\xi_t$ and $\eta_t$ as though they were normal and maximizing the prediction-error decomposition form of the likelihood obtained via the Kalman filter. As noted by Harvey et al. (1994), the QML estimators are asymptotically normal with covariance matrix given by applying the theory of Dunsmuir (1979, p. 502). This assumes that $\eta_t$ and $\xi_t$ have finite fourth moments and that the parameters are not on the boundary of the parameter space.

If a normal or $t$ distribution is assumed for $\varepsilon_t$, it is no longer necessary to estimate $\sigma_\xi^2$, as it is known, and the scale factor, $\sigma^2$, can be obtained from the estimator of $\omega$. We will refer to this as "restricted" QML.

## 2. ESTIMATION OF CORRELATION

We now turn to the question of how our estimation procedures may be "modified" to take account of the possible dependence between the disturbances $\varepsilon_t$ and $\eta_t$, while retaining the assumption that they are independent of each other in different time periods. We will assume that the joint distribution of $\varepsilon_t$ and $\eta_t$ is symmetric. Under these conditions, as we have noted, the state-space form, (5) and (3), still has uncorrelated disturbances because the information on the dependence between $\varepsilon_t$ and $\eta_t$ is lost when observations are squared. We will show, however, that the information may be recovered by carrying out inference conditional on the signs of the observations. For expositional reasons, it will be assumed that $h_t$ is stationary.

### 2.1 A Criterion Function

Let $\rho$ denote any parameters that appear in the joint distribution of $\varepsilon_t$ and $\eta_t$ and are 0 when they are mutually independent. Then write $\theta = (\sigma_\xi^2, \sigma_\eta^2, \rho, \phi)'$ so that the joint density of the observations can be expressed as

$$f(y_1, \ldots, y_T; \theta) = \prod_{t=1}^{T} f(y_t | Y_{t-1}; \theta),$$

where $Y_{t-1}$ denotes the information available at time $t-1$. If $s_t$ is the sign of $y_t$—that is, $s_t$ is 1 $(-1)$ if $y_t$ is positive (negative)—the joint density may be factorized as

$$f(y_1, \ldots, y_T; \theta) = \prod_{t=1}^{T} f(y_t | s_t, Y_{t-1}; \theta) f(s_t | Y_{t-1}; \theta).$$

Because the $\varepsilon_t$ are serially independent, the distribution of $s_t$ conditional on $Y_{t-1}$ is just the marginal distribution of $s_t$. Furthermore, if $\varepsilon_t$ is symmetric around 0, the distribution of $s_t$ does not depend on $\theta$, so $f(s_t | Y_{t-1}; \theta) = f(s_t) = .5$. This implies that all that is relevant for inference on $\theta$ is the distribution of the observations conditional on the (observed) signs. That is, the signs are ancillary statistics (see Barndorff-Nielsen and Cox 1994, chap. 4). Because the observation can be obtained from its absolute value and its sign, the distribution of $|y_t|$ conditional on $s_t$ is also a valid basis for inference; that is,

$$\begin{aligned} f(y_t | Y_{t-1}; \theta) &= f(y_t | s_t, Y_{t-1}; \theta) f(s_t) \\ &= f(|y_t| | s_t, Y_{t-1}; \theta) f(s_t), \end{aligned}$$

so an appropriate criterion function can be formed by using the conditional density

$$\prod_{t=1}^{T} f(|y_t| | s_t, Y_{t-1}; \theta).$$

When $\varepsilon_t$ and $\eta_t$ are independent, the distribution of $y_t$ conditional on $s_t$ is the same as the marginal distribution of the absolute values of the observations. Thus, $f(|y_t| | s_t, Y_{t-1}; \theta) = f(|y_t| | Y_{t-1})$. This was the situation addressed by Harvey et al. (1994) in their study of exchange-rate volatility. If we ignore the role of the signs of the observations in the more general case when $\rho$ is not 0, we not only lose all the information on $\rho$ but also impair our inferences on other elements in $\theta$.

### 2.2 Implementation

It follows from the preceding that, even though we are not proposing estimation by maximum likelihood, the state-space form, being based on absolute values of $y_t$, is a sensible basis for inference when we condition on the signs of the observations. These signs are, of course, the same as the signs of the $\varepsilon_t$'s. Let $E_+ (E_-)$ denote the expectation taken conditional on $\varepsilon_t$ being positive (negative), and assign a similar interpretation to variance and covariance operators. The distribution of $\xi_t$ is not affected by conditioning on the signs of the $\varepsilon_t$'s, but, remembering that $E(\eta_t|\varepsilon_t)$ is an odd function of $\varepsilon_t, \mu^* = E_+(\eta_t) = E_+[E\eta_t|\varepsilon_t] = -E_-(\eta_t)$, and $\gamma^* = \text{cov}_+(\eta_t, \xi_t) = E_+(\eta_t\xi_t) - E_+(\eta_t)E(\xi_t) = E_+(\eta_t\xi_t) = -\text{cov}_-(\eta_t\xi_t)$ because the expectation of $\xi_t$ is 0 and $E_+(\eta_t\xi_t) = E_+[E(\eta_t|\varepsilon_t)\log\varepsilon_t^2] - \mu^*E(\log\varepsilon_t^2) = -E_-(\eta_t\xi_t)$. Finally $\text{var}_+\eta_t = E_+(\eta_t^2) - [E_+(\eta_t)]^2 = \sigma_\eta^2 - \mu^{*2}$. The linear state-space form is now

$$\log y_t^2 = \omega + h_t + \xi_t$$

$$h_{t+1} = \phi h_t + s_t\mu^* + \eta_t^*$$

$$\left.\begin{pmatrix}\xi_t\\ \eta_t^*\end{pmatrix}\right| s_t \sim \text{ID}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_\xi^2 & \gamma^* s_t\\ \gamma^* s_t & \sigma_\eta^2 - \mu^{*2}\end{pmatrix}\right). \quad (6)$$

Although some Kalman-filter algorithms are written for state-space models with correlated measurement and transition errors, it may be more convenient to adopt the corresponding form with uncorrelated measurement and transition equation disturbances. This is

$$\log y_t^2 = \omega + h_t + \xi_t$$

$$h_{t+1} = \left(\phi - \frac{\gamma^* s_t}{\sigma_\xi^2}\right) h_t + s_t\left\{\mu^* + \frac{\gamma^*}{\sigma_\xi^2}(\log y_t^2 - \omega)\right\} + \eta_t^+$$

$$\left.\begin{pmatrix}\xi_t\\ \eta_t^+\end{pmatrix}\right| s_t \sim \text{ID}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_\xi^2 & 0\\ 0 & \sigma_\eta^2 - \mu^{*2} - \left(\frac{\gamma^{*2}}{\sigma_\xi^2}\right)\end{pmatrix}\right). \quad (7)$$

The state may still be initialized by taking $h_0$ to have mean 0 and variance $\sigma_\eta^2/(1-\phi^2)$.

*2.2.1 Filtered Volatility Estimate.* The filtered estimate of the log volatility $h_t$, written as $h_{t+1|t}$, takes the form

$$h_{t+1|t} = \phi h_{t|t-1} + \frac{\phi(p_{t|t-1} + \gamma^* s_t)}{p_{t|t-1} + 2\gamma^* s_t + \sigma_\xi^2} \times (\log y_t^2 - \omega - h_{t|t-1}) + s_t\mu^*,$$

where $p_{t|t-1}$ is the corresponding mean squared error of the $h_{t|t-1}$. If $\gamma^* < 0$, it will behave in a similar way to the EGARCH model estimated by Nelson (1991) (see also Engle and Ng 1993), with negative $y_t$ increasing the sensitivity of the filtered value of $h_t$ to the size of $(\log y_t^2 - \omega - h_{t|t-1}) \approx \log\varepsilon_t^2$. Thus, in this model a large value of $y_t$ will increase the filtered estimate of volatility, but not by as much as a corresponding negative value. Finally, the $s_t\mu^*$ term plays a similar role to the ARCH threshold effect of Glosten et al. (1993).

The parameters $\sigma_\xi^2, \sigma_\eta^2, \phi, \mu^*$, and $\gamma^*$ can be estimated by unrestricted QML via the Kalman filter without any distributional assumptions, apart from the existence of fourth moments of $\eta_t$ and $\xi_t$. That the model is identified can be demonstrated by showing that simple consistent estimators can be constructed from the information set used in (6). Indeed these estimators could be used as starting values in the QML procedure. First, we know from Section 1 that $\sigma_\xi^2, \sigma_\eta^2$, and $\phi$ can be estimated consistently by QML without conditioning on the signs. (They could also be estimated from the autocorrelation function of $\log y_t^2$.) Second, if we condition on signs, the measurement equation in (6) can be written as $w_t = \log y_t^2 - \phi\log y_{t-1}^2 = \omega(1-\phi) + \mu^* s_t + \eta_{t-1}^* + \xi_t - \phi\xi_{t-1}$. A regression of $w_t$ on $s_t$ gives a consistent estimator of $\mu^*$, and the variance of the disturbance term is $\text{var}(\eta_{t-1}^* + \xi_t - \phi\xi_{t-1}) = \sigma_\eta^2 - \mu^{*2} + (1+\phi^2)\sigma_\xi^2 - \phi\gamma^* s_{t-1}$, so a consistent estimator of $\gamma^*$ can be extracted by summing the squared residuals separately over positive and negative $y_{t-1}$'s.

Unrestricted QML estimation therefore enables us to estimate the parameters $\sigma_\xi^2, \sigma_\eta^2, \phi, \mu^*$, and $\gamma^*$. It does not give us an estimate of $\rho$, however. Such an estimate can be constructed by making a distributional assumption about $\varepsilon_t$ as well as $\eta_t$.

*2.2.2 Distributional Assumptions.* When $\varepsilon_t$ and $\eta_t$ are bivariate normal with $\text{corr}(\varepsilon_t, \eta_t) = \rho$, $E(\eta_t|\varepsilon_t) = \rho\sigma_\eta\varepsilon_t$, so

$$\mu^* = E_+(\eta_t) = \rho\sigma_\eta E_+(\varepsilon_t) = \rho\sigma_\eta\sqrt{2/\pi} = .7979\rho\sigma_\eta. \quad (8)$$

Furthermore, using the results in Appendix A,

$$\gamma^* = \rho\sigma_\eta E(|\varepsilon_t|\log\varepsilon_t^2) - .7979\rho\sigma_\eta E(\log\varepsilon_t^2) = 1.1061\rho\sigma_\eta. \quad (9)$$

A similar type of argument can be used if a $t$ distribution is assumed instead of a normal one.

These parameter restrictions give a fully specified model that allows a modified form of restricted QML estimation of the parameters $\sigma, \rho, \phi$, and $\sigma_\eta^2$ of the original model. This is the approach used in our simulations and empirical work. Notice that in this form it is relatively easy to impose on the optimization routine the constraint that $|\rho| < 1$.

## 3. PROPERTIES OF ESTIMATOR

As we saw in Section 1, the results of Dunsmuir (1979, p. 502) can be applied to the basic QML estimator to show that it is consistent and asymptotically normal. Conditioning on the signs of the observations, the $s_t$ of Section 2, complicates matters somewhat in that the covariance between the two disturbances in (7) varies according to $s_t$. Because the $s_t$ are treated as nonstochastic, however, (7) is still a linear state-space form but no longer a time-invariant one. An argument along the lines of Ljung and Caines (1979) can be invoked to demonstrate consistency and asymptotic normality; that is,

$$\sqrt{T}(\hat\theta - \theta) \xrightarrow{d} N(0, E^{-1}DE^{-1}),$$

where

$$\frac{1}{T}\frac{\partial^2\log L}{\partial\theta\,\partial\theta'} \xrightarrow{p} E > 0, \quad \frac{1}{\sqrt{T}}\frac{\partial\log L}{\partial\theta} \xrightarrow{d} N(0, D).$$

Table 1. Simulations for the Modified QML Estimator

| Parameter values | Sample size 500 | 1,000 | 3,000 | 6,000 |
|---|---|---|---|---|
| $\phi = .995$ | .973 (.071) | .989 (.029) | .994 (.004) | .994 (.002) |
| $\log\sigma_\eta^2 = -5.298$ | −4.954 (1.201) | −5.257 (.805) | −5.313 (.443) | −5.303 (.292) |
| $\rho = .000$ | .006 (.496) | .005 (.335) | .006 (.167) | .003 (.110) |
| $\phi = .995$ | .971 (.077) | .986 (.042) | .994 (.003) | .995 (.002) |
| $\log\sigma_\eta^2 = -5.298$ | −4.962 (1.210) | −5.175 (.852) | −5.309 (.424) | −5.317 (.279) |
| $\rho = -.300$ | −.339 (.450) | −.308 (.298) | −.312 (.153) | −.304 (.105) |
| $\phi = .995$ | .986 (.043) | .993 (.006) | .995 (.002) | .995 (.001) |
| $\log\sigma_\eta^2 = -5.298$ | −5.186 (.832) | −5.327 (.530) | −5.339 (.291) | −5.315 (.191) |
| $\rho = -.900$ | −.906 (.182) | −.913 (.124) | −.911 (.068) | −.907 (.046) |
| $\phi = .975$ | .917 (.142) | .943 (.098) | .969 (.028) | .973 (.011) |
| $\log\sigma_\eta^2 = -4.605$ | −4.223 (1.643) | −4.387 (1.267) | −4.582 (.727) | −4.618 (.474) |
| $\rho = .000$ | .042 (.497) | −.010 (.325) | −.008 (.153) | .004 (.106) |
| $\phi = .975$ | .923 (.133) | .948 (.088) | .970 (.028) | .973 (.009) |
| $\log\sigma_\eta^2 = -4.605$ | −4.302 (1.549) | −4.437 (1.185) | −4.590 (.692) | −4.604 (.432) |
| $\rho = -.300$ | −.337 (.465) | −.335 (.298) | −.304 (.156) | −.303 (.101) |
| $\phi = .975$ | .954 (.072) | .968 (.034) | .974 (.007) | .975 (.005) |
| $\log\sigma_\eta^2 = -4.605$ | −4.446 (1.044) | −4.596 (.708) | −4.617 (.353) | −4.612 (.249) |
| $\rho = -.900$ | −.896 (.208) | −.909 (.132) | −.911 (.079) | −.907 (.058) |
| $\phi = .995$ | .977 (.063) | .990 (.016) | .994 (.003) | .994 (.002) |
| $\log\sigma_\eta^2 = -4.605$ | −4.474 (1.054) | −4.613 (.656) | −4.609 (.321) | −4.610 (.216) |
| $\rho = .000$ | .013 (.418) | .002 (.260) | −.007 (.140) | .006 (.095) |
| $\phi = .995$ | .981 (.051) | .990 (.020) | .994 (.003) | .994 (.002) |
| $\log\sigma_\eta^2 = -4.605$ | −4.502 (1.026) | −4.597 (.670) | −4.608 (.325) | −4.604 (.213) |
| $\rho = -.300$ | −.323 (.401) | −.315 (.254) | −.305 (.124) | −.302 (.090) |
| $\phi = .995$ | .989 (.028) | .994 (.004) | .995 (.002) | .995 (.001) |
| $\log\sigma_\eta^2 = -4.605$ | −4.620 (.717) | −4.669 (.438) | −4.623 (.223) | −4.613 (.151) |
| $\rho = -.900$ | −.913 (.154) | −.914 (.098) | −.907 (.055) | −.903 (.037) |

Some simulation results for the QML estimator are given in Table 1. It is important to note that they are constructed under the normality assumption using Model (7) and the moment constraints (8) and (9). Finally, $h_t$ is assumed stationary, and the Kalman filter is initialized using the unconditional distribution for $h_0$. This setup will also be used in all our empirical work.

The selection of the parameters for the simulation was made so as to be empirically reasonable. We report values for the estimates of $\log\sigma_\eta^2$ instead of $\sigma_\eta^2$ because they are closer to being normally distributed. The numbers in brackets are root mean squared errors, not standard errors. Table 1 was computed using 1,000 replications, NAG random-number generation, and the E04JAF numerical optimization routine, which hill-climbs without analytic derivatives. Estimates of the $\phi$ parameter were constrained to be between .5 and 1.0. This was to ensure that the root mean squared error estimates were not overly influenced by occasional outliers. In addition, estimates of $\log\sigma_\eta^2$ were constrained to be bigger than $-7$ for the estimates of $\rho$ to be meaningful.

The table indicates that, as $\rho$ falls, the estimates of the other parameters become more precise. The mean squared errors indicate that the procedure can give useful estimates of $\rho$ provided the sample size is bigger than 1,000. Smaller sizes are likely to be rather uninformative.

Table 2 presents the corresponding results for the basic QML estimator of Section 1 when $\rho$ is 0. Its performance is unaffected when $\rho$ is not 0 because, as we have seen, the state-space form in (5) is independent of $\rho$. The main point about Table 2 is that it demonstrates that the use of the modified procedure appears to lead to no loss in efficiency

Table 2. Simulations for the Restricted QML Estimator, With $\rho$ Set to Zero

| Parameter values | Sample size 500 | 1,000 | 3,000 | 6,000 |
|---|---|---|---|---|
| $\phi = .995$ | .969 (.084) | .986 (.042) | .994 (.004) | .994 (.002) |
| $\log\sigma_\eta^2 = -5.298$ | −4.918 (1.282) | −5.191 (.879) | −5.311 (.426) | −5.301 (.274) |
| $\phi = .975$ | .913 (.147) | .944 (.099) | .967 (.039) | .973 (.011) |
| $\log\sigma_\eta^2 = -4.605$ | −4.141 (1.689) | −4.420 (1.268) | −4.562 (.740) | −4.610 (.464) |
| $\phi = .995$ | .978 (.058) | .991 (.013) | .994 (.003) | .994 (.002) |
| $\log\sigma_\eta^2 = -4.605$ | −4.493 (1.034) | −4.597 (.626) | −4.616 (.318) | −4.598 (.221) |

Table 3. SV Model Fitted to Four Financial Series

| Dataset | Parameters $\hat{\phi}$ | $\widehat{\log \sigma_\eta^2}$ | $\hat{\rho}$ | Box–Ljung statistics $\hat{u}_t$ | $\hat{u}_t^2$ | $\log \hat{u}_t^2$ | $\tilde{\nu}_t$ | $\tilde{\nu}_t^2$ | log L |
|---|---|---|---|---|---|---|---|---|---|
| CRSP | .9877 | −4.13 | −.66 | 37.8 | 812 | 1,523 | 21.8 | 18.3 | −8,455.9 |
| (SE) | (.0033) | (.29) | (.05) | | | | | | |
| | .9919 | −4.51 | .00 | 37.8 | 812 | 1,523 | 21.3 | 17.4 | −8,500.6 |
| | (.0045) | (.49) | (.15) | | | | | | |
| ATT | .9978 | −6.05 | −.49 | 28.5 | 382 | 215 | 17.8 | 26.9 | −3,339.8 |
| (SE) | (.0016) | (.45) | (.19) | | | | | | |
| | .9956 | −5.80 | .00 | 28.5 | 382 | 215 | 17.6 | 25.6 | −3,342.7 |
| | (.0026) | (.54) | (.20) | | | | | | |
| GM | .9969 | −5.84 | −.43 | 32.2 | 279 | 199 | 20.9 | 7.9 | −3,511.3 |
| (SE) | (.0023) | (.59) | (.23) | | | | | | |
| | .9953 | −5.55 | .00 | 32.2 | 279 | 199 | 21.1 | 8.1 | −3,513.8 |
| | (.0031) | (.62) | (.19) | | | | | | |
| IBM | .9975 | −6.33 | −.26 | 30.1 | 209 | 212 | 13.2 | 15.9 | −3,344.1 |
| (SE) | (.0015) | (.43) | (.21) | | | | | | |
| | .9969 | −6.16 | .00 | 30.1 | 209 | 212 | 13.4 | 16.1 | −3,345.0 |
| | (.0018) | (.44) | (.20) | | | | | | |

NOTE: The Box–Ljung statistics were computed based on 20 lags. The prefiltered returns, taking out an intercept and an autoregressive term, are denoted by $\hat{u}_t$. The $\hat{\nu}_t$ denote the scaled one-step-ahead forecast errors from the Kalman filter, which has been run on the log $\hat{u}_t^2$. As a guide to statistical significance, we compare the Box–Ljung statistic for the errors with a $\chi^2_{18}$ distribution, though it should be noted that the standard regularity conditions do not hold in this case.

when the true value of $\rho$ is 0. It also shows that basic QML entails a loss in efficiency in estimating $\log \sigma_\eta^2$ and $\phi$ when $\rho$ is nonzero.

An interesting feature of the tables is that the root mean squared errors occasionally fall very dramatically when $T$ increases, not at a constant $T^{-.5}$ rate as might be expected. The reason is the reduced probability of very poor estimates of $\phi$ or $\log \sigma_\eta^2$. When $T$ is as small as 500, much of the contribution to the mean squared error comes from occasional outlying estimates, say $\hat{\phi} = .5$ or $\widehat{\log \sigma_\eta^2} = -7.0$. As $T$ increases, the probability of having such estimates falls, falling more rapidly as $\phi$ increases. This partially explains why Ruiz (1994) found that the asymptotic variance considerably underestimated the actual variance in small samples and why Jacquier, Polson, and Rossi (1994) reported such an unfavorable performance for the QML estimator when $\phi$ is around .9 and $T$ is small.

## 4. EMPIRICAL APPLICATIONS

The QML estimation procedure was used to estimate the following SV model with correlation for several datasets:

$$\begin{aligned} y_t &= \mu + \delta y_{t-1} + \sigma \varepsilon_t e^{h_t/2} \\ h_{t+1} &= \phi h_t + \eta_t \\ \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} &\sim \mathrm{NID}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{pmatrix}\right]. \end{aligned}$$

The mean parameter $\mu$ and autoregressive parameter $\delta$ were estimated by ordinary least squares, and we fitted our SV model to the residuals $\hat{u}_t = y_t - \hat{\mu} - \hat{\delta} y_{t-1}$ and used the mean of the $\log \hat{u}_t^2$'s as an estimate of $\omega$. We estimated $\phi, \sigma_\eta^2$, and $\rho$ using a restricted QML estimator.

### 4.1 Nelson's CRSP U.S. Market-Index Data

The first dataset we analyze is the Center for Research in Security Prices (CRSP) daily returns on a value-weighted U.S. market index for July 3, 1962, to December 31, 1987, used by Nelson (1991) to illustrate his EGARCH model. This has a sample size of 6,409.

The estimated parameter values, together with the corresponding fit when $\rho$ is set to 0, are given in Table 3. Figures in brackets are standard errors of the QML estimator. The Box–Ljung statistics use 20 lags. The headings under BL denote the series that is being analyzed, so BL of $\hat{u}_t$ is the Box–Ljung statistic of the residuals from fitting an AR(1) to the raw data. QML was fitted using NAG E04JAF numerical optimization.

The Box–Ljung statistic for $\hat{u}_t$ indicates that there is some linear structure left in the model. This could be removed by fitting a more complex dynamic model to the mean of the process, but experiments along these lines indicate little change to the results reported here.

The squared OLS residuals have a substantial degree of serial correlation, which is increased when we take the logarithmic transformation. The $\tilde{v}_t$ are the standardized innovations, and their Box–Ljung statistic indicates that the fitting of this simple SV model has removed all the important structure in the data. We also record the computed Box–Ljung statistic for $\tilde{v}_t^2$ to allay fears that we have not removed higher-moment dependence in the data.

The high negative estimate, −.62, is consistent with Nelson's findings and suggests that the Black–Scholes option-pricing equation will be quite badly biased. The persistence parameter, $\phi$, gives a half life of about 56 working days.

### 4.2 Scott's SP30 Individual Stock Data

The second set of data is individual (Standard and Poor) SP30 shares from 1974 to 1983 used by Scott (1991). This has a smaller sample size, $T = 2{,}528$. The results for this dataset are given in Table 3.

The diagnostic statistics show the same picture as the Nelson dataset, with serial correlation in the residuals themselves but considerable serial correlation in squares and log squares. Again the SV model seems to remove this structure from the data.

Each of the stocks has high persistence rates of the volatility, with half lives for shocks being around 300 working days. As in the Nelson CRSP dataset, there is negative correlation between $\varepsilon_t$ and $\eta_t$, though it is much less pronounced.

## 5. CONCLUSION

The QML method for estimating the parameters in an SV model is relatively simple. It is based on writing the relevant part of the model as a linear state-space system, and we have shown that it can be modified to handle correlation between the two disturbances by conditioning on signs. Applications to stock-returns data used in the work of Nelson (1991) and Scott (1991) produce plausible empirical results.

## ACKNOWLEDGMENTS

We thank the editor, the associate editor, and the referees for their comments on the previous version of this article. An earlier draft, entitled "The Econometrics of Stochastic Volatility," was presented at Harvard-MIT, INSEE, London School of Economics, Nuffield College, Princeton, and at conferences at Bristol, Paris, Warwick, and Lancaster. We thank Soren Johansen for some useful comments. Finally, we thank the Economic and Social Research Council (ESRC) for financial support.

## APPENDIX: CHI-SQUARED EXPECTATIONS

If $X_v$ has a chi-squared distribution with $v$ df, the expectation of some function of $X_v, g(X_v)$, is given by

$$E[g(X_v)] = \int_0^\infty g(x)\, \frac{x^{(v-2)/2} e^{-x/2}}{2^{v/2}\Gamma(v/2)}\, dx.$$

Thus, using the definition of the gamma function,

$$EX_\nu^a = 2^a\, \frac{\Gamma\left(a + \frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}, \qquad a > -\nu/2. \tag{A.1}$$

If $Y$ is a standard normal variable, this result can be used to evaluate its absolute moments because $E(|Y|^b) = E(X_1^{b/2})$. Furthermore, if $t_\nu$ denotes a $t$ variable with $\nu$ df, we can write it as $t_\nu = Y\nu^{.5}X_\nu^{-.5}$, where $Y$ and $X_v$ are independent, so

$$E|t_\nu|^b = \nu^{b/2} E|Y|^b EX_\nu^{-b/2}, \qquad -1 < b < \nu, \tag{A.2}$$

Finally,

$$EX_\nu^{1/2}\log X_\nu = 2^{1/2} E(\log X_{v+1})\Gamma\left(\frac{\nu+1}{2}\right) \Big/ \Gamma\left(\frac{\nu}{2}\right)$$

and thus takes on the value .0925 when $v = 1$.

*[Received November 1994. Revised November 1995.]*

## REFERENCES

Abramowitz, M., and Stegun, N. C. (1970), *Handbook of Mathematical Functions*, New York: Dover.

Barndorff-Nielsen, O. E., and Cox, D. R. (1994), *Inference and Asymptotics*, London: Chapman and Hall.

Campbell, J. Y., and Kyle, A. S. (1993), "Smart Money, Noise Trading and Stock Price Behaviour," *Review of Economic Studies*, 60, 1–34.

Dunsmuir, W. (1979), "A Central Limit Theorem for Parameter Estimation in Stationary Vector Time Series and its Applications to Models for a Signal Observed With Noise," *The Annals of Statistics*, 7, 490–506.

Engle, R. F., and Ng, V. K. (1993), "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1749–1801. (Reprinted in Engle, R. F. (ed.) (1995), *ARCH: Selected Readings*, Oxford: Oxford University Press, pp. 145–175.)

Glosten, L. R., Jagannathan, R., and Runkle, D. (1993), "Relationship Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1802.

Harvey, A. C., Ruiz, E., and Shephard, N. (1994), "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264. (Reprinted in Engle, R. F. (ed.) (1995), *ARCH: Selected Readings*, Oxford: Oxford University Press, pp. 256–276.)

Harvey, A. C., and Shephard, N. (1993), "Estimation and Testing of Stochastic Variance Models," STICERD Econometrics Discussion Paper, London School of Economics.

Hull, J., and White, A. (1987), "The Pricing of Options on Assets With Stochastic Volatilities," *Journal of Finance*, 42, 281–300.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994), "Bayesian Analysis of Stochastic Volatility Models" (with discussion), *Journal of Business & Economic Statistics*, 12, 371–417.

Ljung, G. M., and Caines, P. E. (1979), "Asymptotic Normality and Prediction Error Estimators for Approximate System Models," *Stochastics*, 3, 29–46.

Nelson, D. B. (1991), "Conditional Heteroskedasticity in Asset Pricing: A New Approach," *Econometrica*, 59, 347–370. (Reprinted in Engle, R. F. (ed.) (1995), *ARCH: Selected Readings*, Oxford, U.K.: Oxford University Press, pp. 87–113.)

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

Schwert, G. W. (1989), "Why Does Stock Market Volatility Change Over Time," *Journal of Finance*, 44, 1115–1153.

Scott, L. (1991), "Random-Variance Option Pricing," *Advances in Future and Options Research*, 5, 113–135.

Shephard, N. (1996), "Statistical Aspects of ARCH and Stochastic Volatility," in *Time Series Models in Econometrics, Finance and Other Fields*, eds. D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielson, London: Chapman & Hall, pp. 1–67.

Taylor, S. J. (1986), *Modelling Financial Time Series*, Chichester, U.K.: John Wiley.