

Unobserved component time series models with ARCH disturbances*



Andrew Harvey, Esther Ruiz, and Enrique Sentana

London School of Economics, London WC2A 2AE, UK

This paper considers how ARCH effects may be handled in time series models formulated in terms of unobserved components. A general model is formulated, but this includes as special cases a random walk plus noise model with both disturbances subject to ARCH effects, an ARCH-M model with a time-varying parameter, and a latent factor model with ARCH effects in the factors. Although the model is not conditionally Gaussian, an approximate filter can be obtained and used as the basis for estimation. The performance of this method is examined on real data sets and Monte Carlo experiments are carried out. The method is extended to handle GARCH and a latent factor model based on the t -distribution.

1. Introduction

The autoregressive conditional heteroscedasticity (ARCH) model was introduced by Engle (1982) and developed further by Bollerslev (1986), who proposed the generalized ARCH, or GARCH, model. This article examines ways in which ARCH and GARCH disturbances may be incorporated in time series models with unobserved components, and the implications this has for estimation. A general model is first set up, and the main issues discussed. Three special cases are then studied in some detail. The first is a simple univariate time series model, consisting of a random walk observed with a white noise error, in which ARCH effects are present in both disturbances.

*This article is based on two earlier papers, one by the first two authors, the other by the third. The Harvey/Ruiz paper was presented at the ARCH conference held at INSEE in Paris in June 1990 and at the World Congress of the Econometric Society held in Barcelona in August 1990. We would like to thank several of the participants for helpful comments. We are also grateful to Neil Shephard, Mervyn King, Sushil Wadhvani, Manuel Arellano, Herman van Dijk, Rob Engle, and several anonymous referees for their comments. In addition we would like to thank Ray Chou, Frank Diebold, and Charles Goodhart for supplying us with the data used in the applications. The second author acknowledges financial support from the Basque Government; the third author acknowledges support from the LSE Financial Markets Group and the Spanish Ministry of Education and Science.

The second is the ARCH-M model with a time-varying parameter, studied by Chou et al. (1992). The third is the latent factor model with ARCH effects in the factors of Diebold and Nerlove (1989). It is also shown how this last model can be handled if it is generalised so as to have disturbances generated by a multivariate Student's t -distribution.

2. The general model

We consider the following model for the $N \times 1$ vector of observations y_t :

$$y_t = Z_t \alpha_t + X_t \beta + \Lambda \varepsilon_t + \varepsilon_t^*, \quad t = 1, \dots, T, \quad (1)$$

where α_t is an $m \times 1$ state vector which evolves according to

$$\alpha_t = T_t \alpha_{t-1} + \Psi \eta_t + \eta_t^*. \quad (2)$$

The matrices Z_t ($N \times m$) and T_t ($m \times m$) may be nonstochastic, but may, more generally, depend on information available at time $t - 1$. The matrix X_t ($N \times k$) is a matrix of observable exogenous variables, while β ($k \times 1$) is a vector of corresponding parameters. The disturbances vectors ε_t^* ($N \times 1$) and η_t^* ($m \times 1$) are normally distributed with mean zero and are serially independent, that is $\varepsilon_t^* \sim \text{NID}(0, H_t^*)$ and $\eta_t^* \sim \text{NID}(0, Q_t^*)$. The ARCH effects are introduced *via* the scalar disturbances, ε_t and η_t , and

$$\varepsilon_t = h_t^{1/2} \varepsilon_t^\dagger \quad \text{and} \quad \eta_t = q_t^{1/2} \eta_t^\dagger,$$

where $\varepsilon_t^\dagger \sim \text{NID}(0, 1)$, $\eta_t^\dagger \sim \text{NID}(0, 1)$, and

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \quad (3)$$

$$q_t = \gamma_0 + \gamma_1 \eta_{t-1}^2. \quad (4)$$

Finally, all the disturbances ε_t , ε_t^* , η_t , and η_t^* are mutually independent.

If the terms involving ε_t and η_t were dropped, we would have a standard linear state space model. Autoregressive, ARIMA, and structural time series models can all be cast in this form. Regression models are also included by virtue of the term $X_t \beta$.

In eqs. (3) and (4) we have first-order ARCH models. The extensions to higher-order ARCH and GARCH models are discussed in subsection 2.4. Another generalization is to have several factors, and the different state variables in (2) driven by different ARCH effects so that ε_t and η_t are vectors and Λ and Ψ are corresponding matrices. This raises no new issues as regards the statistical handling of the model and so we stick to the single

disturbance formulation for simplicity. The identifiability of such models is discussed in Sentana (1991). It is also possible to generalize the model by letting the individual elements of ε_t^* and η_t^* follow ARCH processes.

2.1. A quasi-optimal filter

We now consider a filter for the above model which provides the basis for (a) the estimation of unknown parameters, such as α_0 , α_1 , γ_0 , and γ_1 , and (b) the prediction of future observations. The filter is based on an augmented state space formulation in which the disturbances ε_t and η_t are treated as state variables. The measurement equation is

$$y_t = Z_t \alpha_t + X_t \beta + \Lambda \varepsilon_t + \varepsilon_t^* = [Z_t \ 0 \ \Lambda] \alpha_t^+ + X_t \beta + \varepsilon_t^*, \quad (5)$$

while the transition equation is

$$\alpha_t^+ = \begin{bmatrix} \alpha_t \\ \eta_t \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} T_T & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_{t-1} \\ \eta_{t-1} \\ \varepsilon_{t-1} \end{bmatrix} + \begin{bmatrix} I & \Psi & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_t^* \\ \eta_t \\ \varepsilon_t \end{bmatrix}. \quad (6)$$

Having ε_t and η_t playing the role of both disturbances and state variables is somewhat unusual. Diebold and Nerlove (1989) use the same device in their latent factor model.

If past values of the disturbances ε_t and η_t were directly observable, the model would be a conditionally Gaussian one. The distribution of y_t , conditional on the information available at time $t - 1$ would be normal, and as a result, the Kalman filter would yield minimum mean square estimates (MMSEs) of the state; see, for example, Harvey (1989, ch. 3). As it is, the model is not conditionally Gaussian, since knowledge of past observations does not, in general, imply knowledge of past disturbances. Nevertheless we may proceed on the basis that the model can be treated as though it were conditionally Gaussian, and we will refer to the Kalman filter as being *quasi-optimal*. Of course, if no ARCH effects are present, the model is linear and the filter is optimal.

The distribution of η_t conditional on η_{t-1} is assumed to be normal with mean zero and variance, q_t , given by (4). The optimal filter requires the distribution of η_t conditional on past observations. Since η_{t-1} is not observed, the distribution of η_t , conditional on past observations cannot be determined, but given the conditional mean and variance of the state vector at time $t - 1$, we can evaluate its first two moments. The mean of η_t is easily

seen to be zero. As regards its variance, we have

$$\text{var}_{t-1}(\eta_t) = \text{E}_{t-1}(\eta_t^2) = \gamma_0 + \gamma_1 \text{E}_{t-1}(\eta_{t-1}^2), \quad (7)$$

where the $t - 1$ under the expectation operator indicates that it is conditional on the observations up to and including y_{t-1} . Now

$$\eta_{t-1} = \hat{\eta}_{t-1} + (\eta_{t-1} - \hat{\eta}_{t-1}),$$

and so

$$\text{E}_{t-1}(\eta_{t-1}^2) = \hat{\eta}_{t-1}^2 + p_{t-1}^\eta, \quad (8)$$

where the last term is the variance of η_{t-1} conditional on the information at time $t - 1$. Note that $\hat{\eta}_{t-1}$, the estimate from the Kalman filter, is fixed and known at time $t - 1$ and so the cross-product is zero. Substituting in (7) gives

$$\text{var}_{t-1}(\eta_t) = \gamma_0 + \gamma_1(\hat{\eta}_{t-1}^2 + p_{t-1}^\eta). \quad (9)$$

It is straightforward to show that the distribution of η_t conditional on past observations is symmetric. However, it is not, in general, normal. The same is true for the distribution of ε_t conditional on past observations. Again this has a mean of zero and a conditional variance given by an expression analogous to (9). This expression involves p_{t-1}^ε , which is the conditional variance of ε_{t-1} . We will refer to these conditional variance terms as *correction factors*. When the Kalman filter is applied, the last two diagonal elements in the conditional covariance matrix of the augmented state vector provide values for these factors.

When all the observations have been processed, the filter delivers estimates of the mean, $\hat{\alpha}_T$, and the covariance matrix, p_T , of the conditional distribution of α_T . If the model really were conditionally Gaussian, so that the filter was exact, the MMSE's of future α_t 's would be given by the recursion

$$\hat{\alpha}_{T+l|T} = T_{T+l} \hat{\alpha}_{T+l-1|T}, \quad l = 1, 2, \dots,$$

and so

$$\hat{y}_{T+l|T} = Z_{T+l} \hat{\alpha}_{T+l|T} + X_{T+l} \beta. \quad (10)$$

As regards the MSE, this can be built up recursively, since

$$P_{T+l|T} = T_{T+l} P_{T+l-1|T} T_{T+l}' + Q_t^* + \Psi \Psi' \text{E}_T(\eta_{T+l}^2), \quad l = 1, 2, \dots,$$

and

$$E_T(\eta_{T+l}^2) = \gamma_0 + \gamma_1 E_T(\eta_{T+l-1}^2),$$

with the conditional expectation of η_T^2 at time T given by (8), and so

$$\text{MSE}(\hat{y}_{T+l|T}) = Z_{T+l} P_{T+l|T} Z'_{T+l} + H_{T+l}^* + \Lambda \Lambda' E_T(\varepsilon_{T+l}^2), \quad (11)$$

$$l = 1, 2, \dots$$

The conditional expectation of ε_{T+l}^2 can be built up recursively or evaluated directly as

$$E_T(\varepsilon_{T+l}^2) = \alpha_0 (1 + \alpha_1 + \dots + \alpha_1^{l-1}) + \alpha_1^l (\hat{\varepsilon}_T^2 + p_T^\varepsilon).$$

2.2. Quasi-maximum likelihood estimation

An approximate likelihood function can be constructed from the filter of subsection 2.1 by regarding the distribution of y_t , conditional on previous observations, as being Gaussian with mean $Z_t T_t \hat{\alpha}_{t-1} + X_t \beta$ and variance

$$\begin{aligned} \text{var}_{t-1}(y_t) = F_t = Z_t \left[T_t P_{t-1} T_t' + \Psi \Psi' \text{var}_{t-1}(\eta_t) + Q_t^* \right] Z_t' \\ + \Lambda \Lambda' \text{var}_{t-1}(\varepsilon_t) + H_t^*. \end{aligned} \quad (12)$$

The likelihood function is then formed as the product of the conditional distributions of the y_t 's as in the usual prediction error decomposition. It must be maximized numerically with respect to the unknown parameters; these consist of the ARCH parameters α_0 , α_1 , γ_0 , and γ_1 , together with any parameters appearing in Z_t , H_t^* , T_t , Q_t^* , β , Λ , and Ψ .

2.3. Properties of the model and alternative formulations

Although the ε_t 's are not independent of each other, it follows from standard ARCH theory that they are serially uncorrelated with zero mean and variance $\alpha_0/(1-\alpha_1)$ provided that $\alpha_0 > 0$ and $0 \leq \alpha_1 < 1$. Thus the composite disturbance term $\Lambda \varepsilon_t + \varepsilon_t^*$ is multivariate white noise with covariance matrix $\{\alpha_0/(1-\alpha_1)\} \Lambda \Lambda' + H_t^*$. In a similar way the η_t 's are serially uncorrelated with zero mean and variance $\gamma_0/(1-\gamma_1)$ provided $\gamma_0 > 0$ and $0 \leq \gamma_1 < 1$, and $\Psi \eta_t + \eta_t^*$ is multivariate white noise. If the ARCH effects were ignored, and the Kalman filter were based on the unconditional

covariance matrices of the composite disturbances, it would, in the absence of any unknown parameters, yield minimum mean square *linear* estimators and predictors. Furthermore we would expect any unknown parameters in Z_t , T_t , and β , together with the unconditional covariance matrices of the composite disturbances, to be estimated consistently. This is merely a generalization of a result in Weiss (1984) which shows that the parameters in an ARMA model are estimated consistently, but not efficiently, when ARCH effects in the disturbance term are ignored and the implied ML estimator is computed.

Returning to the fully specified model, we have seen that the Kalman filter is not, in general, optimal and the proposed quasi-ML estimation procedure is only an approximation. However, suppose we were to modify the model by replacing the squared lagged disturbances in (3) and (4) by their conditional expectations. Thus,

$$h_t = \alpha_0 + \alpha_1 E_{t-1}(\varepsilon_{t-1}^2), \quad (13)$$

$$q_t = \gamma_0 + \gamma_1 E_{t-1}(\eta_{t-1}^2). \quad (14)$$

Since the conditional expectations of the squared disturbances are given directly by the Kalman filter, as in (8), h_t and q_t are observable from the information at time $t - 1$ and the model is conditionally Gaussian. The filter and the resulting likelihood function are therefore exact. This is a much happier state of affairs, although one could never distinguish between the two models since they are estimated in exactly the same way.

The disadvantage of (13) and (14) is that standard ARCH theory cannot be used to determine the properties of ε_t and η_t . Nevertheless, it can be shown that ε_t and η_t are serially uncorrelated with zero means and the same variances as for (3) and (4); see appendix. However, the unconditional distribution, although leptokurtic, has fourth moments smaller than the fourth moments of (3) and (4) and the autocorrelations of their squares are not easily derived; see subsection 3.4 for the relevance of these statistics.

A final variation is to assume that estimates of ε_t and η_t based on the information at time $t - 1$ appear in place of the true disturbances in (3) and (4). Thus,

$$h_t = \alpha_0 + \alpha_1 \hat{\varepsilon}_{t-1}^2, \quad (15)$$

and similarly for q_t . The usual assumption would be that $\hat{\varepsilon}_{t-1}$ is the conditional expectation of ε_{t-1} obtained from the Kalman filter, although it is not necessary for it to be defined in this way. In any case the model is conditionally Gaussian by construction, and the Kalman filter yields the exact likelihood.

Once more, the problem with (15) is that standard ARCH results cannot be used. However, it is shown in the appendix that ε_t is still uncorrelated with zero mean but with a smaller unconditional variance, and that its unconditional distribution is leptokurtic, although even less than in the previous case. The squares of the ε_t 's are serially correlated, but the autocorrelation function is now bounded from above by the autocorrelation function of the process generated by (13). The intuition again is that (15) implies an even smoother process for ε_t^2 than does (13). The same obviously applies to η_t .

There are arguments for adopting a specification like that in (15). However, if the true model is (3), then assuming (15), that is not including the correction factor from the state covariance matrix, may yield unsatisfactory results. This matter is investigated in some detail in section 3.

2.4. Higher-order ARCH and GARCH

The ARCH(p) generalisation of (3) is simply

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2, \quad (16)$$

and similarly for (4). The generalized ARCH (GARCH) model of Bollerslev (1986), or at least the GARCH(1, 1) case, has

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 h_{t-1} \quad (17)$$

and

$$q_t = \gamma_0 + \gamma_1 \eta_{t-1}^2 + \gamma_2 q_{t-1}. \quad (18)$$

We will assume that $\alpha_1 + \alpha_2 < 1$ and $\gamma_1 + \gamma_2 < 1$. The integrated GARCH model has the sum of these parameters equal to one; see, for example, Engle and Bollerslev (1986) and Nelson (1990).

The filter requires the variance of the disturbances conditional on previous observations. If ε_t and η_t follow ARCH(p) processes, then p lagged values of each must be included in the state vector. Then analogously to (8),

$$\mathbf{E}_{t-1} (\varepsilon_{t-j}^2) = \hat{\varepsilon}_{t-j|t-1}^2 + p_{t-1}^{(j)}, \quad j = 1, \dots, p, \quad (19)$$

where $\hat{\varepsilon}_{t-j|t-1}$ appears in the estimated state at time $t-1$, and the last term in (19) is its variance. A similar expression holds for the lagged values of η_t .

In the GARCH model,

$$\mathbf{E}_{t-1} (\varepsilon_t^2) = \alpha_0 + \alpha_1 \mathbf{E}_{t-1} (\varepsilon_{t-1}^2) + \alpha_2 \mathbf{E}_{t-1} (h_{t-1}). \quad (20)$$

In order to evaluate the last term on the right-hand side we need to

substitute for h_{t-1} from (17). Doing this repeatedly gives, for $J \geq 1$,

$$\begin{aligned}\text{var}_{t-1}(\varepsilon_t) &= \text{E}_{t-1}(\varepsilon_t^2) \\ &= \alpha_0(1 + \alpha_2 + \cdots + \alpha_2^{J-1}) \\ &\quad + \alpha_1 \sum_{j=1}^J \alpha_2^{j-1} \text{E}_{t-1}(\varepsilon_{t-j}^2) + \alpha_2^J \text{E}_{t-1}(h_{t-j}).\end{aligned}\quad (21)$$

If we are prepared to assume that α_2^J is negligible for a particular value of J , an approximation to (21) may be computed by augmenting the state vector by lagged values of ε_t , and replacing the conditional expectations of the squared ε_{t-j} 's by quantities computed from the right-hand side of (19). We can, in a similar way, compute an approximation to the conditional variance of η_t .

The drawback to the above method is that α_2^J may not be negligible for moderate values of J . An amendment is made by setting $J = \infty$ in (21), multiplying both sides of the expression for $t-1$ by α_2 and subtracting from the expression for t . This yields

$$\begin{aligned}\text{var}_{t-1}(\varepsilon_t) &= \alpha_0 + \alpha_1 \left[\text{E}_{t-1}(\varepsilon_{t-1}^2) + \sum_{j=1}^{\infty} \alpha_2^j \left[\text{E}_{t-1}(\varepsilon_{t-j-1}^2) - \text{E}_{t-2}(\varepsilon_{t-j-1}^2) \right] \right] \\ &\quad + \alpha_2 \text{E}_{t-2}(\varepsilon_{t-1}^2).\end{aligned}\quad (22)$$

Again we can truncate this expression, but we are now neglecting

$$\alpha_1 \sum_{j=J+1}^{\infty} \alpha_2^j \left[\text{E}_{t-1}(\varepsilon_{t-j-1}^2) - \text{E}_{t-2}(\varepsilon_{t-j-1}^2) \right].$$

This is likely to be much smaller than $\alpha_2^J \text{E}_{t-1}(h_{t-j})$, firstly because $\alpha_1 \alpha_2^J < \alpha_2^J$, and secondly because the term in square brackets is likely to be very small. The extra computation, as compared with (21), simply involves storing the quantities in (19) and carrying them over to the next period.

The above filter may be rather cumbersome to implement if J is set to a moderately large value. An alternative approach is therefore to amend the formulation of the structural GARCH model to something which is easier to handle. A natural amendment is to define h_t as

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \text{E}_{t-2}(h_{t-1}).\quad (23)$$

Then

$$\text{E}_{t-1}(\varepsilon_t^2) = \alpha_0 + \alpha_1 \text{E}_{t-1}(\varepsilon_{t-1}^2) + \alpha_2 \text{E}_{t-2}(h_{t-1}).\quad (24)$$

Since

$$E_{t-1}(\varepsilon_t^2) = E_{t-1}(h_t),$$

expression (24) can be evaluated recursively. A similar formulation may be adopted for η_t .

When the model only contains a single disturbance term, (23) is the same as (17) since h_{t-1} is known at time $t - 2$. However, when this is not the case, substituting repeatedly for the expectation of lagged h_t on the right-hand side of (24) points to the contrast between the two formulations. In the analogous expression to (21), the expectation of each ε_{t-j}^2 is taken at time $t - j$, not at time $t - 1$. Note that in terms of (22) expression (24) ignores the summation term within the square brackets.

One disadvantage of (23) is that it is difficult to obtain the autocorrelation function of ε_t^2 , although it can be shown that ε_t is white noise. A more fundamental drawback to the model is that ε_t^2 appears to display a smaller degree of serial dependence than is shown by (17) with the same parameter values. Based on simulation evidence, it appears that although the first-order autocorrelation in (23) is usually only slightly below the one for (17), the higher-order autocorrelations tend to be much smaller. A similar situation arises with the partial autocorrelation function.

3. STARCH models

Unobserved components models are normally set up in such a way that they have a direct interpretation. As a result they are often known as structural time series models; see Harvey (1989). The components may have a purely statistical rationale insofar as their role is to pick up trend, seasonal, and cyclical movements which evolve over time. Alternatively there may be economic reasons behind the model formulation. An example is the 'fads' model in financial markets, where different types of traders give rise to different unobserved components; see, for example, Poterba and Summers (1988). The statistical treatment of linear unobserved component models is relatively straightforward, with the Kalman filter playing a central role. However, the presence of ARCH effects gives rise to the kind of issues discussed in the previous section.

The simplest structural time series model is the local level, or random walk plus noise, model

$$y_t = \mu_t + \varepsilon_t, \tag{25}$$

$$\mu_t = \mu_{t-1} + \eta_t, \tag{26}$$

where y_t , $t = 1, \dots, T$, is the observed series and ε_t and η_t are mutually uncorrelated white noise disturbances. That is they are serially uncorrelated with zero means and variances σ_ε^2 and σ_η^2 , respectively. The reduced form of (25) and (26) is the ARIMA (0, 1, 1) model. Following Weiss (1984), it is possible to set up ARIMA models with ARCH disturbances. However, it is quite possible to envisage situations where it makes sense to consider the transitory disturbance term, ε_t , as being subject to an ARCH mechanism which is independent of the ARCH mechanism for the disturbance, η_t , driving the permanent component. For example, in the context of stock prices Black (1986) has argued that different components in a model, arising from market fundamentals and noisy traders, may be subject to different degrees of volatility. In some situations it may be argued that one of the disturbances does not exhibit ARCH at all. For example if ε_t is interpreted as being a measurement error, there may be no reason for it to be subject to ARCH effects.

Models in which one or more of the disturbances displays ARCH effects will be said to exhibit structural ARCH, or 'STARARCH'. In general such models do not have a reduced form which is an ARIMA model with a (time-invariant) ARCH representation for the disturbance, although, of course, they may be reasonably well approximated by such models.

As discussed in subsection 2.3, there are a number of ways of introducing ARCH effects into models with unobserved components. We will concentrate on the direct formulation in which the variance of a disturbance is related to its own past values as in (3) and (4). This has the attraction that the properties of the disturbances can be obtained from known results on ARCH models, and these properties are briefly set out below. In subsection 3.3 we examine the performance of the quasi-optimal filter in a series of Monte Carlo experiments. We also suggest ways of testing for ARCH effects in one or both of the disturbances. The discussion throughout is in terms of the local level model, (25) and (26), but extensions to more complicated structural time series models, and models with explanatory variables, are relatively straightforward.

3.1. Properties of the local level with ARCH

Even with ARCH disturbances, (3) and (4), model (25) and (26) still has white noise disturbances. The first differences of the observations, that is the Δy_t 's, are stationary with mean zero and first-order autocorrelation

$$\rho(1) = \frac{-\sigma_\varepsilon^2}{\sigma_\eta^2 + 2\sigma_\varepsilon^2} = \frac{-\alpha_0/(1-\alpha_1)}{\{\gamma_0/(1-\gamma_1) + 2\alpha_0/(1-\alpha_1)\}}, \quad (27)$$

and $\rho(\tau) = 0$ for $\tau \geq 2$.

From standard ARCH theory, the fourth moments of the ε_t and η_t disturbances exist if $\alpha_1, \gamma_1 < 0.578$. The autocorrelation functions (a.c.f.'s) of ε_t^2 and η_t^2 then behave like the a.c.f.'s of first-order autoregressive processes, with α_1 and γ_1 playing the roles of the autoregressive parameters. The autocovariance function of $(\Delta y_t)^2$ is

$$\begin{aligned} \text{var}[(\Delta y_t)^2] &= (\gamma_0/(1-\gamma_1))^2 \left(\frac{2}{1-3\gamma_1^2} \right) + 4 \left(\frac{\alpha_0}{1-\alpha_1} \right)^2 \\ &\quad \times \left(\frac{2-3\alpha_1^2+3\alpha_1}{1-3\alpha_1^2} \right) + 8 \frac{\gamma_0 \alpha_0}{(1-\gamma_1)(1-\alpha_1)}, \end{aligned} \quad (28)$$

$$\text{cov}[(\Delta y_t)^2, (\Delta y_{t-1})^2] = \frac{2\gamma_0^2\gamma_1}{(1-\gamma_1)^2(1-3\gamma_1^2)} + \frac{2\alpha_0^2(1+\alpha_1)^2}{(1-\alpha_1)^2(1-3\alpha_1^2)}, \quad (29)$$

and, for $\tau > 1$,

$$\begin{aligned} \text{cov}[(\Delta y_t)^2, (\Delta y_{t-\tau})^2] &= \alpha_1 \text{cov}[(\Delta y_t)^2, (\Delta y_{t-\tau+1})^2] \\ &\quad + (\gamma_1 - \alpha_1) \text{cov}[\eta_t^2, \eta_{t-\tau+1}^2]. \end{aligned} \quad (30)$$

When $\gamma_1 = \alpha_1 = 0$, all the autocorrelations at lags greater than one are zero, while the autocorrelation at lag one is

$$\left(\frac{\alpha_0}{\gamma_0 + 2\alpha_0} \right)^2 = [\rho(1)]^2, \quad (31)$$

where $\rho(1)$ is the lag one autocorrelation of the Δy_t 's defined in (27). This is just a special case of the result in Maravall (1983), which states that, for a linear process, the autocorrelation function of the squared observations is equal to the square of the autocorrelation function of the observations themselves.

The autocorrelation functions of the squared Δy_t 's for higher-order ARCH models and GARCH models can be obtained by using the results of Milhoj (1985) and Bollerslev (1986), respectively.

3.2. The filter

We now consider some special features of the quasi-optimal filter for the local level STARCh model. The state space model of section 2 can be

specialised to

$$y_t = \mu_t + \varepsilon_t, \quad (32)$$

$$\alpha_t^+ = \begin{bmatrix} \mu_t \\ \eta_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \eta_{t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \eta_t. \quad (33)$$

Notice that in terms of eq. (6) we have set Ψ equal to one.

The conditional variance of η_t is given by expression (9). However, there is no need to include ε_t in the state vector. From (3),

$$E_{t-1}(\varepsilon_t^2) = \alpha_0 + \alpha_1 E_{t-1}(\varepsilon_{t-1}^2),$$

but

$$\varepsilon_{t-1} = y_{t-1} - \mu_{t-1} = (\hat{\varepsilon}_{t-1}) + (\hat{\mu}_{t-1} - \mu_{t-1}),$$

where $\hat{\mu}_{t-1}$ is the estimate of the mean of the conditional distribution of μ_{t-1} at time $t-1$, and

$$\hat{\varepsilon}_{t-1} = y_{t-1} - \hat{\mu}_{t-1}.$$

Thus

$$E_{t-1}(\varepsilon_{t-1}^2) = \hat{\varepsilon}_{t-1}^2 + p_{t-1},$$

where p_{t-1} is the estimated conditional variance of μ_{t-1} . Hence

$$\text{var}_{t-1}(\varepsilon_t) = \alpha_0 + \alpha_1(\hat{\varepsilon}_{t-1}^2 + p_{t-1}). \quad (34)$$

As regards initialization of the filter, set $\hat{\mu}_1 = y_1$, with $\text{MSE}(\hat{\mu}_1) = \alpha_0/(1 - \alpha_1)$, that is, the unconditional variance of ε_t . If the conditional variance of η_t at time $t=1$ is also set equal to its unconditional variance, the Kalman filter may be applied with

$$\begin{aligned} \text{var}_{t=1}(\varepsilon_2) &= E_1(\varepsilon_2^2) = \alpha_0 + \alpha_1 E_1(\varepsilon_1^2) \\ &= \alpha_0 + \alpha_1(\alpha_0/(1 - \alpha_1)) = \alpha_0/(1 - \alpha_1) \end{aligned} \quad (35)$$

and

$$\text{var}_{t=1}(\eta_2) = \gamma_0/(1 - \gamma_1). \quad (36)$$

A switch to the augmented filter is then made for $t \geq 3$.

The general method of handling higher-order ARCH and GARCH models is to augment the state vector by past values of ε_t and η_t . However, for our

univariate STAR model, it is more efficient to augment by $\mu_{t-1}, \dots, \mu_{t-J}$, where J depends on the order of the ARCH model, or the approximation employed in GARCH. Estimates of both sets of disturbances and their associated correction factors may then be made since

$$\hat{\varepsilon}_{t-j|t-1} = y_{t-j} - \hat{\mu}_{t-j|t-1} \quad (37)$$

and

$$\hat{\eta}_{t-j|t-1} = \hat{\mu}_{t-j|t-1} - \hat{\mu}_{t-j-1|t-1}, \quad j = 1, \dots, J. \quad (38)$$

3.3. Small sample simulation of the quasi-maximum likelihood estimator

An approximate likelihood function can be constructed from the filter in the way described in subsection 3.2. This quasi-likelihood function must be maximized numerically with respect to the unknown parameters $\alpha_0, \alpha_1, \gamma_0$, and γ_1 . However, one of the parameters, say α_0 , can be concentrated out of the likelihood function by reparameterizing (3) and (4) as

$$h_t = \alpha_0(1 + \alpha_1' \varepsilon_{t-1}^2) \quad (39)$$

and

$$q_t = \alpha_0 \gamma_0' (1 + \gamma_1' \eta_{t-1}^2). \quad (40)$$

Table 1 shows the RMSE's of the quasi-maximum likelihood estimators of the parameters of a random walk plus noise model with ARCH(1) disturbances based on the results of a set of Monte Carlo experiments each with

Table 1

RMSE's of quasi-maximum likelihood estimation of parameters in random walk plus noise model with ARCH(1) disturbances, based on Monte Carlo experiments with 1000 replications.

Parameter	Sample size (T)				Naive estimation for $T = 3000$
	150	500	1000	3000	
$\alpha_0 = 1$	0.488	0.335	0.257	0.169	0.288
$\alpha_1 = 0.3$	0.279	0.226	0.184	0.123	0.192
$\gamma_0 = 1$	0.614	0.373	0.287	0.199	0.372
$\gamma_1 = 0.5$	0.330	0.218	0.165	0.103	0.191
$\alpha_0 = 1$	0.526	0.369	0.301	0.204	0.279
$\alpha_1 = 0.3$	0.304	0.258	0.219	0.157	0.302
$\gamma_0 = 1$	0.785	0.423	0.313	0.222	0.741
$\gamma_1 = 0.8$	0.348	0.175	0.121	0.074	0.127
$\alpha_0 = 1$	0.567	0.348	0.252	0.149	0.473
$\alpha_1 = 0.5$	0.316	0.211	0.151	0.088	0.142
$\gamma_0 = 1$	0.516	0.372	0.315	0.240	0.218
$\gamma_1 = 0.3$	0.308	0.257	0.218	0.163	0.290

1000 replications. The numerical optimization was carried out using the NAG subroutine E04JBF with α_1 and γ_1 restricted to be between 0 and 1, and α_0 and γ_0 restricted to be nonnegative. As one would hope, the RMSE's tend to get smaller as the sample size increases. For small sample sizes, the variation in the estimates can be quite considerable, with the distributions being far from normality. This can be seen quite clearly in fig. 1a, which shows histograms for $T = 150$. Fig. 1b shows the corresponding distributions for $T = 3000$, and it can be seen that they are not too far from normality. The conclusion would therefore seem to be that the proposed estimation procedure has reasonable statistical properties, not unlike those of maximum likelihood. However, it is also clear that fairly large sample sizes are needed to estimate the parameters in STARCH models with a reasonable degree of precision.

The last column in table 1 shows the results from the 'naive' estimation procedure in which the correction terms are omitted from expressions such as (34). As can be seen the resulting estimations are markedly inferior to those produced by quasi-maximum likelihood.

3.4. Testing

We now consider the question of testing for structural ARCH and GARCH effects when a linear time series model has been fitted. A test for ARCH in the innovations of the model, that is the reduced form disturbances, can be carried out using the Box-Ljung statistic formed from the autocorrelations of the squares of residuals. If the first P autocorrelations are used, the Box-Ljung statistic will have a chi-square distribution with P degrees of freedom under the null hypothesis of no ARCH; see McLeod and Li (1983). Such a test would be the Lagrange multiplier (LM) test against an ARCH(P) process in the reduced form disturbance. It may well have good power against ARCH in both the measurement and transition equations, (25) and (26).

An alternative approach is to specifically construct tests against STARCH effects by setting up tests based on the LM principle, treating the quasi-likelihood as if it were exact. A test of the hypothesis $H_0: \alpha_1 = \gamma_1 = 0$ may be constructed as may tests that either of these parameters is zero. The test statistics can be set up in terms of the uncentred coefficient of determination, R^2 , of a regression involving the innovation, its variance, and their respective derivatives, all of them evaluated under the null hypothesis; see Harvey (1989, pp. 240–241). Although the tests cannot be expressed in a particularly simple way, it turns out that the regressions are trying to link movements in the squared innovations to movements in either or both of $\hat{\varepsilon}_t^2$ and $\hat{\eta}_t^2$, as appropriate.

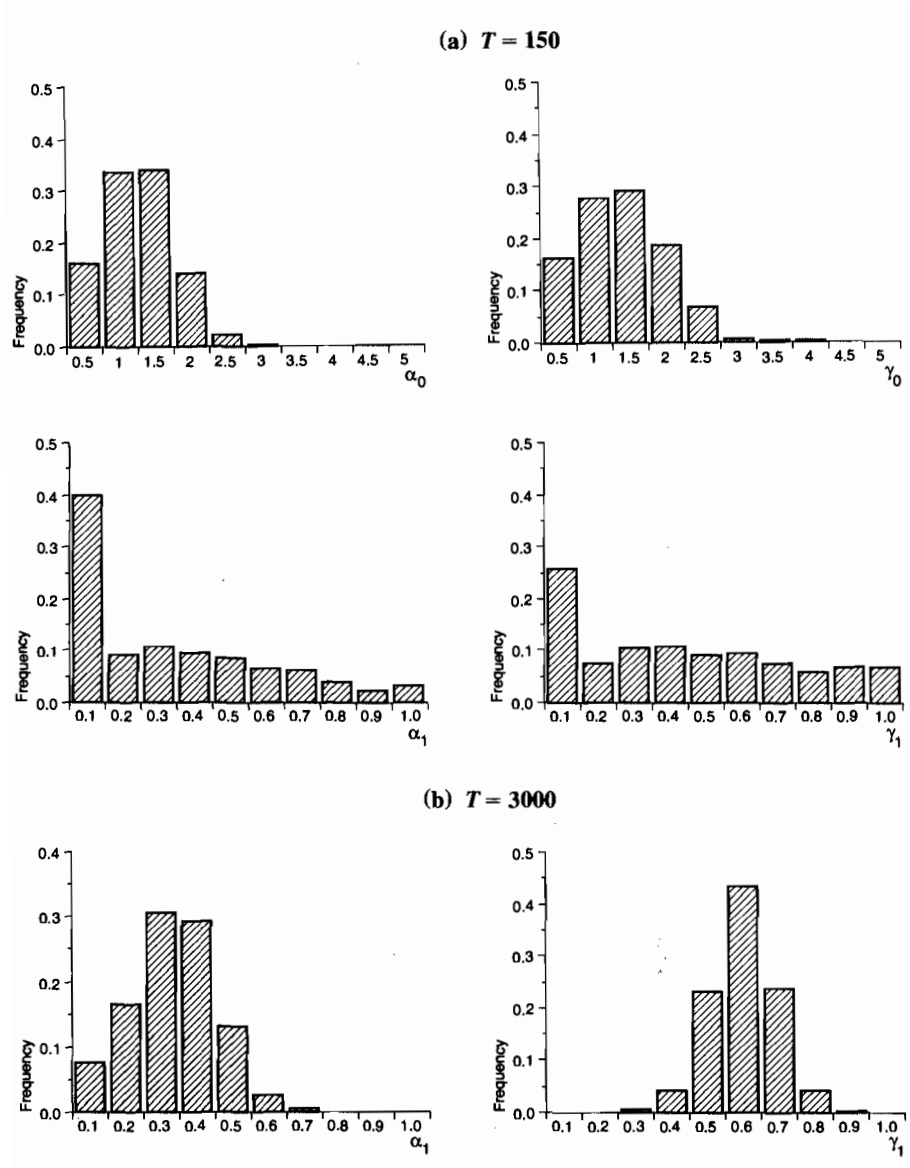


Fig. 1. Empirical distributions for the estimated parameters in a random walk plus noise with ARCH disturbances: $\alpha_0 = \gamma_0 = 1$, $\alpha_1 = 0.3$, and $\gamma_1 = 0.5$.

Table 2

Estimated size and power of tests against ARCH(1) disturbances in random walk plus noise model, with $T = 1000$, based on Monte Carlo experiments with 1000 replications.

(a) No ARCH in random walk ($\gamma_1 = 0$)					
$\gamma_1 = 0$	α_1				
	0.0	0.2	0.4	0.6	0.8
$Q^2(1)$	0.03	0.21	0.59	0.86	0.93
LM*	0.06	0.26	0.70	0.92	0.97
LM(ε)	0.06	0.31	0.74	0.91	0.95
LM(η)	0.06	0.11	0.36	0.69	0.88
(b) No ARCH in measurement noise ($\alpha_1 = 0$)					
$\alpha_1 = 0$	γ_1				
	0.0	0.2	0.4	0.6	0.8
$Q^2(1)$	0.03	0.25	0.67	0.91	0.95
LM*	0.06	0.26	0.71	0.92	0.97
LM(ε)	0.06	0.09	0.29	0.55	0.70
LM(η)	0.06	0.29	0.74	0.93	0.97

Table 3

Estimated power of tests against ARCH(1) distribution in random walk plus noise model based on Monte Carlo experiments with 1000 replications.

	$\alpha_1 = 0.3$	$\gamma_1 = 0.5$	$T = 1000$
	$T = 150$	$T = 500$	
$Q^2(1)$	0.30	0.75	0.95
LM*	0.29	0.76	0.95
LM(ε)	0.18	0.54	0.84
LM(η)	0.25	0.71	0.92

A test can also be set up against GARCH disturbances. The null hypothesis is then $H_0: \alpha_1 = \alpha_2 = \gamma_1 = \gamma_2 = 0$, but a straightforward application of the LM principle fails because of multicollinearity amongst the regressors formed from the derivatives. This is exactly what happens when testing the null hypothesis of white noise against an ARMA(1, 1) alternative; see Poskitt and Tremayne (1980). Obviously the solution is similar and just as the LM test against an ARMA(1, 1) alternative reduces to an LM test against AR(1), so the test against GARCH(1, 1) disturbances reduces to a test against ARCH(1) disturbances.

Tables 2 and 3 show the results of some Monte Carlo experiments based on 1000 replications each. The test statistics are the McLeod-Li test, $Q^2(P)$,

based on the first P squared residuals, the quasi-LM test of the hypothesis that $\alpha_1 = \gamma_1 = 0$, denoted LM* and based on a chi-square with two degrees of freedom, and two one degree of freedom quasi-LM tests against ARCH(1) in the measurement equation disturbance only and the transition equation disturbance only; these are denoted LM(ε) and LM(η) respectively. All the tests are set up at the 5% level of significance.

When there is only ARCH in one disturbance term, the appropriate one degree of freedom quasi-LM test has the highest power. The other one degree of freedom test has a relatively low power, but not so low as to enable the tests to reliably distinguish cases where only one of the disturbances is subject to ARCH. The LM* test always has a relatively high power, though in many cases the power of $Q^2(1)$ is similar.

3.5. Application

A number of studies have found evidence of ARCH effects in exchange rates. The data set we consider consists of 3195 observations on hourly exchange rates for various currencies recorded from 0.00 a.m., January 2, 1987 to 11.00 a.m., July 15, 1987. Baillie and Bollerslev (1991) fitted ARIMA(0, 1, 1) models with GARCH disturbances to the logarithms of these exchange rates. The fact that the ARIMA(0, 1, 1) model is the reduced form of the random walk plus noise model suggests that it might be fruitful to fit the random walk plus noise with GARCH in either or both of the disturbances. Baillie and Bollerslev also included time of day dummies in their GARCH models, but we have not, as yet, attempted to incorporate such variables in our model, though there is no difficulty, in principle, in doing so.

For the Deutschmark/dollar exchange rate, fitting the basic random walk plus noise model gives

$$\begin{aligned}\hat{\sigma}_\varepsilon^2 &= 6.65 \times 10^{-8}, & \hat{\sigma}_\eta^2 &= 333.0 \times 10^{-8}, \\ \log L &= 18,484.3, & Q(10) &= 8.81, & Q^2(10) &= 54.4, \\ BIC &= -36,950.46.\end{aligned}$$

There is no evidence of serial correlation in the residuals, but for the squared residuals the Box-Ljung statistic, $Q^2(P)$ with $P = 10$, is 54.4, indicating a clear rejection of the underlying model specification. The LM* test for ARCH is also significant having a value of 34.34.

The preferred STARARCH model is one in which there is GARCH(1,1) in the transition equation disturbance, but no ARCH in the measurement equation disturbance. For the modified GARCH specification, (23), the

results are

$$\tilde{\alpha}_0 = 4.69 \times 10^{-8},$$

$$\tilde{\gamma}_0 = 100.8 \times 10^{-8}, \quad \tilde{\gamma}_1 = 0.163, \quad \tilde{\gamma}_2 = 0.553,$$

$$\log L = 18,564.5, \quad Q(10) = 8.25, \quad Q^2(10) = 8.10,$$

$$BIC = -37,096.73.$$

The unconditional variance of η_t is 355.5×10^{-8} . Thus the variances implied by the STARCh model are roughly consistent with those estimated for the linear model. The Box–Ljung diagnostics are satisfactory, and there is a clear increase in the log-likelihood. The only unsatisfactory feature of the model is that the residuals still show some evidence of heavy tails, the kurtosis being 11.22. Baillie and Bollerslev also found excess kurtosis in their estimated models.

The results for the yen/dollar exchange rate are very similar, with

$$\tilde{\alpha}_0 = 10.07 \times 10^{-8},$$

$$\tilde{\gamma}_0 = 201.2 \times 10^{-8}, \quad \tilde{\gamma}_1 = 0.240, \quad \tilde{\gamma}_2 = 0.314.$$

However, in the case of the dollar/pound, the variance of the measurement equation disturbance was estimated to be zero.

4. Time-varying parameter ARCH-M model

In the ARCH-M model of Engle et al. (1987), the conditional variance of the disturbance term affects the conditional mean of the observations. In Chou et al. (1992) this model is extended to allow the coefficient of the conditional variance to change over time according to a random walk. Thus consider the model:

$$y_t = h_t \mu_t + \varepsilon_t, \quad (41)$$

$$\varepsilon_t = \varepsilon_t^\dagger h_t^{1/2}, \quad \varepsilon_t^\dagger \sim \text{NID}(0, 1), \quad (42)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \quad (43)$$

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim \text{NID}(0, \sigma_\eta^2), \quad (44)$$

where, more generally, (43) may be a higher-order ARCH or GARCH

process. Chou et al. propose replacing ε_{t-1}^2 in (43) by the squared innovation obtained from the Kalman filter. The model is then conditionally Gaussian, just as in the original ARCH-M model where μ_t is time-invariant. However, this may not be satisfactory if we regard the ARCH effect as stemming from ε_{t-1} . To handle this problem we may proceed as with the local level model of the previous section and evaluate the conditional expectation of ε_{t-1}^2 at time $t-1$ when the conditional variance of ε_t is required for the Kalman filter. The appearance of h_t as the coefficient of μ_t , however, introduces a new complication since h_t is now unobservable at time $t-1$. A solution is to replace h_t by the conditional variance of ε_t^2 at time $t-1$. Thus (41) becomes

$$y_t = z_t \mu_t + \varepsilon_t, \quad (45)$$

with

$$z_t = E_{t-1}(\varepsilon_t^2). \quad (46)$$

Proceeding as in subsection 3.2, with (45) as the measurement equation and (44) as the transition equation, we have

$$\varepsilon_t = y_t - z_t \mu_t = (y_t - z_t \hat{\mu}_t) + z_t (\hat{\mu}_t - \mu_t),$$

and so

$$E_t(\varepsilon_t^2) = \hat{\varepsilon}_t^2 + z_t^2 p_t.$$

Thus

$$\text{var}_{t-1}(\varepsilon_t) = z_t = \alpha_0 + \alpha_1 (\hat{\varepsilon}_{t-1}^2 + z_{t-1}^2 p_{t-1}). \quad (47)$$

In the application presented in Chou et al. (1992), the variance of η_t is relatively small, and for this reason the results obtained from our proposed estimation procedure turned out to be very close to those reported by Chou et al.

5. Latent factor models

With H_t^* and Q_t^* specified to be diagonal, the model in (5) and (6) is a generalisation of the latent factor model of Diebold and Nerlove (1989), which allows for general dynamics in the mean. In their formulation such dynamics are not included directly. Thus α_t does not appear, the state vector contains only ε_t and the state space form reduces to

$$y_t = \Lambda \varepsilon_t + \varepsilon_t^* \quad (48)$$

and

$$\varepsilon_t = \varepsilon_t, \quad (49)$$

where ε_t now represents the common factor and ε_t^* the specific factors. This implies that the conditional covariance matrix of y_t will have the typical factor analytic form $F_t = \Lambda p_{t|t-1} \Lambda' + H_t^*$, where $p_{t|t-1}$ now denotes the conditional variance of ε_t .

Diebold and Nerlove (1989) propose this model as a natural way of capturing the co-movements in the variances of seven dollar exchange rates with a more parsimonious representation than the multivariate ARCH model of Kraft and Engle (1983). But it can also be given a direct economic interpretation in the context of the Arbitrage Pricing Theory of Ross (1976); see King, Sentana, and Wadhvani (1990) for an application along these lines to 16 world stock markets.

When the conditional variance of ε_t is generated by a higher-order ARCH process, the augmentation of the state vector by lagged values of ε_t is unnecessary. The reason is that the degenerate nature of the transition equation means that the smoothed estimates of these disturbances are the same as the filtered estimates. This fact also implies that GARCH-type effects can be handled without much effort and in the manner of subsection 2.4, as eqs. (20) and (24) coincide in this case.

5.1. Application to exchange rates

Diebold and Nerlove (1989, p. 19) do not estimate their model using the method of section 2, but simply replace the expectation of past square values of ε_t by the squares of the estimates obtained from the Kalman filter. That is, they do not apply the correction factor involving p_{t-1} . They remark: ‘while we modify the model associated with our simultaneous estimation procedure to accommodate the fact that F_t is never observed, it would be preferable (but much harder) to preserve the original model and instead appropriately modify the Kalman filter recursions’. (Note that their F_t is our ε_t .)

We re-estimated the Diebold–Nerlove model with and without the correction term. The two sets of parameter estimates are presented in table 4 together with their asymptotic standard errors. As it can be seen, both sets of estimates are remarkably close, especially those corresponding to the standard deviations of the specific terms.¹ In fact, the likelihood functions are almost identical: $-15,576.79$ without the correction versus $-15,575.72$ with

¹The parameter estimates presented here differ slightly from the ones reported in Diebold and Nerlove (1989). The standard deviations of the specific factors are almost identical, but the common factor loadings are higher by a factor of approximately 1.69. At the same time the persistence parameter is estimated to be slightly lower than their 0.975, whereas the log-likelihood function (excluding the constant) is $-11,614.33$ vs. $-11,604.29$ in their paper. These differences are due to the different treatment of the initial conditions. However, for our purposes what matters is the difference between the estimates with and without the correction.

Table 4
Parameter estimates for the latent factor model.

	CD	FF	DM	LIR	YEN	SF	BP
(a) Without correction term							
$\hat{\lambda}$	2.62 (0.77)	20.38 (5.81)	22.21 (6.30)	15.60 (4.46)	13.09 (3.74)	24.16 (6.88)	-15.83 (4.52)
$\hat{h}^{*1/2}$	5.05 (0.14)	6.78 (0.25)	4.57 (0.25)	8.84 (0.26)	10.33 (0.30)	8.42 (0.29)	10.24 (0.29)
Persistence: 0.9535				Log-likelihood: -15576.79			
(b) With the correction term							
$\hat{\lambda}$	2.41 (0.78)	18.80 (5.87)	20.48 (6.38)	14.40 (4.53)	12.09 (3.82)	22.30 (6.93)	-14.61 (4.58)
$\hat{h}^{*1/2}$	5.05 (0.14)	6.78 (0.24)	4.61 (0.25)	8.83 (0.26)	10.33 (0.31)	8.42 (0.29)	10.23 (0.31)
Persistence: 0.9703				Log-likelihood: -15575.72			

correction. The main difference is that the factor loadings obtained without the correction are somewhat higher than those which use it, whereas the corresponding persistence parameter changes from 0.9535 to 0.9703. This similarity is not surprising as the latent factor has no dynamics in the mean, only in the variance. Hence, the correction term has a one-off effect on the variance of the current observation and does not carry over to further observations.

When the estimates of the time-varying standard deviation of the common factor obtained by both methods are plotted, it is found, as expected, that the conditional variances obtained using the correction term are larger than those obtained without it. However, the correlation between the two sets of figures is very high. Therefore it seems that because of the special structure of this particular model, ignoring the correction does not substantially alter the results, especially if one is mainly interested in decomposing the observed series into common and idiosyncratic components.

5.2. Fat-tailed disturbances

The special structure of the transition equation allows us to generalize the model to allow for disturbances with fat tails. This is important because Gaussian ARCH models do not seem to capture completely the degree of leptokurtosis often observed in practice, particularly in financial data.

As in Bollerslev (1987) and Baillie and Bollerslev (1989) our approach is based on the t -distribution, which includes the normal as a limiting case, but has generally fatter tails. Let

$$\varepsilon_t = h_t^{1/2} \varepsilon_t^\dagger = h_t^{1/2} [\xi_t / (\nu - 2)]^{-1/2} \varepsilon_t^{\dagger\dagger}, \quad (50)$$

$$\varepsilon_t^* = H_t^{*1/2} [\xi_t / (\nu - 2)]^{-1/2} \varepsilon_t^{**}, \quad (51)$$

where $\varepsilon_t^{\dagger\dagger}$ is a univariate standard normal variate, ε_t^{**} a multivariate one, ξ_t a χ^2 variable with ν degrees of freedom, and all three variates are mutually independent. Thus, conditional on the information at time $t - 1$ both the common and specific factors are proportional to a standardized t -distribution with ν degrees of freedom. Note, however, that although ε_t and ε_t^* are conditionally orthogonal, they are not independent because of the common denominator.

The above assumptions imply that each element of y_t has a conditional distribution which is proportional to a Student's t . Therefore, by allowing $\nu < \infty$ we can have not only a leptokurtic unconditional distribution due to ARCH effects, but also a fat-tail conditional distribution.

Using the properties of the multivariate t -distribution, as given, for example, in Zellner (1971, pp. 383–389), the updating equations, which give the mean and variance of ε_t conditional on y_t are

$$\hat{\varepsilon}_t = P_{t|t-1} \Lambda' F_t^{-1} y_t \quad (52)$$

and

$$p_t = (\nu - 2)(\nu + N - 2)^{-1} \left[1 + (\nu - 2)^{-1} y_t' F_t^{-1} y_t \right] \\ \times \left\{ p_{t|t-1} - p_{t|t-1} \Lambda' F_t^{-1} \Lambda p_{t|t-1} \right\}. \quad (53)$$

Notice that no real extra complications arise in this case as eq. (52) is the standard state updating equation under normality [see Harvey (1989, ch. 3)], while (53) is basically the standard covariance matrix updating equation, the term outside the curly brackets being simply a scalar correction factor.

The log-likelihood function now becomes

$$\log L = -(TN/2) \ln[\pi(\nu - 2)] + T \{ \ln \Gamma[(\nu + N)/2] - \ln \Gamma(\nu/2) \} \\ - \frac{1}{2} \sum \{ \ln |F_t| + (\nu + N) \ln [1 + (\nu - 2)^{-1} y_t' F_t^{-1} y_t] \}. \quad (54)$$

Table 5
Parameter estimates for the latent factor model with t -distributed factors.

	CD	FF	DM	LIR	YEN	SF	BP
$\hat{\lambda}$	3.26 (0.95)	22.72 (6.02)	24.62 (6.51)	17.56 (4.67)	14.45 (3.90)	26.54 (7.02)	-16.76 (4.50)
$\hat{h}^{*1/2}$	5.70 (0.25)	5.36 (0.30)	4.40 (0.29)	7.09 (0.35)	11.13 (0.49)	8.43 (0.41)	11.15 (0.50)
Persistence: 0.9305		Log-likelihood: -15070.54		Degrees of freedom: 3.89567			

We re-estimated the Diebold and Nerlove (1989) data with this new distributional assumption, and the results are presented in table 5. The most distinctive feature is that conditional normality is strongly rejected (the LR ratio is 1010.36). In fact, ν is estimated to be just below 4, which implies that not even the conditional fourth moments exist. Nevertheless the other parameter values do not change substantially, and the new estimates of the common factor and its variance are very highly correlated with those obtained under normality. The persistence parameter, however, is now 0.930 compared to 0.97 under normality.

The fact that the estimate of ν is so low is mainly due to the high excess kurtosis of the ε_t^{**} 's. It is possible to remove some of this excess kurtosis by letting the ε_t^{**} 's follow ARCH processes. However, the main source of the problem appears to be a small number of observations which are around ten standard deviations from the mean. If these were treated as outliers and removed from the data set, the estimate of ν would be much higher.

6. Conclusion

The proposed estimation procedure for unobserved components models with ARCH disturbances can be carried out using the Kalman filter. It involves an approximation, but the Monte Carlo results for a simple random walk plus noise model indicate that it works well for reasonably large samples. Monte Carlo results also show that the naive procedure in which the required correction factors are not employed can be quite inefficient. For the two applications examined, the potential inefficiency of the naive procedure is not particularly apparent insofar as the results obtained are similar to those with the proposed procedure. However, other applications could easily show a significant difference, and the fact that the proposed procedure proved to be viable in the cases examined makes us feel confident in recommending that it be adopted.

Appendix

Properties of conditionally Gaussian ARCH disturbances in unobserved component time series models

Consider the following ARCH process, where the lagged disturbances have been replaced by their conditional expectations:

$$\varepsilon_t = \varepsilon_t^\dagger \left(\alpha_0 + \alpha_1 \left(\mathbb{E}_{t-1} \varepsilon_{t-1}^2 \right) \right)^{1/2}, \quad (\text{A.1})$$

where ε_t^\dagger is a standard normal variable and the $t-1$ under the expectation operator indicates that it is conditional on observations up to and including time $t-1$. Notice that, because we are dealing with unobserved component models, ε_{t-1} may not be included in the information set. If ε_{t-1} were included in the information set, then ε_t would be a standard ARCH(1) process, as defined by Engle (1982). For simplicity, we are considering the ARCH(1) case, but the results could be easily extended to higher-order processes.

Taking expectations in (A.1), we get

$$\mathbb{E}(\varepsilon_t) = \mathbb{E}(\varepsilon_t^\dagger) \mathbb{E} \left(\alpha_0 + \alpha_1 \mathbb{E}_{t-1} \varepsilon_{t-1}^2 \right)^{1/2} = 0.$$

The unconditional variance of ε_t is given by

$$\mathbb{E}(\varepsilon_t^2) = \mathbb{E}(\varepsilon_t^{\dagger 2}) \mathbb{E} \left(\alpha_0 + \alpha_1 \mathbb{E}_{t-1} \varepsilon_{t-1}^2 \right) = \alpha_0 + \alpha_1 \mathbb{E} \varepsilon_{t-1}^2,$$

and, therefore, if $\alpha_1 < 1$,

$$\mathbb{E}(\varepsilon_t^2) = \alpha_0 / (1 - \alpha_1).$$

Because of the normality of ε_t^\dagger , all the odd unconditional moments of ε_t are zero, that is,

$$\mathbb{E}(\varepsilon_t^r) = \mathbb{E}(\varepsilon_t^{\dagger r}) \mathbb{E} \left(\alpha_0 + \alpha_1 \mathbb{E}_{t-1} \varepsilon_{t-1}^2 \right)^{r/2} = 0.$$

The fourth moment of ε_t is given by

$$\begin{aligned} \mathbb{E}(\varepsilon_t^4) &= \mathbb{E}(\varepsilon_t^{\dagger 4}) \mathbb{E} \left(\alpha_0 + \alpha_1 \mathbb{E}_{t-1} \varepsilon_{t-1}^2 \right)^2 \\ &= 3\alpha_0^2(1 - \alpha_1^2) / (1 - \alpha_1)^2 + 3\alpha_1^2 \mathbb{E} \left(\mathbb{E}_{t-1} \varepsilon_{t-1}^2 \right)^2. \end{aligned}$$

Using Jensen's inequality,

$$E(\varepsilon_t^4) \leq 3\alpha_0^2(1 - \alpha_1^2)/(1 - \alpha_1)^2 + 3\alpha_1^2 E\varepsilon_{t-1}^4.$$

Then, if $3\alpha_1^2 < 1$,

$$E(\varepsilon_t^2) \leq \frac{3\alpha_0^2}{(1 - 3\alpha_1^2)} \frac{1 - \alpha_1^2}{(1 - \alpha_1)^2}. \quad (\text{A.2})$$

Therefore, the kurtosis of ε_t in (A.1) is bounded from above by the kurtosis of the standard ARCH(1) process; see Engle (1982).

It is easy to see that ε_t is a serially uncorrelated process, since

$$\begin{aligned} E(\varepsilon_t \varepsilon_{t-\tau}) &= E(\varepsilon_t^\dagger) E\left[\left(\alpha_0 + \alpha_1 E_{t-1} \varepsilon_{t-1}^2\right) \varepsilon_{t-\tau}^\dagger \left(\alpha_0 + \alpha_1 E_{t-\tau-1} \varepsilon_{t-\tau-1}^2\right)\right] \\ &= 0. \end{aligned}$$

On the other hand, ε_t^2 is autocorrelated, but its autocorrelation function is bounded from above by the autocorrelation function of the standard ARCH(1) process. To see this, define the variables

$$z_t = \varepsilon_t^2 - E\varepsilon_t^2 \quad \text{and} \quad v_t = \varepsilon_t^2 - E_{t-1} \varepsilon_t^2.$$

Then,

$$z_t = \alpha_1 E_{t-1} z_{t-1} + v_t.$$

The order τ autocorrelation coefficient of ε_t^2 is given by

$$\begin{aligned} \rho_\tau &= E(z_t z_{t-\tau}) / E(z_t^2) = E\left[\left(\alpha_1 E_{t-1} z_{t-1} + v_t\right) z_{t-\tau}\right] / E(z_t^2) \\ &= \left[\alpha_1 E\left(z_{t-\tau} E_{t-1} z_{t-1}\right) + E v_t z_{t-\tau}\right] / E(z_t^2), \quad \tau = 1, 2, \dots \end{aligned}$$

The second term in the numerator of the expression above is zero,

$$\begin{aligned} E v_t z_{t-\tau} &= E\left[\left(\varepsilon_t^2 - E_{t-1} \varepsilon_t^2\right) \left(\varepsilon_{t-\tau}^2 - E\varepsilon_{t-\tau}^2\right)\right] \\ &= E\left[\varepsilon_t^2 \varepsilon_{t-\tau}^2 - \varepsilon_{t-\tau}^2 E_{t-1} \varepsilon_t^2\right] \\ &= E\left[\varepsilon_t^{\dagger 2} \left(\alpha_0 + \alpha_1 E_{t-1} \varepsilon_{t-1}^2\right) \varepsilon_{t-\tau}^2 - \varepsilon_{t-\tau}^2 E_{t-1} \varepsilon_t^2\right] \\ &= E\left[\varepsilon_{t-\tau}^2 \left(\alpha_0 + \alpha_1 E_{t-1} \varepsilon_{t-1}^2 - E_{t-1} \varepsilon_t^2\right)\right] = 0. \end{aligned}$$

Therefore, ρ_τ is given by

$$\rho_\tau = \alpha_1 \mathbb{E} \left[z_{t-\tau} \mathbb{E}_{t-1} z_{t-1} \right] / \mathbb{E}(z_t^2).$$

Using the fact that

$$\begin{aligned} \mathbb{E}_t z_t &= \mathbb{E}_t \left(\alpha_1 \mathbb{E}_{t-1} z_{t-1} + v_t \right) \\ &= \alpha_1 \mathbb{E}_{t-1} z_{t-1} + \mathbb{E}_t v_t = \alpha_1 \mathbb{E}_{t-1} z_{t-1} \\ &= \alpha_1^r \mathbb{E}_{t-r} z_{t-r}, \quad r = 1, 2, \dots \end{aligned}$$

we can express ρ_τ as

$$\begin{aligned} \rho_\tau &= \alpha_1 \mathbb{E} \left[z_{t-\tau} \alpha_1^{\tau-1} \mathbb{E}_{t-\tau} z_{t-\tau} \right] / \mathbb{E}(z_t^2) \\ &= \alpha_1^\tau \mathbb{E} \left[z_{t-\tau} \mathbb{E}_{t-\tau} z_{t-\tau} \right] / \mathbb{E}(z_t^2) \\ &= \alpha_1^\tau \mathbb{E} \left[\left(\mathbb{E}_{t-\tau} z_{t-\tau} \right)^2 \right] / \mathbb{E}(z_t^2). \end{aligned}$$

Using again Jensen's inequality,

$$\rho_\tau \leq \alpha_1^\tau \mathbb{E}(z_{t-\tau}^2) / \mathbb{E}(z_t^2),$$

and so

$$\rho_\tau \leq \alpha_1^\tau. \tag{A.3}$$

The inequality in (A.3) shows that the autocorrelation function of ε_t^2 is bounded from above by the autocorrelation function of the ARCH(1) process; see Milhoj (1985).

Another possible way of introducing ARCH-type behaviour in the disturbances of unobserved component time series models is to replace the lagged disturbances by the square of their conditional expectation as follows:

$$\varepsilon_t^* = \varepsilon_t^\dagger \left(\alpha_0 + \alpha_1 \left(\mathbb{E}_{t-1} \varepsilon_{t-1} \right)^2 \right)^{1/2}.$$

The unconditional variance of ε_t^* is still easily shown to be zero, but the unconditional variance is smaller than in the previous case:

$$\begin{aligned} E\varepsilon_t^{*2} &= E\varepsilon_t^{\dagger 2} E\left[\alpha_0 + \alpha_1 \left(\mathbf{E}_{t-1} \varepsilon_{t-1}\right)^2\right] \\ &= \alpha_0 + \alpha_1 E\left[\left(\mathbf{E}_{t-1} \varepsilon_{t-1}\right)^2\right] \leq \alpha_0 + \alpha_1 E(\varepsilon_{t-1}^2), \end{aligned}$$

and, therefore,

$$E\varepsilon_t^{*2} \leq \alpha_0 / (1 - \alpha_1).$$

As in the previous case, ε_t^* is serially uncorrelated,

$$E\varepsilon_t^* \varepsilon_{t-\tau}^* = E\varepsilon_t^{\dagger} E\left[\left(\alpha_0 + \alpha_1 \left(\mathbf{E}_{t-1} \varepsilon_{t-1}\right)^2\right)^{1/2} \varepsilon_{t-\tau}^*\right] = 0.$$

The fourth moment of ε_t^* is bounded from above by the fourth moment of ε_t in (A.1),

$$\begin{aligned} E\varepsilon_t^{*4} &= E\left(\varepsilon_t^{\dagger 4} \left(\alpha_0 + \alpha_1 \left(\mathbf{E}_{t-1} \varepsilon_{t-1}\right)^2\right)^2\right) \\ &= 3E\left[\alpha_0 + \alpha_1 \left(\mathbf{E}_{t-1} \varepsilon_{t-1}\right)^2\right]^2 \leq 3E\left[\alpha_0 + \alpha_1 \mathbf{E}_{t-1} \varepsilon_{t-1}^2\right]^2 = E\varepsilon_t^4. \end{aligned}$$

Similarly, the autocorrelation function of ε_t^{*2} is bounded from above by the autocorrelation function of ε_t^2 . If ρ_τ^* is the order τ autocorrelation coefficient of ε_t^* , then

$$\rho_\tau^* = E(z_t z_{t-\tau}) / E(z_t^2),$$

where

$$z_t = \varepsilon_t^2 - E\varepsilon_t^2 = \alpha_1 \left[\left(\mathbf{E}_{t-1} \varepsilon_{t-1}\right)^2 - E\varepsilon_{t-1}^2 \right] + v_t$$

and

$$v_t = \varepsilon_t^2 - \mathbf{E}_{t-1} \varepsilon_{t-1}^2.$$

As before, $E v_t z_{t-\tau} = 0$, and therefore,

$$\begin{aligned}\rho_\tau^* &= \alpha_1 E \left[z_{t-\tau} \left(\left(E_{t-1} \varepsilon_{t-1} \right)^2 - E \varepsilon_{t-1}^2 \right) \right] / E(z_t^2) \\ &\leq \alpha_1 E \left[z_{t-\tau} \left(E_{t-1} \varepsilon_{t-1}^2 - E \varepsilon_{t-1}^2 \right) \right] / E(z_t^2) \\ &= \alpha_1 E \left(z_{t-\tau} E_{t-1} z_{t-1} \right) / E(z_t^2) = \rho_\tau.\end{aligned}$$

References

- Baillie, R.T. and T. Bollerslev, 1989, The message in daily exchange rates: A conditional-variance tale, *Journal of Business and Economic Statistics* 7, 297–305.
- Baillie, R.T. and T. Bollerslev, 1991, Inter-day and inter-market volatility in foreign exchange markets, *Review of Economic Studies* 58, 565–585.
- Black, F., 1986, Noise, *Journal of Finance* 3, 529–543.
- Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 51, 307–327.
- Bollerslev, T., 1987, A conditional heteroskedastic time series model for speculative prices and rates of return, *Review of Economics and Statistics* 69, 542–547.
- Bollerslev, T., 1988, On the correlation structure for the generalized autoregressive conditional heteroskedastic process, *Journal of Time Series Analysis* 9, 121–131.
- Chou, R., R. Engle, and A. Kane, 1992, Measuring risk aversion from excess returns on a stock index, *Journal of Econometrics*, this issue.
- Diebold, F.X. and M. Nerlove, 1989, The dynamics of exchange rate volatility: A multivariate latent factor ARCH model, *Journal of Applied Econometrics* 4, 1–21.
- Engle, R.F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* 50, 987–1007.
- Engle, R.F. and T. Bollerslev, 1986, Modelling the persistence of conditional variances, *Econometric Reviews* 5, 1–50.
- Engle, R.F. and M. Watson, 1981, A one-factor multivariate time series model of Metropolitan wage rates, *Journal of the American Statistical Association* 76, 774–781.
- Engle, R.F., D.M. Lilien, and R.P. Robins, 1987, Estimating time varying risk premia in the term structure: The ARCH-M model, *Econometrica* 55, 391–407.
- Harvey, A.C., 1989, *Forecasting, structural time series models and the Kalman filter* (Cambridge University Press, Cambridge).
- King, M.A., E. Sentana, and S.B. Wadhvani, 1990, A heteroskedastic factor model of asset returns and risk premia with time-varying volatility: An application to sixteen world stock markets, *Financial Markets Group discussion paper no. 90* (London School of Economics, London).
- Kraft, D.F. and R.F. Engle, 1983, Autoregressive conditional heteroscedasticity in multiple time series models, Mimeo. (Department of Economics, University of California, San Diego, CA).
- Maravall, A., 1983, An application of nonlinear time series forecasting, *Journal of Business and Economic Statistics* 1, 66–74.
- McLeod, A.J. and W.K. Li, 1983, Diagnostic checking ARMA time series models using squared-residual autocorrelations, *Journal of Time Series Analysis* 4, 269–273.
- Milhoj, A., 1985, The moment structure of ARCH processes, *Scandinavian Journal of Statistics* 12, 281–292.
- Nelson, D.B., 1990, Stationarity and persistence in the GARCH(1,1) model, *Econometric Theory* 6, 318–334.

- Poskitt, D.S. and A.R. Tremayne, 1980, Testing the specification of a fitted autoregressive-moving average process, *Biometrika* 67, 359–363.
- Poterba, J.M. and L.H. Summers, 1988, Mean reversion in stock process: Evidence and implications, *Journal of Financial Economics* 2, 27–59.
- Ross, S.A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.
- Sentana, E., 1991, Identification of multivariate conditionally heteroskedastic latent factor models, Mimeo. (London School of Economics, London).
- Weiss, A., 1984, ARMA models with ARCH errors, *Journal of Time Series Analysis* 5, 129–143.
- Zellner, A., 1971, *An introduction to Bayesian inference in econometrics* (Wiley, New York, NY).