CONTINUOUS TIME GARCH-BASED MODELING AND FILTERING: EVIDENCE FROM A SHORT-TERM RATE PROCESS

Fabio Fornari *
Bank of Italy, Research Department

Antonio Mele
Queen Mary and Westfield College, University of London

Abstract

Aim of this article is to judge the empirical performance of Arch as diffusion approximations to models of the short-term rate with stochastic volatility and as filters of the unobserved volatility. We show that the estimation of the continuous time scheme to which a discrete time Arch model converges can be safely based on simple moment conditions linking the discrete time to the continuous time coefficients. A natural substitute of a global specification test for just-identified problems based on indirect inference shows in fact that this approximation to diffusions gives rise to a negligible disaggregation bias. Unlike previous literature in which standard Arch models approximated only specific diffusions, our estimation strategy relies on a new Arch that approximates any Cev-diffusion model for the conditional volatility. A Monte-Carlo study reveals that the filtering performances of this model is remarkably good, even in the presence of an important kind of misspecification.

JEL classification: C15, E43, G12.

Keywords: stochastic volatility, Arch filtering, indirect inference.

* Fabio Fornari, Banca d'Italia - Studi, Via Nazionale 91, I-00184 Roma, Italy. Phone: +39 06 47923189. Fax: +39 06 47924118. Email: *fornari.fabio@insedia.interbusiness.it*

1. Introduction(*)

Great progresses have been recently made in the estimation of stochastic differential equations. Aït-Sahalia (2000) developed a maximum likelihood estimator for scalar diffusions whereby the unknown transitional density of the model can be approximated, in closed-form, with great accuracy. Brandt and Santa-Clara (2001) and Durham and Gallant (2001), both building on Pedersen (1995), proposed to resort to simulating high-frequency paths of the state variables of the continuous time model and by means of these, recover the unknown transitional densities, hence the approximate likelihood; the same methodology is applied in a nonparametric fashion by Nicolau (1999). The nonparametric framework is employed also by Altissimo et al. (2001), who develop a simulated nonparametric estimator based on matching the true density of the data and a density simulated conditionally on a continuous time model and given values of its parameters. The unanimous conclusion of these papers is that, as the short interest rate dynamics is concerned, traditional univariate diffusions perform poorly relative to bivariate continuous time models where the interest rate dynamics is coupled with its conditional volatility dynamics. This finding, not unexpected, represents the continuous time counterpart of the universal finding of Arch-type effects[1] in time series of financial price changes; it finds theoretical justification in the initial contribution of Nelson (1990), where some basic ARCH models are shown to be reasonable approximations to the diffusion processes frequently used in theoretical finance models.[2]

In this paper we wish to study the implications of using Arch-type models as i) continuous time approximations to diffusions as well as ii) filters in continuous time models

---

0 (*) This paper was written while the first author was at the University of Cambridge and the second at the Princeton University. We thank Yacine Aït-Sahalia, Pippo Altissimo, Ron Gallant, Steven Satchell, José Scheinkman and seminar participants at Princeton University and Cambridge University, the 1998 Econometric Society European Meeting at Berlin and the 1999 Society for Computational Economics Conference at Boston College for helpful comments. Responsibility for any views or errors in the paper rests with the authors.

1 See, e.g., Bollerslev et al. (1994), for a survey of the Arch literature. The unanimous finding of conditional heteroskedasticity in financial data has led researchers (e.g., Hull and White, 1987; Wiggins, 1987; Longstaff and Schwartz, 1992; Heston, 1993) to extend early asset pricing theories (e.g., Black and Scholes, 1973; Merton, 1973; Vasicek, 1977) to the case in which volatility evolves in a stochastic manner.

2 The major contribution of Nelson to this strand of research can be found in part II of the book edited by Rossi (1996).

with unobservable state variables.[3] More precisely, a continuous time models of the short-term rate is used as a benchmark to test whether Arch-type models are indeed useful devices to approximate and/or support the estimation of its parameters and to recover the dynamics of the unobservable volatility, a research topic started by Nelson (1990) and subsequently, for no apparent reason, abandoned. Indeed, according to Campbell et al. (1997, p. 381), the empirical properties of Arch as approximations to continuous time stochastic volatility processes "have yet to be explored but will no doubt be the subject of future research".

Given the empirical success of constant elasticity of variance models (Durham, 2001), our concern is first to broaden the Arch class by developing a model that approximates any diffusion model where volatility follows a constant elasticity of variance process (henceforth Cev-Arch) and then to check the functioning of the model to the aims stated above. The main steps of the estimation phase will then be to show i) that the Cev-Arch specification that we wish to employ as a reference model converges to a continuous time model for the short term interest rate as the sampling interval shrinks to zero, ii) that one can easily use the likelihood function of the Cev-Arch in lieu of the true likelihood function (which cannot be computed analytically), iii) that, for the proposed continuous time model, the discrete time Arch approximation is indeed successful in an application based on fitting the dynamics of the short term rate and, last, iv) that the model makes no significant error in recovering the 'true' volatility of the short rate.

Let us make clear from the beginning that the main difficulty faced in the estimation of our continuous time reference model is in the second and in the third of the above mentioned four steps. Though we will provide closed-form moment conditions linking the discrete time to the continuous time parameters and which guarantee the weak convergence of the Cev-Arch model toward the continuous time reference model, Arch schemes are typically not closed under temporal aggregation (Drost and Nijman, 1993 and Drost and Werker, 1996) and, for this reason, we need to test and correct potential 'disaggregation' biases; in this additional step, Arch models will be viewed as *auxiliary* devices in simulation-based (indirect

[3] As widely recognized, Arch models are very appealing for statistical reasons, even though there exist alternative econometric formulations that are surveyed, for instance, in Shephard (1996).

inference) schemes.[4] To anticipate, we find that the correction made by indirect inference is not statistically significant, a result obtained via a global specification test for just-identified models that was originally suggested by Gouriéroux et al. (1993).[5]

As stated in item iv) above, beyond caring about the ability of the Cev-Arch at correctly estimating the parameters of its continuous time limit, we also want to make sure that the model provides a good volatility filtering. This is an important step of our analysis since, for example, if one takes the two-factor model proposed below (see eq. (1)) as data generating process for the short rate, then bond prices will depend both on the interest rate level and on the level of the volatility, and this latter variable becomes an essential ingredient in the practical implementation of a term structure model with stochastic volatility. We already know, from a theoretical standpoint, that appropriate sequences of ARCH models are able to consistently estimate the volatility of a continuous time stochastic process as the sample frequency gets larger and larger, even in the presence of serious misspecifications (Nelson, 1992; Nelson and Foster, 1994).[6] As put by Bollerslev and Rossi (1996), "one could regard the ARCH model as merely a device which can be used to perform filtering or smoothing estimation of unobserved volatilities" (p. xiv). We provide evidence that the desiderable filtering performances of *standard* Arch models are also shared by the Cev-Arch, as one might have expected by a suitable interpretation of the theory (see Nelson and Foster, 1994, theorem 4.1).

---

[4] See Gouriéroux and Monfort (1996) for a full account of simulation-based inference methods. See also Fornari and Mele (2001) for further work related to the diffusion approximation property of Arch in a derivatives context.

[5] Our empirical findings are obtained with the same data set as Andersen and Lund (1997a), who rely on the efficient method of moments (EMM) estimation proposed by Gallant and Tauchen (1996). The advantage of the EMM estimator is that it achieves the same efficiency as the true (intractable) maximum likelihood estimator when the auxiliary model generates a density that 'smoothly embeds' the true likelihood function of the discretely sampled diffusion. It should be clear that our estimation strategy has the aim of ascertaining whether our auxiliary model is a reasonable approximation to the continuous time model. In technical terms, we are going to focus on the empirically difficult just-identified case, a strategy originally suggested in Gouriéroux et al. (1993) (p. S108): "[Indirect inference] methods seem particularly promising when the criterion is based on approximations of the likelihood function, time discretization, range discretizations, linearizations, etc. In this case the method is simpler [...] and appears as an automatic correction for the asymptotic bias implied by the approximation". In our context, indeed, "the asymptotic bias implied by the approximation" is given by a disaggregation bias. While not closed under temporal aggregation, Arch models still have a natural interpretation in terms of the continuous time models that they approximate, being very close (in terms of probability distribution) to the approximated continuous time models when the sampling frequency is high. Furthermore, the auxiliary criteria that we construct are based on approximations that create a natural one-to-one interpretation of the sequence of the parameters of the auxiliary discrete time model in terms of the parameters of the continuous time model (see section 3).

[6] See Bollerslev and Rossi, 1996, (p. xiii-xvii) for a brief account on the filtering performances of ARCH models as applied to continuous time stochastic volatility models.

The paper is organized as follows. Next section presents the basic structure of our continuous time model; it also provides intuition and preliminary results on the estimation and filtering methods to be implemented with the help of Arch models that do not constrain the elasticity of variance to one (the Cev-Arch). The econometric strategy is fully detailed in section 3 while empirical results are in section 4; section 5 concludes and technical considerations and proofs are gathered in the Appendices.

2. The reference continuous time model and the Cev-Arch process

The continuous time model that we wish to use in this paper has an instantaneous volatility of the short-term rate which is a constant elasticity of variance process:

$$\left\{ \begin{array}{lcl} dr(\tau) & = & (\iota - \theta r(\tau))d\tau + \sigma(\tau)\sqrt{r(\tau)}dW^{(1)}(\tau) \\ \\ d\sigma(\tau)^{\delta} & = & (\omega - \varphi\sigma(\tau)^{\delta})d\tau + \psi\sigma(\tau)^{\delta\cdot\eta}d\left(\rho W^{(1)}(\tau) - \sqrt{1-\rho^2}W^{(2)}(\tau)\right), \end{array} \right. \tag{1}$$

where $a = (\iota, \theta, \delta, \omega, \varphi, \psi, \eta, \rho)$ is a vector of parameters, $W^{(i)}$, $i = 1, 2$, are standard Brownian motions, and $\delta \geq 1$. The $\sqrt{r(\cdot)}$-term included in the diffusion term of the short rate equation restricts this variable to positive values only and captures an empirical regularity known as 'level effect', i.e., *coeteris paribus*, the short-term rate volatility rises with the level of the short-term rate. Allowing for more general diffusion terms such as for instance $\sigma(\cdot)\,|r(\cdot)|^d$ $(d \geq 0.5)$ is possible, though it would not change dramatically our empirical results.

As stated, objective of the paper is to use Arch-type models that allow i) the estimation of the above continuous time parameters and ii) the extraction of the unobserved short-term rate volatility process $\sigma(\cdot)$.

To this aim, consider the following Euler-Maruyama discrete time approximation of (1):

$$\left\{ \begin{array}{lcl} {}_hr_{h(k+1)} - {}_hr_{hk} & = & (\iota - \theta \cdot {}_hr_{hk})h + {}_h\sigma_{hk}\sqrt{{}_hr_{hk}} \cdot {}_hu_{h(k+1)} \\ \\ {}_h\sigma^{\delta}_{h(k+1)} - {}_h\sigma^{\delta}_{hk} & = & (\omega - \varphi \cdot {}_h\sigma^{\delta}_{hk})h + \psi \cdot {}_h\sigma^{\delta\cdot\eta}_{hk}\sqrt{h}\,{}_h\widetilde{\xi}_{h(k+1)} \end{array} \right. \tag{2}$$

where h denotes the discretization step,

$$\left(\begin{array}{c} {}_hu_{hk} \\ {}_h\widetilde{\xi}_{hk} \end{array} \right) \sim NID\left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right); \left(\begin{array}{cc} h & \sqrt{h}\rho \\ \sqrt{h}\rho & 1 \end{array} \right) \right),$$

and $({}_hr_{hk}, {}_h\sigma_{hk})_{k=1}^{\infty}$ are the discretized short-term rate and volatility processes.

It is well known that when $h \downarrow 0$ (2) converges weakly (or in distribution) to (1).[7] Hence, the higher the sampling frequency, the higher should be the accuracy of, say, maximum likelihood (ML) estimates of a obtained with (2). Unfortunately (2) represents a discrete time stochastic variance model for which ML methods are rather hard to implement; in addition to this, and with reference to the second of our aims, there are no obvious techniques to filter the actual volatility path out of (2).

A natural alternative to the estimation problem is represented by Arch models when thought of as diffusion approximations, though not every diffusion can be approximated by an Arch scheme. To get an intuition of how the approximating property works, consider the standard Garch(1,1) of Bollerslev (1986):

$$\sigma_{n+1}^2 = w + \beta\sigma_n^2 + \alpha\epsilon_n^2,\ \epsilon_n \equiv (u \cdot \sigma)_n,\ n = 0, 1, \cdots$$

where w, β and α are parameters, $(w, \beta, \alpha) \in R_+^3$, ϵ is the residual of an observation equation, and the index n is an abstract notation for sample points at discrete time intervals (a more precise notation will be introduced in the next section). Rewrite the preceding equation as:

$$\sigma_{n+1}^2 - \sigma_n^2 = w - \left(1 - \alpha E(u^2) - \beta\right)\sigma_n^2 + \alpha\sigma_n^2\left(u_n^2 - E(u^2)\right), \tag{3}$$

and suppose that $u \sim N(0, 1)$. Chop time so as to make $n : hk \leq n \leq h(k+1)$, $k = 1, 2, \cdots$ and let the parameters w, β, α vary with h by introducing sequences w_h, β_h, α_h, and then let $h \downarrow 0$; the resulting volatility process converges in distribution to:[8]

$$d\sigma(\tau)^2 = \left(\omega - \varphi\sigma(\tau)^2\right)d\tau + \psi\sigma(\tau)^2 dW^{(2)}(\tau), \tag{4}$$

---

[7] If (1) has a unique strong solution denoted as $\{r(\tau), \sigma(\tau)^\delta\}_{\tau \geq 0}$, *weak convergence* of $\{{}_h r_{hk}, {}_h\sigma_{hk}^\delta\}_{k=1,2,\ldots}$ in (2) to $\{r(\tau), \sigma(\tau)^\delta\}_{\tau \geq 0}$ means that the finite dimensional distributions of $\{{}_h r_{hk}, {}_h\sigma_{hk}^\delta\}_{k=1,2,\ldots}$ converge to those of $\{r(\tau), \sigma(\tau)^\delta\}_{\tau \geq 0}$ as $h \downarrow 0$. See Stroock and Varadhan (1979). It turns out that the conditions demanded by Stroock and Varadhan (1979) are difficult to verify when studying the convergence of ARCH-type models. One then may wish to make reference to the conditions suggested by Nelson (1990).

[8] To obtain an intuition of this result, notice that the sequence $(\xi_n)_{n=1}^\infty \equiv \left(u_n^2 - E(u^2)\right)_{n=1}^\infty$ is an i.i.d. sequence of centered chi-square variates with one degree of freedom and represents the discrete version of the Brownian motion increments $dW^{(2)}(\cdot)$. On the other side, the re-normalizing $\sqrt{2}$-term in the last equation of (5) is explained by the fact that $\xi = u^2 - E(u^2) = u^2 - 1$ is a chi-square variate with one degree of freedom and has a variance equal to two. The normality assumption for u is not needed to obtain the convergence.

where

$$\begin{cases} \lim_{h\downarrow 0} h^{-1} w_h & = & \omega \\ \lim_{h\downarrow 0} h^{-1} \left(1-\alpha_h-\beta_h\right) & = & \varphi \\ \lim_{h\downarrow 0} h^{-1/2} \sqrt{2}\alpha_h & = & \psi. \end{cases} \tag{5}$$

Equation (4) may correspond to the volatility dynamics in (1) when $\delta = 2$, $\eta = 1$ and $\rho = 0$. Similarly, it is possible to show that under conditions similar to (5), the so called Taylor-Schwert model:

$$\sigma_{n+1} - \sigma_n = w - \left(1 - \alpha E(|u|) - \beta\right)\sigma_n + \alpha\sigma_n \left(|u_n| - E(|u|)\right),$$

also converges in distribution to the following diffusion limit:

$$d\sigma(\tau) = (\omega - \varphi\sigma(\tau))d\tau + \psi\sigma(\tau)dW^{(2)}(\tau). \tag{6}$$

Equation (6) may now correspond to the volatility dynamics of (1) when $\delta = \eta = 1$ and $\rho = 0$.

As these two basic examples should make clear, standard Arch models do not converge in distribution to any unrestricted CEV process. Rather, in their diffusion limit, Arch models typically make the variance of volatility proportional to the square of volatility, thus restricting the elasticity of variance to unity. Since recent evidence in Durham (2001) shows that Cev models are indeed successful in fitting interest rates data, we introduce an Arch scheme that does not force the elasticity of variance to one.[9] Consider, for instance, the following model:

$$\sigma_{n+1}^2 = w + \alpha\sigma_n^{2\eta} |u_n|^{2\eta} + \beta\sigma_n^2 + \alpha E(|u|^{2\eta}) \left(\sigma_n^2 - \sigma_n^{2\eta}\right), \tag{7}$$

which can also be written as:

$$\sigma_{n+1}^2 - \sigma_n^2 = w - \left(1 - \alpha E(|u|^{2\eta}) - \beta\right)\sigma_n^2 + \alpha\sigma_n^{2\eta} \left(|u_n|^{2\eta} - E(|u|^{2\eta})\right),$$

and which collapses to the Garch(1,1) (3) when $\eta = 1$. In the next section and in Appendix A we show that under conditions similar to those of Nelson (1990), this model converges in

[9] This class of models can be shown to satisfy the most salient theoretical properties of an optimal volatility filter as developed in the optimal filtering theory of Nelson and Foster (1994, theorems 4.1 and 5.2).

distribution to:

$$d\sigma(\tau)^2 = \left(\omega - \varphi\sigma(\tau)^2\right) d\tau + \psi\sigma(\tau)^{2\eta} dW^{(2)}(\tau).$$

Finally, to obtain convergence results closer to model (1), we shall be considering a generalization of (7) that sets the volatility propagation mechanism to:

$$\sigma_{n+1}^{\delta} = w + \alpha\sigma_n^{\delta\eta} |u_n|^{\delta\eta} + \beta\sigma_n^{\delta} + \alpha E(|u|^{\delta\eta}) \left(\sigma_n^{\delta} - \sigma_n^{\delta\eta}\right). \tag{8}$$

As before, we will show that at a high sampling frequency, the volatility process in (8) converges in distribution to

$$d\sigma(\tau)^{\delta} = \left(\omega - \varphi\sigma(\tau)^{\delta}\right) d\tau + \psi\sigma(\tau)^{\delta\cdot\eta} dW^{(2)}(\tau),$$

which may correspond to the volatility dynamics in (1) when $\rho = 0$. Complications arising from the presence of correlation will be treated by introducing asymmetries in the volatility dynamics of (8).[10]

2.1 *Filtering and invariance properties of the Cev-Arch: preliminary Monte Carlo evidence*

The practical relevance of the filtering theory for Arch models can be grasped very simply from Figure 1, which depicts the typical filtering of an Arch model as applied to a simplified version of (1). The straight line is one weekly sampled trajectory of the volatility,

---

[10] In the same way one can introduce nonlinear volatility dynamics into discrete time models that match any desidered feature of the resulting diffusion limit. Consider, for instance, the following model:

$$\sigma_{n+1} = (1+w)\sigma_n - (1 - \alpha E(|u|) - \beta)\,\sigma_n^2 + \alpha\,(|u_n| - E(|u|))\,\sigma_n^{3/2}.$$

Using the methods of section 3, it can then be shown that this model converges in distribution toward:

$$d\sigma(\tau) = \{\sigma(\tau)\,(\omega - \varphi\sigma(\tau))\}\,d\tau + \psi\sigma(\tau)^{3/2} dW^{(2)}(\tau),$$

as the sampling frequency gets higher and higher. Likewise, one can adjust both the short-term and the volatility equation to include both variables. In this paper, however, we will only test the adequacy of Arch-type models in the estimation and filtering of system (1).

$\sigma(\tau)$, simulated within the following model:

$$\begin{cases} dr(\tau) &= (\iota - \theta \cdot r(\tau))d\tau + \sqrt{r(\tau)} \cdot \sigma(\tau) \cdot dW^{(1)}(\tau) \\ d\sigma(\tau) &= (\omega - \varphi \cdot \sigma(\tau))d\tau + \psi \cdot \sigma(\tau) \cdot dW^{(2)}(\tau) \end{cases} \tag{9}$$

where $W^{(i)}$, $i = 1, 2$, are standard Brownian motions, $\iota, \theta, \omega, \varphi$ and ψ are real-valued parameters fixed at their estimates obtained with US data (see section 4). The dotted line represents instead the (rescaled) volatility obtained via an Arch model fitted to the weekly sampled trajectory of the short-term rate $r(\tau)$, as simulated by (9); of course, in estimating the Arch model, we considered ourselves constrained to *only* knowing the realization of the simulated $r(\tau)$. In fact, figure 1 visualizes one of the simulations performed in the Monte Carlo experiment of section 4, but such a performance is typical of the overall experiment; this can be gauged by the very tiny RMSE between the two trajectories computed over all the simulations.[11] More precisely, when we compared the volatility trajectories filtered with equation (8) and conditional on model (1) (with parameters set at the values in Table 5) to those directly simulated from (1), we find that their patterns are very similar and similar also to those of figure 1. Table 1 reports precise results assessing the performance of this volatility filtering based on (8), where the common concept of volatility adopted to make comparisons is the standard deviation. The result is what we call the 'volatility filtering error', which is defined precisely in section 4. The findings reported in Table 1 are of the same order of magnitude as those derived from a much more detailed analysis and illustrated in section 4. Notice also that to compare the simulated volatility to the filtered volatility, the latter has to be "rescaled for diffusions"; techniques for treating this issue are introduced and explained in great detail in Appendix C.

When ascertaining whether (8) is able to deliver reliable parameter estimates beyond a consistent filtering of the unobservable volatility, it would be useful if some of the parameters of the continuous time model could be eliminated from the estimation procedure. Our guess is that δ and η can be fixed at their discrete-time values, after assuming a sort of time-scale invariance. To prove this we can only resort to numerical arguments: we considered model (1) and fixed $\rho = 0$ (consistently with susequent empirical evidence reported in section 4) and the other coefficients at the values in Table 1. We then simulated (1) 1,000 times with

---

[11] In addition to section 4, see Schwartz et al., 1993, for previous related work on similar models.

an Euler-Maruyama approximation and sampled the simulated data at a weekly frequency (in this simulation we allow for 25 intra-week observations). All the simulated weekly paths have 1,135 points, thus matching the sample size used in the empirical analysis (see section 4). Finally, all weekly simulated short-term rate paths were fitted by a conditionally Gaussian AR(1) model of the form:

$$r_n = \phi_0 + \phi_1 r_{n-1} + \sqrt{r_{n-1}}\epsilon_n \quad (\phi_0, \phi_1 \text{ constants}),$$

with (8) as volatility propagation equation.

Table 1 reports the results of the experiment. We begin with the case related to the empirical evidence provided in section 4: there, we find that fitting (8) to actual US short-term rate data produces estimates of δ and η that are both statistically not distinguishable from unity. Now Table 1 shows that when the data generating process in (1) has $\delta = \eta = 1$, then (8) also reproduces, on average, approximately the same ML estimates of δ and η. Results not reported here reveal that the same phenomenon occurs with other possible combinations of δ and η. As an example, Table 1 reports Monte Carlo results concerning the case in which $\delta = 2$ and $\eta = \frac{1}{2}$ in (8). Based on this evidence we remove δ and η from the parameter vector a and fix them at their discrete time ML-based estimates.

2.2 *Some additional characteristics of the continuous time Cev model*

Beyond providing a framework for CEV-type volatility modeling, (1) differs significantly from previous stochastic volatility models, since it does not constrain the 'volatility concept' to be 'variance' or 'standard deviation'; rather, in (1) δ is a new parameter that must be estimated from data. In the empirical section of the paper, for instance, we uncover evidence that $\delta \cong 1$ and, as already stated, that $\eta \cong 1$.[12] To understand the influence of δ on the dynamics of σ_t^δ it may be interesting to recall that with $\eta = 1$ and positive mean-reversion, the volatility process σ^δ, $\delta \geq 1$, has a steady state distribution that is an inverted Gamma with mean $\frac{\varpi}{\varphi}$ (e.g., lemma 3.1 p. 217 in Fornari and Mele, 1997a); the stationary distribution of σ

[12] Engle and Lee (1996) fitted a restricted version of the volatility equation of model (1) to stock returns, namely for $\delta = 2$, and supported a model in which the volatility of volatility raised linearly with the square of volatility, as our empirical findings do.

is consequently given by

$$f_\delta(\sigma) \equiv \frac{\delta \cdot \left(\frac{2\omega}{\psi^2}\right)^{\frac{2\varphi+\psi^2}{\psi^2}}}{\Gamma\left(\frac{2\varphi+\psi^2}{\psi^2}\right)} \sigma^{-\frac{2\delta\varphi+(\delta+1)\psi^2}{\psi^2}} \exp\left(-\frac{2\omega}{\psi^2}\sigma^{-\delta}\right) \tag{10}$$

(see lemma A.2, p. 227, in Fornari and Mele, 1997a). As shown by Fornari and Mele (2000;chapter 5), the density $f_\delta(\cdot)$ tends to shrink to the left as δ decreases.

The volatility equation in (1) encompasses other formulations already encountered in the stochastic volatility literature (see, for instance, Ball and Roma, 1994, and Taylor, 1994, for a list of the typical models in the stochastic volatility option pricing area). This is the case, for instance, of the non-stationary models of Hull and White (1987) or Johnson and Shanno (1987), to which our volatility equation reduces when $\omega \equiv 0$. By Itô's lemma, indeed, $\mathcal{V} \equiv \log \sigma^2$ is solution of

$$d\mathcal{V}(\tau) = \left(-\frac{2\varphi+\psi^2}{\delta} + 2\frac{\omega}{\delta}\exp\left(-\frac{\delta}{2}\mathcal{V}(\tau)\right)\right)d\tau + \frac{2\psi}{\delta}d\left(\rho W^{(1)} + \sqrt{1-\rho^2}W^{(2)}\right). \tag{11}$$

Hence log-volatility mean-reverts in a *non-linear* manner when $\omega \neq 0$. Therefore, (11) is rather different from the *linear* mean-reverting process for the log-volatility adopted in Wiggins (1987) in a stochastic volatility option pricing domain and in Andersen and Lund (1997a) or Gallant and Tauchen (1998) in a interest rate framework. To see this in more detail, consider the linear mean reverting model utilized in Andersen and Lund,

$$d\mathcal{V}(\tau) = \left(\overline{\alpha} - \overline{\beta}\mathcal{V}(\tau)\right)d\tau + \overline{\xi}dW(\tau)$$

where W is a standard Brownian motion and $\overline{\alpha}, \overline{\beta}, \overline{\xi}$ are real constants. By Itô's lemma, in this model σ^δ is the solution of

$$d\sigma(\tau)^\delta = \left(\frac{4\overline{\alpha}\delta + \overline{\xi}^2\delta^2}{8}\sigma(\tau)^\delta - \overline{\beta}\sigma(\tau)^\delta \cdot \log\sigma(\tau)^\delta\right)d\tau + \frac{\overline{\xi}\delta}{2}\sigma(\tau)^\delta dW(\tau), \tag{12}$$

which becomes of course also the starting point of Wiggins (1987 eq. (2) p. 353 and eq. (15) p. 361) when $\delta \equiv 1$. Although the volatility of volatility in (12) rises linearly with $\sigma^{2\delta}$, as in (1) when $\eta = 1$, the drift behaves rather differently in the two volatility equations.

Figure 2 (panel A) depicts a comparison between the stationary densities that are generated by (11) and (12). The first is given by (10) and has been produced using the estimates of section 4; the latter is just a log-normal density, and has been produced using the estimates of Andersen and Lund (1997b). While the two models approximately put the same probability masses on low levels of volatility, our model puts relatively more masses on high values of volatility. An explanation of such a phenomenon can be found by comparing the drift functions of the two models: as is clear from figure 2 (panel B), the two drift functions are of the same order of magnitude when volatility is low; once volatility visits higher regions, however, the Andersen-Lund linear drift function pulls volatility towards its steady state expected value more rapidly than the drift function of our model. This implies that our model generates relatively more frequent episodes of high volatility than the Andersen-Lund model. Naturally, our model does not encompass the Andersen-Lund scheme, but it should be more flexible in practice due to the presence of the additional parameter δ in the volatility equation: should the volatility equation in (1) be misspecified, such an additional parameter might give the model additional flexibility in fitting the statistical properties of the true volatility generating mechanism.

3. Statistical inference

As reported in the Introduction, various methods have been recently proposed to estimate the parameters of a diffusion when sampling is not continuous, a situation in which the main difficulty of ML methods arises from the likelihood function implied by the measure induced by a discretely sampled diffusion being unavailable in explicit form.[13]

In this paper we follow the natural alternative consisting in making use of a (tractable) exact likelihood function of a class of approximating models. The main idea, presented in the previous section, consists in resorting to a suitably chosen class of ARCH models converging in distribution to the solution of (1) as the sampling frequency gets infinite. However, since the resulting likelihood function refers to a model converging in distribution to the solution of (1)

---

[13] Following Lo (1988), ML estimation might turn out to be feasible if the transition density of $\{r(\tau)\}_{\tau \geq 0}$ in (1) could be computed easily. Since this is not the case here — as in virtually all continuous time stochastic volatility models — ML is computationally demanding, since it would require to implement a numerical solution to a multi-dimensional partial differential equation at each iteration of the optimization algorithm. The likelihood would then be recovered by integrating out with respect to volatility.

that is *not* an Euler approximation of (1), we call the resulting criterion 'quasi-approximated' likelihood function.

The advantage of the quasi-approximated ML estimator is that it requires negligible computational efforts. Its main drawback is that it is not necessarily consistent, as Arch models are typically not closed under temporal aggregation; owing to this, a one-to-one correspondence between convergence in distribution of the discrete time models and disaggregation from a diffusion is not guaranteed.[14] To quantify such potential drawbacks of the procedure, we show how to construct a very precise testing procedure of the validity of the moment conditions needed to guarantee the convergence to well-defined diffusion limits; as it turns out, such a testing procedure also gives information about the relevance of the disaggregation bias. Our strategy is based on the consistency test originally suggested by Gouriéroux et al. (1993, section 4.2), and it can be viewed as the natural substitute of a global specification test in just-identified problems.

3.1 *Quasi-approximated likelihood functions*

The rationale behind the quasi-approximated ML estimator that we propose lies in the weak convergence of a class of Arch models towards the solution of (1). We start with considering the restricted version of (1) that sets $\eta \equiv 1$; theorem 3.2 below treats the general case. With $\eta = 1$, a model approximating (1) can be a discrete time approximation of the short-term rate equation in (1) modified by introducing the so-called asymmetric-power Arch model of Ding et al. (1993):

$$\begin{cases} {}_{\Delta}r_{n+1} &= {}_{\Delta}r_n + \iota_{\Delta} - \theta_{\Delta} \cdot {}_{\Delta}r_n + {}_{\Delta}\sigma_{n+1}\sqrt{{}_{\Delta}r_n} \cdot {}_{\Delta}u_{n+1} \\ {}_{\Delta}\epsilon_n &= {}_{\Delta}u_n \cdot {}_{\Delta}\sigma_n, \quad \frac{{}_{\Delta}u_n}{\sqrt{\Delta}} \sim N(0,1) \\ {}_{\Delta}\sigma^{\delta}_{n+1} &= w_{\Delta} + \alpha_{\Delta}(|{}_{\Delta}\epsilon_n| - \gamma \cdot {}_{\Delta}\epsilon_n)^{\delta} + \beta_{\Delta} \cdot {}_{\Delta}\sigma^{\delta}_n \end{cases} \tag{13}$$

where the indexing $n = 0, 1, \cdots$ refers to consecutive observations sampled at the same frequency Δ (weekly, say), $\iota_{\Delta}, \theta_{\Delta}, w_{\Delta}$ are of the form $x_{\Delta} = x^{(\Delta)} \cdot \Delta$, with $\iota^{(\Delta)}, \theta^{(\Delta)}$ real parameters and $w^{(\Delta)} > 0, \alpha_{\Delta}, \beta_{\Delta} \geq 0, \gamma \in (-1, 1), \delta > 0$. Finally, γ allows for the

[14] Theoretically, such a correspondence exists only when the concept of an ARCH model is weakened (Drost and Nijman, 1993, and Drost and Werker, 1996). Furthermore, Corradi (2000) recently criticized the conditions in Nelson (1990) , necesary to achieve the convergence of the basic GARCH(1,1) to a diffusion; see footnote 16 for details on how to adapt Corradi's critique to our setup.

leverage effect originally observed by Black (1976), and incorporated by Nelson (1991) in Arch-type models. To keep things relatively simple, we take advantage of the assumed time-scale invariance for (δ, η); we also assume that γ shares the same property.

Heuristically, to obtain the weak convergence towards the solution of (1), chop time as $hk \leq n \leq h(k+1)$:

$$\begin{cases} {}_h r_{h(k+1)} & = {}_h r_{hk} + \iota_h - \theta_h \cdot {}_h r_{hk} + {}_h\sigma_{h(k+1)}\sqrt{{}_h r_{hk}} \cdot {}_h u_{h(k+1)} \\ {}_h \epsilon_{hk} & = {}_h u_{hk} \cdot {}_h\sigma_{hk}, \quad \frac{{}_h u_{hk}}{\sqrt{h}} \sim N(0,1) \\ {}_h\sigma^{\delta}_{h(k+1)} - {}_h\sigma^{\delta}_{hk} & = w_h - (1 - \alpha_h \left|{}_h u_{hk}\right|^{\delta} (1 - \gamma s_k)^{\delta} h^{-\frac{\delta}{2}} - \beta_h) {}_h\sigma^{\delta}_{hk} \end{cases} \tag{14}$$

(with $s_k = \text{sign}({}_h u_{hk})$ and, $\forall h > 0, (\{\iota_h\}, \{\theta_h\}\{w_h\}, \{\alpha_h\}, \{\beta_h\}) \in \mathbb{R}^5_+$ and $\gamma \in (-1, +1)$), and impose suitable Lipschitz conditions on the 'h-drift' as well as non-explosion conditions on the '$\sqrt{h}$-diffusion' terms of volatility.

Nelson (1996, p. 19) was one of the first to suggest a model of the kind of (14) as a discrete time approximation of a continuous time model for the short-term rate. More specifically, Nelson (1996) took $\delta \equiv 2$ and $\gamma \equiv 0$ in (14), and pointed out that the resulting scheme is the model of Brenner et al. (1996), slightly changed to admit a diffusion limit. While the empirical results of this paper suggest a simplification of (1) in which δ is one and ρ is nil, we provide here more general results that can be useful when applied to different data sets and/or related problems. As originally remarked by Nelson (1996), the kind of results that we are going to provide can be useful especially when a researcher is interested in the filtering performances of model (13) when ρ is not nil in (1).[15]

To save space, we shall be avoiding technical discussion on the construction of the measure space in (14): technical details can be found in Nelson (1990) and are those exploited in Fornari and Mele (1997a, b and 2000) (see also Duan, 1997, for related work). We only introduce notation for the filtration generated by $\{{}_h r_{h(i-1)}, {}_h\sigma^{\delta}_{hi}\}_{i=1}^{k}$, which is $\mathcal{F}_{hk}$, and which will be used in appendix A. Let the symbol $\Rightarrow$ denote weak convergence. Recall that if a random variable x is general error distributed then its density is written as $\frac{\upsilon \exp(-\frac{1}{2}\nabla_{\upsilon}^{-\upsilon}|x|^{\upsilon})}{2^{1+\upsilon^{-1}}\nabla_{\upsilon}\Gamma(\upsilon^{-1})}$,

---

[15] We slightly complicate the theoretical analysis allowing standardized residuals to be general error distributed; however, such a possibility is not subsequently considered in the empirical section of the paper, since Engle and Lee (1996) (see their tables 2 and 4) obtained indirect estimates that seemed to be dependent on the distributional assumption made for the auxiliary model.

where $\nabla_\upsilon^2 \equiv \frac{\Gamma(\upsilon^{-1})}{2^{2/\upsilon}\Gamma(3\upsilon^{-1})}, \upsilon > 0$ and $\Gamma(.)$ is the Gamma function. The following convergence result is an extension of theorem 2.3 p. 211 in Fornari and Mele (1997a) that allows for the presence of the instantaneous correlation between $\{{}_h r_{hk}\}_{k=0,1,\cdots}$ and $\{{}_h\sigma_{hk}^\delta\}_{k=0,1,\cdots}$ as h shrinks to nil:

THEOREM 3.1: *Let* $m_{\delta,\upsilon} = \frac{2^{\frac{2\delta}{\upsilon}-1}\nabla_\upsilon^{2\delta}\Gamma(\frac{2\delta+1}{\upsilon})}{\Gamma(\upsilon^{-1})}$, $n_{\delta,\upsilon} = \frac{2^{\frac{\delta}{\upsilon}-1}\nabla_\upsilon^{\delta}\Gamma(\frac{\delta+1}{\upsilon})}{\Gamma(\upsilon^{-1})}$, *and let* $\frac{{}_h u_{hk}}{\sqrt{h}}$ *be general error distributed. Let*:

$$\text{(15)} \quad \varphi_h \equiv 1 - n_{\delta,\upsilon}((1-\gamma)^\delta + (1+\gamma)^\delta)\alpha_h - \beta_h,$$

$$\psi_h \equiv \sqrt{(m_{\delta,\upsilon} - n_{\delta,\upsilon}^2)((1-\gamma)^{2\delta} + (1+\gamma)^{2\delta}) - 2n_{\delta,\upsilon}^2(1-\gamma)^\delta(1+\gamma)^\delta}\cdot\alpha_h,$$

$$\text{(16)} \quad \rho \equiv \frac{2^{\frac{\delta-\upsilon+1}{\upsilon}}\nabla_\upsilon^{\delta+1}\Gamma(\frac{\delta+2}{\upsilon})\cdot((1-\gamma)^\delta - (1+\gamma)^\delta)}{\Gamma(\upsilon^{-1})\sqrt{(m_{\delta,\upsilon} - n_{\delta,\upsilon}^2)((1-\gamma)^{2\delta} + (1+\gamma)^{2\delta}) - 2n_{\delta,\upsilon}^2(1-\gamma)^\delta(1+\gamma)^\delta}}$$

and suppose that $\lim_{h\downarrow 0} h^{-1}\iota_h = \iota, \lim_{h\downarrow 0} h^{-1}\theta_h = \theta$ *and*:

$$\begin{array}{lcl} \lim_{h\downarrow 0} h^{-1}w_h & = & \omega \in (0,\infty), \\ \lim_{h\downarrow 0} h^{-1}\varphi_h & = & \varphi < \infty, \\ \lim_{h\downarrow 0} h^{-1/2}\psi_h & = & \psi < \infty. \end{array} \qquad (17)$$

Then, $\{{}_h r_{h(k-1),h}\,\sigma_{hk}^\delta\}_{k=0,1,\cdots} \Rightarrow \{r(\tau), \sigma(\tau)^\delta\}_{\tau\geq 0}$ *as* $h \downarrow 0$, *where* $\{r(\tau), \sigma(\tau)^\delta\}_{\tau\geq 0}$ *are solutions of* (1) *when* $\eta \equiv 1$.

Let, in addition:

$${}_h\xi_{hk} \equiv \frac{\left|\frac{{}_h u_{hk}}{\sqrt{h}}\right|^\delta (1-\gamma s_k)^\delta - E(\left|\frac{{}_h u_{hk}}{\sqrt{h}}\right|^\delta (1-\gamma s_k)^\delta)}{\sqrt{(m_{\delta,\upsilon} - n_{\delta,\upsilon}^2)((1-\gamma)^{2\delta} + (1+\gamma)^{2\delta}) - 2n_{\delta,\upsilon}^2(1-\gamma)^\delta(1+\gamma)^\delta}},$$

then the preceding approximation result says that when h shrinks to zero and the moment conditions in (17) are fulfilled, the distribution of the sample paths generated by the following

model,

$$\begin{cases} {}_h r_{h(k+1)} - {}_h r_{hk} &= (\iota_h - \theta_h \cdot {}_h r_{hk}) + {}_h\sigma_{h(k+1)}\sqrt{{}_h r_{hk}} \cdot {}_h u_{h(k+1)} \\ {}_h\sigma^{\delta}_{h(k+1)} - {}_h\sigma^{\delta}_{hk} &= (w_h - \varphi_h \cdot {}_h\sigma^{\delta}_{hk}) + \psi_h \cdot {}_h\sigma^{\delta}_{hk} \cdot {}_h\xi_{hk} \end{cases} \tag{18}$$

gets 'closer and closer' to the distribution generated by the sample paths generated by (1), with ρ given by (16). Comparing (13) to (18) then suggests an estimator based on moment conditions; specifically, the *quasi-approximated* ML (q-aml) estimators of $\overline{\omega}, \varphi, \psi$ that we propose are

$$\begin{array}{rcl} \omega_{\text{q-aml}} &\equiv& \Delta^{-3/2}\widehat{w}_{\Delta}, \\ \varphi_{\text{q-aml}} &\equiv& \Delta^{-1}\widehat{\varphi}_{\Delta}, \\ \psi_{\text{q-aml}} &\equiv& \Delta^{-1/2}\widehat{\psi}_{\Delta}, \end{array} \tag{19}$$

where $\widehat{\varphi}_{\Delta}, \widehat{\psi}_{\Delta}$ are obtained by means of (15)-(16) computed in correspondence of the qml estimator of model (13), $\widehat{w}_{\Delta}$ is the qml estimator of w_{Δ} of model (13). The q-aml estimator of δ is the qml estimator of δ in model (13), and the q-aml estimators of ι and θ are as those of ω and φ above. Finally, the q-aml estimator of ρ is obtained by plugging the qml estimators of $(\delta, \upsilon, \gamma)$ in formula (16).[16]

While recognizing that weak convergence results such as those contained in theorem 3.1 are obviously related to parametrization issues, in the empirical section we find that not only the parametrization in (14) provides a reasonably good picture of the volatility dynamics, consistently with the theoretical results of Nelson and Foster (1994), but it even passes the

---

[16] The estimators in (19) are based on the moment conditions (17) and as we noted before, they may be affected by a disaggregation bias; furthermore, Corradi (2000) questions the realism of the moment conditions that Nelson (1990) originally imposed to show the weak convergence of the Garch(1,1) towards a continuous time stochastic volatility model. Her reasoning can be generalized here as follows. In the third equation of (14), the term generating the diffusion terms of volatility is proportional to $(h^{-\frac{\delta}{2}}\alpha_h) \cdot |{}_h u_{hk}|^{\delta}$, which is of course $O_p(\sqrt{h})$ under the third moment condition in (17). In other terms, a condition for a diffusion to be obtained is to scale the variance of $|{}_h u_{hk}|^{\delta}$ with a diverging sequence. In general, one would generate diffusion terms with $\alpha_h \cdot |{}_h u_{hk}|^{\delta}$, where $\alpha_h \approx O(h^q), q \in \mathbb{R}$. This leaves three alternatives:

- a) $q = \frac{1-\delta}{2}$.
- b) $q < \frac{1-\delta}{2}$.
- c) $q > \frac{1-\delta}{2}$.

The first condition is another way to express the condition under which (14) has a well-defined diffusion limit; the second condition implies that (14) does not converge to any diffusion limit; the third condition implies a 'degenerate' diffusion limit, i.e. with identically zero diffusion terms.

global consistency test that checks, ex-post, the accuracy of the approximation in (15)-(16) and that we present below.

Consider now the case of interest to this paper based on generalizing both (7) and (8) by means of the following model:

$$\begin{cases} {}_{\Delta}r_{n+1} &= {}_{\Delta}r_n + \iota_{\Delta} - \theta_{\Delta} \cdot {}_{\Delta}r_n + {}_{\Delta}\sigma_{n+1}\sqrt{{}_{\Delta}r_n} \cdot {}_{\Delta}u_{n+1} \\ {}_{\Delta}\epsilon_n &= {}_{\Delta}u_n \cdot {}_{\Delta}\sigma_n, \quad \frac{{}_{\Delta}u_n}{\sqrt{\Delta}} \sim ged_{\upsilon} \\ {}_{\Delta}\sigma^{\delta}_{n+1} &= w_{\Delta} + \alpha_{\Delta}(|{}_{\Delta}\epsilon_n| - \gamma \cdot {}_{\Delta}\epsilon_n)^{\delta\eta} + \beta_{\Delta} \cdot {}_{\Delta}\sigma^{\delta}_n \\ & \quad +\alpha_{\Delta} \cdot E\left\{(|{}_{\Delta}u_n| - \gamma \cdot {}_{\Delta}u_n)^{\delta\eta}\right\} \cdot \left\{{}_{\Delta}\sigma^{\delta}_n - {}_{\Delta}\sigma^{\delta\eta}_n\right\}. \end{cases} \tag{20}$$

Chopping time in (20) as in (13), and rearranging, yields:

$$\begin{cases} {}_{h}r_{h(k+1)} &= {}_{h}r_{hk} + \iota_h - \theta_h \cdot {}_{h}r_{hk} + {}_{h}\sigma_{h(k+1)}\sqrt{{}_{h}r_{hk}} \cdot {}_{h}u_{h(k+1)} \\ {}_{h}\epsilon_{hk} &= {}_{h}u_{hk} \cdot {}_{h}\sigma_{hk}, \quad \frac{{}_{h}u_{hk}}{\sqrt{h}} \sim ged_{\upsilon} \\ {}_{h}\sigma^{\delta}_{h(k+1)} - {}_{h}\sigma^{\delta}_{hk} &= w_h - \left(1 - h^{-\frac{\delta\eta}{2}} E\left\{|{}_{h}u_{hk}|^{\delta\eta}(1-\gamma s_k)^{\delta\eta}\right\}\alpha_h - \beta_h\right) {}_{h}\sigma^{\delta}_{hk} \\ & \quad +\alpha_h \cdot \left(|{}_{h}u_{hk}|^{\delta\eta}(1-\gamma s_k)^{\delta\eta} - E\left\{|{}_{h}u_{hk}|^{\delta\eta}(1-\gamma s_k)^{\delta\eta}\right\}\right) h^{-\frac{\delta\eta}{2}}\sigma^{\delta\eta}_{hk}. \end{cases} \tag{21}$$

We have:

THEOREM 3.2: *Let*

$$\varphi_h \equiv 1 - n_{\delta\eta,\upsilon}((1-\gamma)^{\delta\eta} + (1+\gamma)^{\delta\eta})\alpha_h - \beta_h, \tag{22}$$

$$\psi_h \equiv \sqrt{(m_{\delta\eta,\upsilon} - n^2_{\delta\eta,\upsilon})((1-\gamma)^{2\delta\eta} + (1+\gamma)^{2\delta\eta}) - 2n^2_{\delta\eta,\upsilon}(1-\gamma)^{\delta\eta}(1+\gamma)^{\delta\eta}} \cdot \alpha_h, \tag{23}$$

and

$$\rho \equiv \frac{2^{\frac{\delta\eta-\upsilon+1}{\upsilon}}\nabla^{\delta\eta+1}_{\upsilon}\Gamma(\frac{\delta\eta+2}{\upsilon}) \cdot ((1-\gamma)^{\delta\eta} - (1+\gamma)^{\delta\eta})}{\Gamma(\upsilon^{-1})\sqrt{(m_{\delta\eta,\upsilon} - n^2_{\delta\eta,\upsilon})((1-\gamma)^{2\delta\eta} + (1+\gamma)^{2\delta\eta}) - 2n^2_{\delta\eta,\upsilon}(1-\gamma)^{\delta\eta}(1+\gamma)^{\delta\eta}}}. \tag{24}$$

Suppose that $\lim_{h\downarrow 0} h^{-1}\iota_h = \iota$, $\lim_{h\downarrow 0} h^{-1}\theta_h = \theta$ *and*:

$$\begin{array}{lcl} \lim_{h\downarrow 0} h^{-1} w_h & = & \omega \in (0,\infty), \\ \lim_{h\downarrow 0} h^{-1}\varphi_h & = & \varphi < \infty, \\ \lim_{h\downarrow 0} h^{-1/2}\psi_h & = & \psi < \infty. \end{array} \tag{25}$$

Then, $\{{}_h r_{h(k-1),h}\,\sigma^{\delta}_{hk}\}_{k=0,1,\cdots} \Rightarrow \{r(\tau),\sigma(\tau)^{\delta}\}_{\tau\geq 0}$ as $h \downarrow 0$, where $\{r(\tau),\sigma(\tau)^{\delta}\}_{\tau\geq 0}$ are solutions of (1) and $\{{}_h r_{h(k-1),h}\,\sigma^{\delta}_{hk}\}_{k=0,1,\cdots}$ are solution of (21).

In the same way one can make a creative use of other asymmetric Arch models to obtain convergence to models with correlated Brownian motions. We briefly show this in Appendix A.

3.2 *Quasi indirect inference*

We test and correct the potential disaggregation bias of the q-aml estimator with the indirect inference principle. The procedure that we follow is a natural generalization of Broze et al. (1995) and allows the volatility of the short-term rate to evolve in a stochastic and *autonomous* manner. Formally, if we replace the normality assumption with the g.e.d. assumption for the innovation process u in (21) (see section 3.1), the q-aml estimator of $b = (\Delta^{-1}\iota_\Delta,\ \Delta^{-1}\theta_\Delta,\ \Delta^{-3/2}w_\Delta,\ \Delta^{-1}\varphi_\Delta,\ \Delta^{-1/2}\psi_\Delta,\ \gamma,\ \delta,\ \eta,\ \upsilon)'$ in (20) is (where υ is the tail-thickness of the ged distribution):

$$a_{\text{q-aml}} \equiv \widehat{b}_N = \arg\max_b \mathfrak{L}_N({}_\Delta r; b),$$

where $\mathfrak{L}_N({}_\Delta r; b)$ is the likelihood function implied by (20), N is the sample size, and ${}_\Delta r$ is the observations set, which is supposed to be a discretely sampled diffusion from (1) when the true parameter vector is a_0. Note that $\dim(b) > \dim(a)$. In the empirical implementation below, however, we shall consider the Gaussian case in which $\upsilon \equiv 2$ and, motivated by the Monte Carlo findings reported before, we impose the time-scale invariance of δ and η. As already stated, we assume the same for γ, leaving for future research the task of ascertaining whether such a time-scale invariance of γ is a reasonable assumption in practice. Accordingly, we re-interpret b as a vector in an open subset of $\mathbb{R}^5$ (with coordinates $\Delta^{-1}\iota_\Delta$, $\Delta^{-1}\theta_\Delta$, $\Delta^{-3/2}w_\Delta$, $\Delta^{-1}\varphi_\Delta$, $\Delta^{-1/2}\psi_\Delta$), $\mathfrak{L}_N(.)$ as a normal likelihood function with δ, η and γ fixed at pre-specified values (e.g. at the preliminary qml estimates obtained by fitting model (20), see section 4), and a as a vector in an open subset of $\mathbb{R}^5$, with coordinates $\iota, \theta, \overline{\omega}, \varphi, \psi$.

It is well known that under standard regularity conditions (appendix B), one has asymptotic normality of the pseudo-ML estimator,

$$\sqrt{N}\left(\widehat{b}_N - b_0(a_0)\right) \xrightarrow{d} \mathrm{N}\left(0, \ddot{\mathfrak{L}}_\infty^{-1}(a_0; b_0(a_0)) \cdot J(a_0) \cdot \ddot{\mathfrak{L}}_\infty^{-1}(a_0; b_0(a_0))\right),$$

where $\ddot{\mathfrak{L}}_\infty(.)$ and $J(.)$ are defined in appendix B, and $b_0(.)$ is the so-called *binding function*:

$$b_0(a_0) = \arg\max_b \mathfrak{L}_\infty(a_0; b).$$

However, the true law of ${}_\Delta r$, as implied by the data generating mechanism, say $\ell_0({}_\Delta r)$, is such that

$$\ell_0({}_\Delta r) \notin \{\mathfrak{L}_N({}_\Delta r; b),\ b \text{ varying}\},$$

and the discrete time model is expected to behave in a way that allows for a discretization bias:

$$b(a_0) \neq a_0.$$

The reason why we may also refer to the preceding inequality as a 'discretization bias' is that when we chop time in (20) by creating sequences of the form $\{\iota_h, \theta_h, w_h, \alpha_h, \beta_h\}$, and substitute the moment conditions (22)-(25) of theorem 3.2 in (21), thereby creating a stochastic process $\{{}_h r_{hk}, {}_h\sigma^\delta_{hk}\}_{k=0,1,\cdots}$ solution of:

$$\begin{cases} {}_h r_{h(k+1)} - {}_h r_{hk} &= (\iota - \theta \cdot {}_h r_{hk})h + {}_h\sigma_{h(k+1)}\sqrt{{}_h r_{hk}} \cdot {}_h u_{h(k+1)} \\ {}_h\sigma^\delta_{h(k+1)} - {}_h\sigma^\delta_{hk} &= (\omega - \varphi \cdot {}_h\sigma^\delta_{hk})h + \psi \cdot {}_h\sigma^{\delta\eta}_{hk}\sqrt{h}\,{}_h\xi_{hk}, \end{cases} \tag{26}$$

then (20) is embedded in $\{{}_h r_{hk}, {}_h\sigma^\delta_{hk}\}_{k=0,1,\cdots}$ (namely for $h \equiv \Delta$), although $\{{}_h r_{hk}, {}_h\sigma^\delta_{hk}\}_{k=0,1,\cdots}$ converges weakly to the solution of (1) under the limit conditions given in theorem 3.2.

Indirect inference methods correct the preceding bias in the following manner. Consider simulating (26) for small h. This is accomplished by setting γ, δ and η to their ML estimates $\widehat{\gamma}, \widehat{\delta}, \widehat{\eta}$, assigning values to $a = (\iota, \theta, \omega, \varphi, \psi)$, and drawing $\frac{{}_h u_{hk}}{\sqrt{h}}$ from the normal distribution; one obtains ${}_{h,h}\widetilde{r}^{(s)}(a) = \{{}_h\widetilde{r}^{(s)}_{hk}(a)\}_{k=0}^{N/h}, s = 1, \cdots, S$, where S is the number of simulations. For each simulation retain the (N) numbers ${}_h\widetilde{r}^{(s)}_{hk}(a)$ that correspond to integer indexes of time,

and estimate the auxiliary model on each series of simulated data:

$$\widehat{b}_{N,s}^{(h)}(a) = \arg\max_{b} \mathfrak{L}_N({}_{\Delta,h}\widetilde{r}^{(s)}(a); b), \quad s = 1, \cdots, S,$$

where ${}_{\Delta,h}\widetilde{r}^{(s)}(.)$ denotes the set of the simulated short-term rate with integer indexes of time at simulation s and interval h. In our specific just-identified problem ($\dim(a) = \dim(b)$), the indirect estimator of a is then the solution (provided it exists) of the following five-dimensional system:

$$0 = \widehat{b}_N - \frac{1}{S}\sum_{s=1}^{S} \widehat{b}_{N,s}^{(h)}(a).$$

If ${}_h\widehat{a}_N(a_0)$ denotes the solution of the preceding system, its asymptotic distribution can be obtained, heuristically, as follows. Expand the preceding system of equalities around a_0:

$$\widehat{b}_N - \frac{1}{S}\sum_{s=1}^{S} \widehat{b}_{N,s}^{(h)}(a_0) = \left(\frac{1}{S}\sum_{s=1}^{S} \frac{\partial \widehat{b}_{N,s}^{(h)}}{\partial a}(a_0)\right) ({}_h\widehat{a}_N(a_0) - a_0)\,.$$

For large N, the preceding is in fact an equality in distribution, and the covariance matrix of $(\frac{1}{S}\sum_{s=1}^{S} \frac{\partial \widehat{b}_{N,s}^{(h)}}{\partial a}(a_0))({}_h\widehat{a}_N(a_0) - a_0)$ is the covariance matrix of $\widehat{b}_N - \frac{1}{S}\sum_{s=1}^{S} \widehat{b}_{N,s}^{(h)}(a_0)$, i.e. $(1 + \frac{1}{S})\text{cov}(\widehat{b}_{N,s}^{(h)}(a_0))$, and one has:

$$\sqrt{N}\,({}_h\widehat{a}_N(a_0) - a_0)\; N \uparrow \infty,\overset{d}{,}\; h \downarrow 0 \rightarrow \text{N}\left(0, \left(1 + \frac{1}{S}\right) V_0^{-1}\Gamma_0 V_0'^{-1}\right), \tag{27}$$

where Γ_0 is the covariance matrix of the simulated estimator and $V_0 \equiv \frac{\partial b}{\partial a}(a_0)$, i.e. the Jacobian of the binding function evaluated at a_0. Broze et al. (1998) proved the preceding result in great generality — i.e. in the case of a general diffusion in $\mathbb{R}^l$ — and to avoid bias due to the discretization step used during the simulations (hence the label 'quasi'-indirect inference), they also suggested to take $h = N^{-d}$ with $d > \frac{1}{2}$. In appendix B we check the conditions of Broze et al. (1998) that ensure that (27) holds for the scheme proposed here.

Notice also that (26) do *not* represent the Euler approximation of (1), but this is not a disturbing feature for it has been known since Broze et al. (1998) that implementing the indirect inference estimator just requires the weak convergence of the high frequency simulator toward the solution of (1); see also appendix B. For reasons of comparisons, however, the

empirical section also considers the case in which the high frequency simulator is the Euler-Maruyama approximation of (1) (i.e. (2)).

Finally, a global specification test for the adequacy of the approximating model (20) is easily implemented. It is sufficient to use the consistency test of Gouriéroux et al. (1993, section 4.2 and appendix 3), designed to verify the existence of a fixed point of the binding function:

$$\mathrm{H}_0 : a_0 = b(a_0).$$

Let I denote the identity matrix in $\mathbb{R}^{5\times 5}$. Under H_0, one has that (see Appendix B for the definition of the terms $\ddot{\mathfrak{L}}_\infty$ and J):

$$\sqrt{N}\left(\widehat{b}_N - \frac{1}{S}\sum_{s=1}^{S}\widehat{b}^{(h)}_{N,s}(\widehat{b}_N)\right)$$

$$\xrightarrow{d} \mathrm{N}\left(0, \left(I - \frac{\partial b}{\partial a}(a_0)\right)\ddot{\mathfrak{L}}_\infty^{-1}\, J\, \ddot{\mathfrak{L}}_\infty^{-1}\left(I - \frac{\partial b'}{\partial a}(a_0)\right) + \frac{1}{S}\ddot{\mathfrak{L}}_\infty^{-1}\, J\, \ddot{\mathfrak{L}}_\infty^{-1}\right). \tag{28}$$

4. Empirical analysis

4.1 *The data*

We use weekly data for the 3-month US T-Bill rates as an approximation to the short-term rate.[17] The motivation for using weekly data lies in an attempt of avoiding problems raised by market microstructure effects. This is the same data set used by Andersen and Lund (1997a,b), but here we restrict attention to the sample going from May 30, 1973 to February 22, 1995, which has 1135 observations.

Raw data are converted into instantaneous figures, hereafter referred to as r. Table 2 contains some preliminary statistics for r and its autocorrelation function, showing high persistence in the data. Nonstationarity is formally tested through an augmented Dickey-Fuller test, according to which data are borderline stationary. As an example, the statistic takes a value of -2.435 at lag 5, which is roughly the threshold value for rejecting nonstationarity with a 90 percent probability; more generally, one rejects nonstationarity at the 85-90 percent

---

[17] See Chapman et al. (1999) for an analysis concerning the validity of such an approximation.

to the extent of the 15-th lag, but given the low power of the test, even such a slight rejection can be symptomatic of stationarity in the data. It is worth noticing that the same kind of results holds for the full sample originally employed in Andersen and Lund.

4.2 *Fitting the short rate: auxiliary discrete time model*

We start by estimating model (20). Consistently with previous results of Andersen and Lund (1997a) we do not find evidence of leverage effects, since the estimate of γ is not statistically significant; further, the model gives rise to stable dynamics for the volatility process. As regards the estimates of δ and η, we find that they are 1.0326 and 1.0014, respectively, statistically not distinguishable from unity, which allows, as mentioned in the last section, an additional simplification to model (1) by fixing $\delta = \eta = 1$. Such restrictions, along with $\gamma = 0$, will propagate into a much faster indirect inference phase. In the model that we select as an auxiliary device, we thus restrict $(\delta, \eta, \gamma) \equiv (1, 1, 0)$. Due to numerical stability issues, model (20) was estimated without explicitly disentangling the sample frequency, i.e. under the restrictions $(\delta, \eta, \gamma) \equiv (1, 1, 0)$, the model was cast as:

$$\begin{cases} r_n &= c_0 + c_1 r_{n-1} + r_{n-1}^{1/2} \cdot \epsilon_n, \quad \epsilon_n \equiv (u \cdot \sigma)_n, \quad u \sim NID(0,1) \\ \sigma_n &= w + \alpha \left|\epsilon_{n-1}\right| + \beta \sigma_{n-1}, \quad n = 2, \cdots, N, \end{cases} \tag{29}$$

where $\{r_n\}_{n=1}^N$ denotes the observed (weekly) series, and $(c_0, c_1, w, \alpha, \beta)$ are real parameters. The correspondence between the estimators of the parameters in (20) and (29) is easily written as:

$$\widehat{b}_N \equiv a_{\text{q-aml}} = \Delta_0 + \Delta_1 \widehat{m}_N,$$

where $\widehat{m}_N$ denotes the vector of the ML estimators of the parameters in (29), $\Delta_0 = (\ 0 \quad \Delta^{-1} \quad 0 \quad \Delta^{-1} \quad 0\ )'$, and

$$\Delta_1 = \begin{pmatrix} \Delta^{-1} & 0 & 0 & 0 & 0 \\ 0 & -\Delta^{-1} & 0 & 0 & 0 \\ 0 & 0 & \Delta^{-3/2} & 0 & 0 \\ 0 & 0 & 0 & -0.798 \cdot \Delta^{-1} & -\Delta^{-1} \\ 0 & 0 & 0 & 0.603 \cdot \Delta^{-1/2} & 0 \end{pmatrix},$$

with $\Delta = \frac{1}{52}$. Similarly, the Jacobian of the binding function that has been used to report the t-statistics and the consistency tests in Table 6 is based on the set of parameters of the auxiliary

model (29): to such a set of parameters is associated a binding function of the form $m = m(a)$, and the relationship between the Jacobians of b and m is

$$\frac{\partial b}{\partial a}(.) = \Delta_1 \frac{\partial m}{\partial a}(.). \tag{30}$$

Model (29) is the absolute-value model of Taylor (1986) and Schwert (1989) with normal errors, studied by Nelson and Foster (1994) and Fornari and Mele (1997a). Its main advantage over the more usual variance specifications is that it delivers estimates of volatility that are relatively more robust to the presence of possible outliers in the data. In this case, we also know that the invariant distribution of the residuals is approximately a generalized Student's-t when $\delta = \upsilon$ (theorem 3.3 p. 218 in Fornari and Mele (1997a)), which reduces to the celebrated Student's-t result of Nelson (1990) when $\delta = \upsilon = 2$.[18]

As mentioned in section 3, we consider normally distributed errors only (i.e. $\upsilon = 2$), since expanding into non-normality makes the resulting model non-stationary.[19] Hence, we are left with a specification in which $(\delta, \eta, \upsilon) = (1, 1, 2)$, and it is possible to show that in this case the invariant distribution of ϵ is more leptokurtic than the Student's-t obtained when $(\delta, \eta, \upsilon) = (2, 1, 2)$. Specifically, by applying theorem 3.5 p. 218 in Fornari and Mele (1997a), the invariant distribution of the residuals of (29) is given by

$$P(\overline{\epsilon}) = \frac{\left(\frac{2\omega}{\psi^2}\right)^{\frac{2\varphi+\psi^2}{\psi^2}}}{\sqrt{2\pi}\cdot\Gamma\left(\frac{2\varphi+\psi^2}{\psi^2}\right)} \int_0^\infty x^{-\frac{2\varphi+3\psi^2}{\psi^2}} \exp\left(-\frac{1}{2}\overline{\epsilon}^2 x^{-2} - \frac{2\omega}{\psi^2}x^{-1}\right) dx,\ \overline{\epsilon} \equiv \frac{\epsilon}{\sqrt{h}}, \tag{31}$$

as $h \downarrow 0$. Figure 3 compares the density in (31) with a normal density with variance equal to $(\omega/\varphi)^2$ where ω, φ and ψ have been fixed at the values shown in the second column of Table 6; its shape suggests that it should capture the usual stylized facts of the unpredictable parts of the vast majority of financial time series.

---

[18] To recall, υ is the tail-thickness parameter of the ged distribution (see section 3.1) and δ is the power to which σ is raised (see eq. (21)).

[19] Such a phenomenon is also noted by Andersen and Lund (1997a), who show that a specification based on EGARCH-type models is more stable when the errors of the model are not normal. Motivated by further empirical findings of Andersen and Lund (1997a), we also tried to include further lags in the volatility equation, but we did not observe any significant improvements.

Table 3 reports the qml estimates of model (29). The condition for covariance-stationarity of this model, reported in Theorem 3.1, i.e. $2 \cdot n_{1,2}\alpha + \beta = 0.798 \cdot \alpha + \beta < 1$ holds for the qml estimates reported in Table 3; the persistence of the the volatility propagating process is 0.993.

Table 4 presents summary statistics of the volatility filtered by the model (not yet "rescaled for diffusions"), and Figure 4 depicts its behavior in the sample. For reasons of comparisons, we also depict the first differences of r. The model appears to successfully capture some stylized features of the data, including the high volatility induced by the 'Monetary Experiment' of the early 80's. It is also worth noticing that perhaps due to such an isolated and yet relatively persistent episode, the long run volatility as implied by the parameter estimates attains the value of $\frac{w}{1-0.798 \cdot \alpha - \beta} = 15.458 \cdot 10^{-3}$, which is more than twice the average value of the filtered volatility for the whole sample. Because the estimated volatility wanders in a range of variation of about 0.026, however, such a difference is negligible: when we compute the ratio of the difference between the long run and average volatility to the range of variation, we find that it equals 0.321.

4.3 *Correction of the discretization bias, consistency tests, and filtering*

The second column of Table 5 reports the q-aml estimates of the continuous time model to which (20) converges based on (8)-(10). To correct their potential disaggregation bias we implement the indirect inference setup by simulating system (1) with the Euler-Maruyama approximation[20] (2) with $h^{-1} \equiv 1300$, which corresponds to generating 25 sub-intervals within a week. With an observations set of $N = 1135$, this implies that $h = N^{-l}$, with $l \simeq 1.0193 > \frac{1}{2}$, guarateeing that the conditions developed in Broze et al. (1998) to avoid simulation biases are fulfilled. We use $S = 50$ simulations. The estimation results are in the third column of Table 5. The correction made by indirect inference does not appear to matter: none of the q-aml estimates lies outside the usual 95 percent probability bands around the corresponding indirect inference estimates and, more importantly, when we formally checked the adequacy of the auxiliary model through the consistency test described in section 3, we found that the adjustment speed of the short-term rate is the only parameter that does not pass the test at the standard 95 percent level.

---

[20] Using (26) as simulation device does not alter our estimation results.

Such findings are of special interest here since Drost and Nijman (1993) showed that Arch models aggregate only when one weakens the very concept of Arch, introducing the so-called weak-Arch process; more importantly, Drost and Werker (1996) generalized the Drost and Nijman setting by introducing the so-called Arch diffusion which is, heuristically, the continuous time stochastic volatility process whose implied discrete differences form a weak-Arch process. A natural interpretation of our empirical findings is that even though the standard Arch models do not aggregate, they still remain, for a given frequency, an excellent approximation to the continuous time models toward which they converge in distribution, at least insofar as they are a natural proxy to the corresponding (discrete time) weak-Arch models. Naturally, these are issues that deserve a deep theoretical investigation that we leave for future research.

To check that the previous estimation results do not depend on the dimension of the simulation experiment ($S = 50$), we implement a sort of reverse exercise that consists in looking for the Arch model that one can expect to estimate if the true data generating mechanism happens to be (1). Specifically, we simulate (1) with parameters fixed at the indirect inference estimates of Table 5, sample the short-term rate at weekly frequency and estimate model (29) with such sampled data. We repeat the experiment 5000 times, removing the simulations for which there was not stationarity for the short-term rate and volatility (i.e.those where the persistence was greater than one). Notice that as a by-product of such an experiment, we will also get an assessment of the filtering performance of model (29).

Table 6 provides some basic statistics of the estimates and Figure 5 displays their relative frequencies. The distributions of the estimates are concentrated around the values of the estimates reported in Table 3: specifically, the standard 95 percent confidence bands of the Monte Carlo estimates are sufficiently tight to ensure statistical significance; yet they contain the figures corresponding to the true estimates reported in Table 3.

The filtering performance of the model is gauged in the following manner. Let $\sigma_{i,n}$ denote the volatility simulated at the $i-$th replication and sampled at n, and $\widehat{\sigma}_{i,n}$ is the corresponding (rescaled) Arch estimate. We are interested in evaluating the average filtering error in all the simulations, $\{\mathcal{E}_i\}_{i=1}^{5000}$, where $\mathcal{E}_i \equiv \frac{1}{1135}\sum_{n=1}^{1135}(\sigma_{i,n} - \widehat{\sigma}_{i,n})$. Figure 6 displays the Monte Carlo distribution of the average filtering error. It has an

average value of $9.610 \cdot 10^{-5}$ and a standard deviation of $3.275 \cdot 10^{-3}$ The RMSE, defined as $\sqrt{\frac{1}{5000}\sum_{i=1}^{5000}(\frac{1}{1135}\sum_{n=1}^{1135}(\sigma_{i,n}-\widehat{\sigma}_{i,n})^2)}$, is equal to 0.0209.

The last objective of this section consists in showing that even in the presence of a kind of misspecification which could be easily faced with in many applications involving the modeling of interest rates, the kind of models considered in this paper still remain a valid reference, at least insofar as one considers volatility filtering issues. Suppose, in other terms, that the data generating process (under the objective measure) is a three-factor model including the short-term rate, stochastic volatility, and a stochastic central tendency factor, this factor being a time varying conditional long-run mean of the short term rate. The question we want to answer to is: are the filtering results of this paper still valid when we attempt at extracting the (unobserved) stochastic volatility of such a data generating process? In addition to its obvious practical content, such a problem is directly related to previous theoretical work by Nelson (1992) and Nelson and Foster (1994). As mentioned in the Introduction, these authors produced many theoretical results based on more or less restrictive assumptions. The message of such results is that even in the presence of serious misspecification, Arch models still remain robust volatility filters. Now we wish to ascertain whether such results hold in an experiment in which Arch models are used to reconstruct the volatility dynamics of a three-factor data generating process.

To this end, we implement a Monte Carlo experiment in which we fit model (29) to 1,000 simulated trajectories of a three-factor model that extends in a natural way model (9) as

$$\left\{\begin{array}{lcl} dr(\tau) & = & \theta\left(l(\tau)-r(\tau)\right)d\tau+\sqrt{r(\tau)}\sigma(\tau)dW^{(1)}(\tau) \\ d\sigma(\tau) & = & \left(\omega-\varphi\sigma(\tau)\right)d\tau+\psi\sigma(\tau)dW^{(2)}(\tau) \\ dl(\tau) & = & \left(b_1-b_2 l(\tau)\right)d\tau+b_3\sqrt{l(\tau)}dW^{(3)} \end{array}\right. \tag{32}$$

where $W^{(i)}$, $i=1,2,3$, are standard Brownian motions, θ, ω, φ and ψ are fixed at the indirect inference estimates of Table 5, and b_i, $i=1,2,3$, are fixed at the values suggested by Andersen and Lund (1997*b*), i.e. $b_1 = 0.0078$, $b_2 = 0.1257$ and $b_3 = 0.0493$, and repeat the same expertiment as before. Table 7 provides the results. Even though model (29) is neglecting one of the factors of (32) (namely, the stochastic central tendency factor, $l(\tau)$), it exhibits remarkable volatility filtering properties. The Monte Carlo volatility filtering error is of the same order of magnitude as in the previous experiment and, again, the resulting dynamics

of simulated vis-à-vis filtered volatility trajectories display the same patterns as in figure 1. Considered as a (stochastic) volatility filter, model (29) would be hardly rejected as a valid tool of analysis, even in the presence of the neglected factor.

5. Conclusion

The intent of this paper was to explore to which extent Arch models can be practically used for the purpose of providing parameter estimates and volatility filtering in diffusion processes. Since the *standard* Arch models that have traditionally been used in the empirical literature do not approximate all diffusion models, we considered a reasonably wide class of models, named CEV-Arch, that converges toward any unrestricted CEV diffusion process as the sample frequency becomes larger and larger. While the searching strategy followed in this paper to the aim of approximating diffusions by means of Arch can be used to construct Arch sequences converging to yet more general diffusion processes, our central focus was the special case of volatility following a CEV-diffusion with linear drift.

Despite the fact that the CEV coefficient of volatility was left unrestricted, we provided empirical evidence supporting a model in which the (stochastic) volatility process of the short-term rate follows a diffusion process with *unit* elasticity of variance. In addition, we made use of simulation-based techniques to implement a global specification test for just-identified problems and provided evidence that (suitably rescaled) ARCH estimates of relevant parameters are statistically not distinguishable from estimates that one obtains with, say, indirect inference methods. Finally, the volatility filtering performances of the models are excellent. Even if one extracts the volatility from a three-factor model and uses a two-factor model only as a filter, the volatility filtering errors have the same order of magnitude as in the absence of misspecification. This finding suggests very simple and yet efficient tools to extract the (unobserved) volatility of a diffusion.

Tables and figures

Table 1

Monte Carlo study [a]

parameter	true	average	median	std.dev.
δ	1	1.0725	1.0206	0.1273
η	1	1.0849	1.0834	0.0961
volatility filtering error	NA	-1.1163·10^{-4} [b]	-2.2082·10^{-4}	4.5025·10^{-3}
δ	2	2.0047	1.9737	0.2474
η	$\frac{1}{2}$	0.6178	0.6132	0.1320
volatility filtering error	NA	-1.4995·10^{-3} [c]	-2.3333·10^{-4}	5.6091·10^{-2}

[a] The third column reports the average ML estimates of δ and η in (1) obtained by fitting an AR(1) model with volatility equation given by eq. (9) to 1000 simulated weekly sampled trajectories from the stochastic differential equation system (2). In these simulations, $\iota = 8{\cdot}10^{-3}$, $\theta = 0.11$, $\varphi = 0.38$ and δ and η are fixed at the values of the second column, with A) $\omega = 0.03$, $\psi = 0.8$ when $\delta = \eta = 1$, and B) $\omega = 2.36 \cdot 10^{-3}$, $\psi = 0.06$ when $\delta = 2$ and $\eta = \frac{1}{2}$. The fourth and fifth columns report the Monte Carlo median and standard deviation of such estimates. The case $\delta = \eta = 1$ corresponds to the actual estimates obtained in section 4. The Table also reports the Monte Carlo average (with the RMSE and the steady state expectation of σ in parentheses), median and standard deviation of the volatility filtering error.

[b] (RMSE: $1.8609 \cdot 10^{-2}$; $(\omega / \varphi = 7.895 \cdot 10^{-2})$;

[c] (RMSE: $1.6692 \cdot 10^{-2}$); $(\sqrt{(\omega - \psi^2/4)} / \varphi = 6.1985 \cdot 10^{-2})$.

Table 2

Summary statistics of r

mean	median	maximum	minimum	std. dev.	skewness	kurtosis
0.070	0.068	0.155	0.026	0.026	0.828	3.681

Autocorrelation function of r

lag	1	2	3	4	5	10	30	50
autocorrelation	0.995	0.998	0.979	0.971	0.961	0.914	0.789	0.696

The time series r is the short-term interest rate as defined in section 4 of the main text. It is a sample of 1,135 observation of the US 3-month T-Bill rate. It is observed between May 30, 1973 and February 22, 1995.

Table 3

QML estimates of (29)[a]

parameter	estimate	t-stat [b]
c_0	$1.555{\cdot}10^{-4}$	$\underset{(2.58)}{2.09}$
c_1	0.9979	$\underset{(9.86{\cdot}10^2)}{6.79{\cdot}10^2}$
w	$1.110{\cdot}10^{-4}$	$\underset{(3.41)}{4.36}$
α	0.1504	$\underset{(13.76)}{13.05}$
β	0.8728	$\underset{(139.18)}{93.97}$

[a] QML is the quasi-maximum likelihood estimation of the short rate dynamics. [b] Bollerslev-Wooldridge (1992) robust t-statistics in parentheses.

Table 4

Summary statistics of the conditional volatility σ as filtered by (29) [a]

mean	median	maximum	minimum	std. dev.	skewness	kurtosis
$7.102{\cdot}10^{-3}$	$5.483{\cdot}10^{-3}$	$2.805{\cdot}10^{-2}$	$2.042{\cdot}10^{-3}$	$4.306{\cdot}10^{-3}$	1.761	6.048

[a] Not rescaled for diffusion (see Appendix C).

Table 5

Parameter estimates [a]

parameter	q-aml	II	II t-stat	consistency tests
ι	0.0081	0.0082	3.04	-0.6727
θ	0.1067	0.1108	2.92	-2.0855
ω	0.0418	0.0301	2.98	-1.2177
φ	0.3736	0.3806	3.01	-0.1275
ψ	0.6540	0.8092	3.23	0.1390

[a] The second column reports the estimates of the parameters in (1) obtained with the moment conditions (18) and (20). The second column reports estimates obtained via the indirect inference (II) strategy explained in section 3, and the third column gives the corresponding t-statistics computed using the variance in (28) and (30) as the Jacobian of the binding function. The last column reports the ratio of each element of $\widehat{b}_N - \frac{1}{S}\sum_{s=1}^{S}\widehat{b}_{N,s}^{(h)}(\widehat{b}_N)$ to the corresponding standard error computed from the variance in (29) and using (30) as the Jacobian of the binding function.

Table 6

Monte Carlo study [a]

parameter	average	median	std. dev.
c_0	$1.640 \cdot 10^{-4}$	$1.610 \cdot 10^{-4}$	$3.340 \cdot 10^{-5}$
c_1	0.9974	0.9976	$1.764 \cdot 10^{-3}$
w	$1.210 \cdot 10^{-4}$	$1.130 \cdot 10^{-4}$	$4.420 \cdot 10^{-5}$
α	0.1548	0.1544	$2.405 \cdot 10^{-2}$
β	0.8665	0.8669	$2.056 \cdot 10^{-2}$

[a] The second column reports the average qml estimates of the parameters in model (29) obtained by fitting model (29) to 5000 simulated weekly sampled trajectories from the stochastic differential equation system (2). In these simulations, parameters are set to their II estimates reported in the third column of Table 5. The third and fourth columns report the Monte Carlo median and standard deviation of the simulated qml estimates.

Table 7

Monte Carlo study [a]

	average	median	std. dev.
volatility filtering error	$-3.6815 \cdot 10^{-5}$ [b]	$-6.0461 \cdot 10^{-5}$	$3.9666 \cdot 10^{-3}$

[a] The second column reports the average volatility filtering error defined in section 4 (with the RMSE and the steady state expectation of σ in parentheses) obtained by fitting model (29) to 1000 simulated weekly sampled trajectories of the following three-factor model:

$$\begin{cases} dr(\tau) &= \theta\left(l(\tau) - r(\tau)\right)d\tau + \sqrt{r(\tau)}\sigma(\tau)dW^{(1)}(\tau) \\ d\sigma(\tau) &= \left(\omega - \varphi\sigma(\tau)\right)d\tau + \psi\sigma(\tau)dW^{(2)}(\tau) \\ dl(\tau) &= \left(b_1 - b_2 l(\tau)\right)d\tau + b_3\sqrt{l(\tau)}dW^{(3)} \end{cases}$$

where $W^{(i)}, i = 1, 2, 3$, are standard Brownian motions, θ, ω, φ and ψ are fixed at the indirect inference estimates of Table 5, and $b_i, i = 1, 2, 3$, are fixed at the values suggested by Andersen and Lund (1997b), i.e. $b_1 = 0.0078, b_2 = 0.1257$ and $b_3 = 0.0493$. The third and fourth columns report the Monte Carlo median and standard deviation of the volatility filtering error.

[b] (RMSE: 0.0285); $(\omega\,/\varphi = 7.895 \cdot 10^{-2})$.

Figure 1

Filtered weekly volatility of $\sigma(t)$ in (1) by means of an Arch model

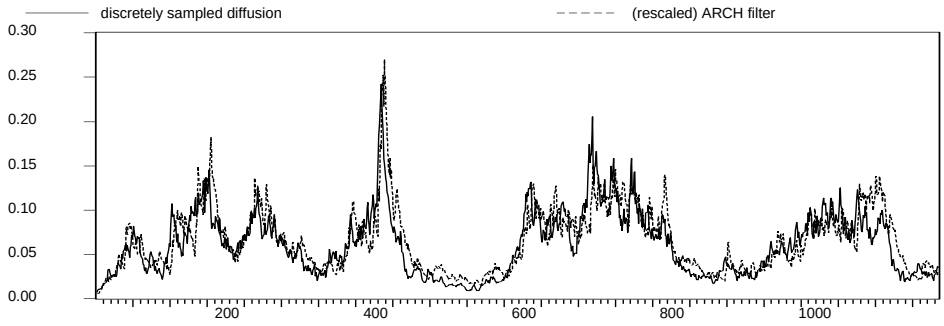
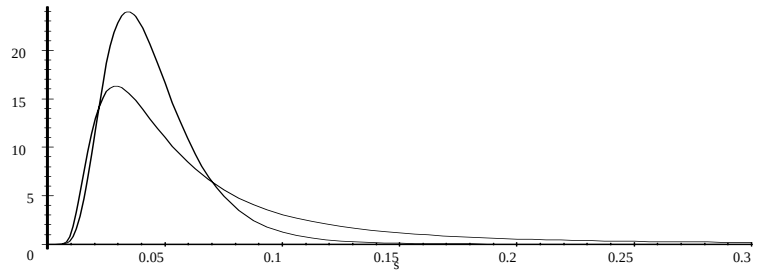
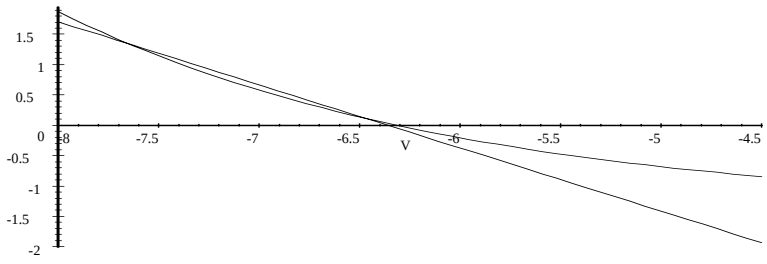


Figure 2

Panel A: Density comparison

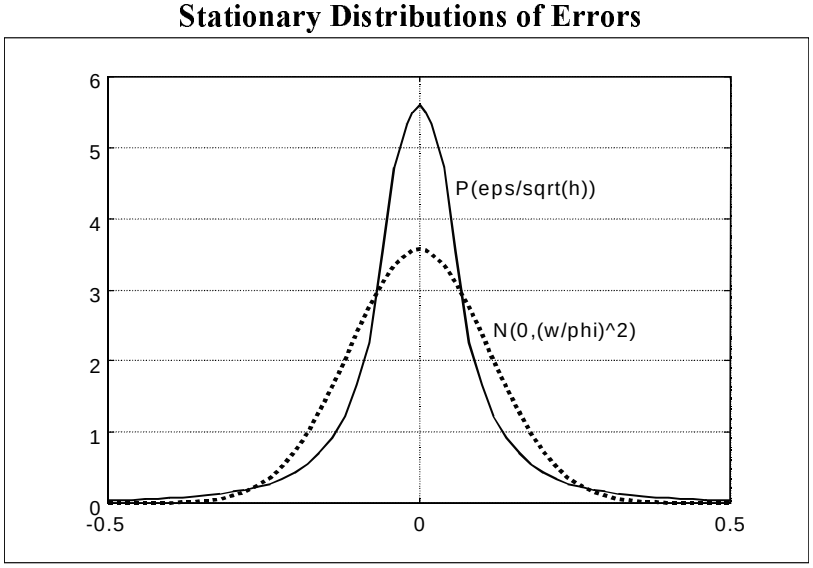


Panel B: Drift comparison



Panel A compares the log-normal density generated by the Andersen-Lund estimates (the curve with higher peak) and the density $f_1(\sigma)$ defined in (10) and generated by the estimates of section 4. Panel B compares the linear drift function generated by the Andersen-Lund estimates to the nonlinear drift function in (11) generated by the estimates of section 4.

Figure 3



In this Figure, $P(\frac{\epsilon}{\sqrt{h}})$ is the approximate invariant distribution of the errors in model (29) rescaled by $\sqrt{h}$ (see (22)). $N(0, (\frac{\omega}{\varphi})^2)$ is instead a normal density with standard deviation fixed at the steady state expectation of the volatility process (see model (1)).

Figure 4

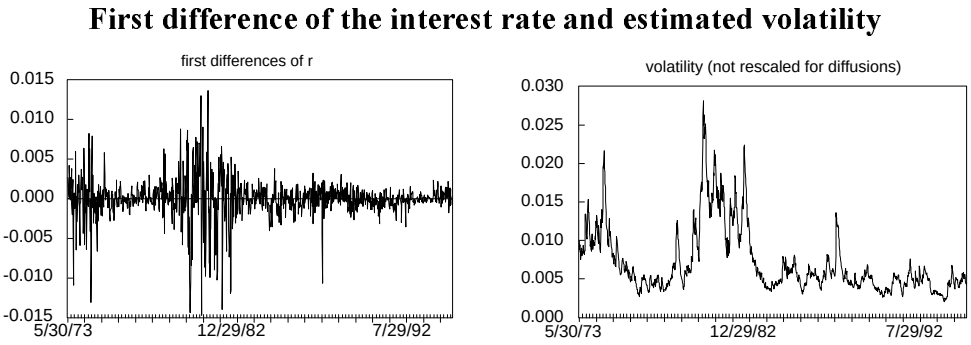


Figure 5

Monte Carlo densities of the ARCH parameters estimates

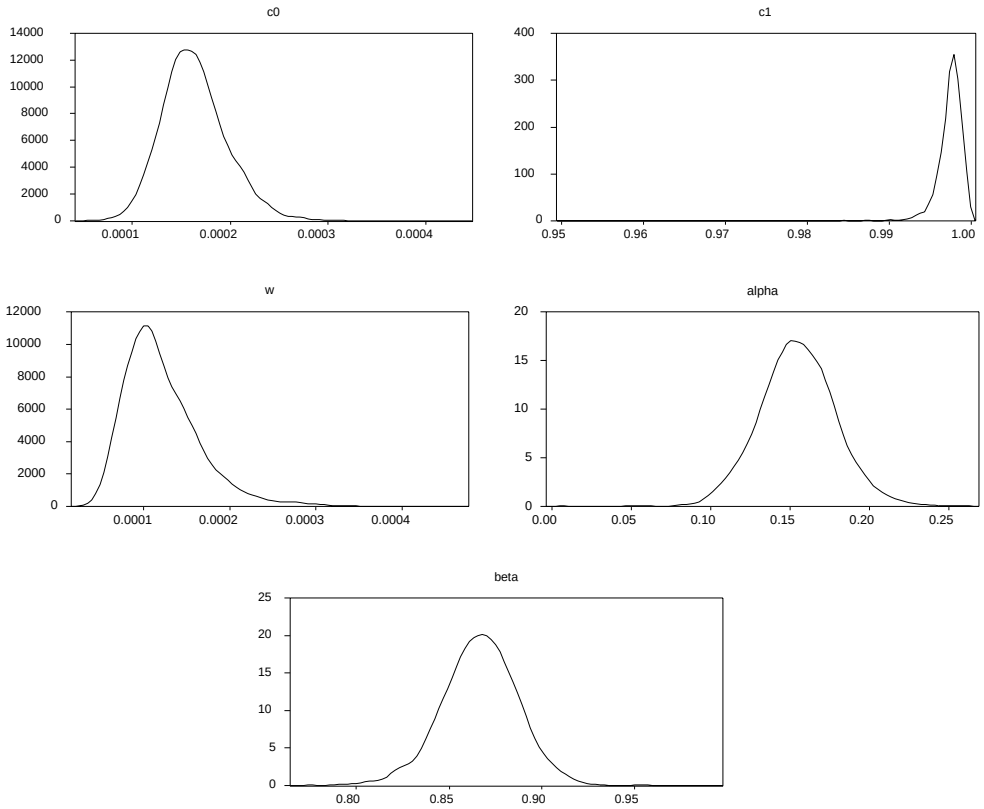
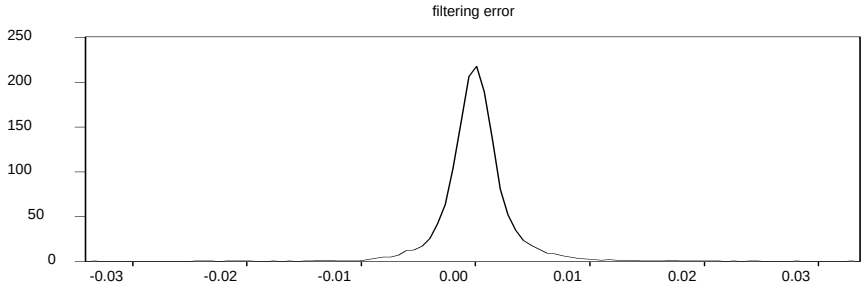


Figure 6

Monte Carlo filtering error



The filtering error of the conditional volatility is evaluated over 5,000 simulations as $\{\mathcal{E}_i\}_{i=1}^{5000}$, where $\mathcal{E}_i = \frac{1}{1135}\sum_{n=1}^{1135}(\sigma_{i,n} - \widehat{\sigma}_{i,n})$ with 1,135 being sample size, $\sigma_{i,n}$ and $\widehat{\sigma}_{i,n}$ the true and the predicted volatility. The Monte Carlo distribution of the average filtering error has an average of $9.61 \cdot 10^{-5}$ and a standard deviation of $3.275 \cdot 10^{-3}$.

Appendix A: Convergence results for section 3

Proof of theorem 3.1

Conditions (17) are sufficient to establish the weak convergence of the short-term rate and volatility processes toward the solutions of the following stochastic differential equations:

$$\begin{cases} dr(\tau) &= (\iota - \theta r(\tau))d\tau + \sigma(\tau)\sqrt{r(\tau)}dW^{(1)}(\tau) \\ d\sigma(\tau)^{\delta} &= (\omega - \varphi\sigma(\tau)^{\delta})d\tau + \psi\sigma(\tau)^{\delta}dW^{(3)}(\tau) \end{cases}$$

where $\{W^{(j)}(\tau)\}_{\tau\geq 0}$, $j = 1$ and 3, are $\mathcal{F}(\tau)$-Brownian motions. This has been shown in thm. 2.3 p. 209-211 of Fornari and Mele (1997a) in the case of a geometric Brownian motion, and the case of a square root process follows easily by an extension of another convergence result.

It remains to show that $W^{(3)}(\tau)$ can be written as:

$$W^{(3)}(\tau) = \rho W^{(1)}(\tau) + \sqrt{1-\rho^2}W^{(2)}(\tau),\ \tau \geq 0$$

with $\{W^{(2)}(\tau)\}_{\tau\geq 0}$ another $\mathcal{F}(\tau)$-Brownian motion. It is sufficient to show that the limit:

$$\lim_{h\downarrow 0} h^{-1}E\left\{\left({}_hr_{hk} - {}_hr_{h(k-1)}\right)\left({}_h\sigma^{\delta}_{h(k+1)} - {}_h\sigma^{\delta}_{hk}\right) \mid \mathcal{F}_{hk}\right\}$$

is not ill-behaved. After that, an identification argument will do the work.

By (18), and the fact that $\frac{{}_hu_{hk}}{\sqrt{h}}$ is g.e.d.$_{(v)}$ for each h,

$$\lim_{h\downarrow 0} h^{-1}E\left\{\left({}_hr_{hk} - {}_hr_{h(k-1)}\right)\left({}_h\sigma^{\delta}_{h(k+1)} - {}_h\sigma^{\delta}_{hk}\right) \mid \mathcal{F}_{hk}\right\}$$

$$= \lim_{h\downarrow 0} h^{-1}E\left\{\left(\iota_h - \theta_h \cdot {}_hr_{h(k-1)} + {}_h\sigma_{hk}\sqrt{{}_hr_{h(k-1)}} \cdot {}_hu_{hk}\right)\right.$$
$$\left.\times(w_h + \left(\alpha_h \left|{}_hu_{hk}\right|^{\delta}(1-\gamma s_k)^{\delta}h^{-\frac{\delta}{2}} + \beta_h - 1\right){}_h\sigma^{\delta}_{hk}) \mid \mathcal{F}_{hk}\right\}$$

$$= \lim_{h\downarrow 0} h^{-1} E\left\{ {}_h u_{hk} \left(\alpha_h \left| {}_h u_{hk} \right|^{\delta} (1-\gamma s_k)^{\delta} h^{-\frac{\delta}{2}} + \beta_h - 1 \right) \cdot {}_h\sigma_{hk}^{\delta+1} \mid \mathcal{F}_{hk} \right\} \cdot \sqrt{{}_h r_{h(k-1)}}$$

$$= \lim_{h\downarrow 0} h^{-1-\frac{\delta}{2}} \alpha_h \cdot E\left\{ {}_h u_{hk} \left| {}_h u_{hk} \right|^{\delta} (1-\gamma s_k)^{\delta} \cdot {}_h\sigma_{hk}^{\delta+1} \mid \mathcal{F}_{hk} \right\} \cdot \sqrt{{}_h r_{h(k-1)}}$$

$$= \lim_{h\downarrow 0} \frac{\alpha_h}{\sqrt{h}} \left\{ \left((1-\gamma)^{\delta} - (1+\gamma)^{\delta} \right) \int_{\mathbb{R}_+} x^{\delta+1} p(dx) \right\} \cdot {}_h\sigma_{hk}^{\delta+1} \cdot \sqrt{{}_h r_{h(k-1)}},$$

where $p(.)$ denotes the g.e.d.$_{(\upsilon)}$ density, or:

$$\lim_{h\downarrow 0} h^{-1} E\left\{ ({}_h r_{hk} - {}_h r_{h(k-1)})({}_h\sigma_{h(k+1)}^{\delta} - {}_h\sigma_{hk}^{\delta}) \mid \mathcal{F}_{hk} \right\}$$

$$= \lim_{h\downarrow 0} \frac{\alpha_h}{\sqrt{h}} \left\{ (1-\gamma)^{\delta} - (1+\gamma)^{\delta} \right\} K \cdot {}_h\sigma_{hk}^{\delta+1} \cdot \sqrt{{}_h r_{h(k-1)}};$$

here,

$$K = \frac{2^{\frac{\delta-\upsilon+1}{\upsilon}} \nabla_{\upsilon}^{\delta+1} \Gamma(\frac{\delta+2}{\upsilon})}{\Gamma(\upsilon^{-1})}.$$

By using (18),

$$\lim_{h\downarrow 0} \frac{\alpha_h}{\sqrt{h}} = \frac{\psi}{\sqrt{Z}},$$

where $Z \equiv (m_{\delta,\upsilon} - n_{\delta,\upsilon}^2)\left((1-\gamma)^{2\delta} + (1+\gamma)^{2\delta} \right) - 2n_{\delta,\upsilon}^2 (1-\gamma)^{\delta}(1+\gamma)^{\delta}$.

Hence,

$$\lim_{h\downarrow 0} h^{-1} E\left\{ \left({}_h r_{hk} - {}_h r_{h(k-1)} \right) \left({}_h\sigma_{h(k+1)}^{\delta} - {}_h\sigma_{hk}^{\delta} \right) \mid \mathcal{F}_{hk} \right\} = \frac{\psi \overline{K}}{\sqrt{Z}} \sigma^{\delta+1} \cdot \sqrt{r},$$

where:

$$\overline{K} = \left((1-\gamma)^{\delta} - (1+\gamma)^{\delta} \right) K.$$

To identify ρ, we note that this has to solve the following equation: $\psi\rho = \frac{\psi}{\sqrt{Z}}\overline{K}$, which yields:

$$\rho = \frac{\overline{K}}{\sqrt{Z}}.$$

The proof is complete.

Proof of theorem 3.2

Nearly identical to the proof of theorem 3.1.

Construction of alternate converging asymmetric models

It is well known that in correspondence with a given diffusion model, there may exist many well-behaved discrete time models converging in distribution to the given continuous time model. Hence, we can find other examples of discrete time ARCH-type models converging to model (1). As an example, consider the following model:

$$\sigma_{n+1}^{\delta} = w + \beta\sigma_n^{\delta} + \alpha(1-\gamma s_n)^{\delta\eta}\left(|u_n|^{\delta\eta} - E\left(|u_n|^{\delta\eta}\right)\right)\sigma_n^{\delta\eta},\ \gamma \in (-1,1). \qquad \text{(A1)}$$

The main difference between model (20) and model (A1) is the way how asymmetries in volatility are modeled. Suppose for instance that $\gamma > 0$ in model (A1). In this case, 'large' negative shocks introduce more volatility than positive shocks of the same size, while 'small' negative shocks introduce less volatility than positive shocks of the same size. Such a phenomenon, referred to as 'volatility reversal' in Fornari and Mele (1997b), seems to be pervasive in many stock markets and in this respect, model (A1) represents another example of the volatility-switching ARCH models that were originally introduced by Fornari and Mele (1997b).

Our objective now is to give a sketch of the proof that (A1) converges in distribution to (1) as the sampling frequency gets higher and higher. Consider the following approximating scheme:

$${}_h\sigma_{h(k+1)}^{\delta} - {}_h\sigma_{hk}^{\delta} = w_h - (1-\beta_h)\,{}_h\sigma_{hk}^{\delta} + \alpha_h(1-\gamma s_k)^{\delta\eta}\left\{|{}_hu_{hk}|^{\delta\eta} - E\left(|{}_hu_{hk}|^{\delta\eta}\right)\right\}h^{-\frac{\delta\eta}{2}}\,{}_h\sigma_{hk}^{\delta},$$

and introduce the following moment conditions:

$$\begin{array}{lcll} \lim_{h\downarrow 0} h^{-1} w_h & = & \omega \in (0,\infty), & \\ \lim_{h\downarrow 0} h^{-1}(1-\beta_h) & = & \varphi < \infty, & \text{(A2)} \\ \lim_{h\downarrow 0} h^{-1/2} \left\{ (1+\gamma)^{2\delta\eta} + (1-\gamma)^{2\delta\eta} \right\} (m_{\delta\eta,\upsilon} - 2n^2_{\delta\eta,\upsilon})\alpha_h & = & \psi < \infty. & \end{array}$$

For each h, we have that

$$E\left\{ (1-\gamma s_k)^{\delta\eta} \left(|{}_h u_{hk}|^{\delta\eta} - E\left(|{}_h u_{hk}|^{\delta\eta} \right) \right) h^{-\frac{\delta\eta}{2}} \mid \mathcal{F}_{hk} \right\} = 0,$$

and so the drift per unit of time is:

$$h^{-1} E\left({}_h\sigma^{\delta}_{h(k+1)} - {}_h\sigma^{\delta}_{hk} \mid \mathcal{F}_{hk} \right) = h^{-1} w_h - h^{-1}(1-\beta_h)\, {}_h\sigma^{\delta}_{hk}.$$

By taking limits for $h \downarrow 0$, and using the moment conditions (A2), we obtain the drift function of volatility in (1).

Now consider the second order moment per unit of time $h^{-1}E\{({}_h\sigma^{\delta}_{h(k+1)} - {}_h\sigma^{\delta}_{hk})^2 \mid \mathcal{F}_{hk}\}$. By taking limits for $h \downarrow 0$, and using again the moment conditions in (A2), yields after tedious computations:

$$\begin{aligned} & \lim_{h\downarrow 0} h^{-1} E\left\{ \left({}_h\sigma^{\delta}_{h(k+1)} - {}_h\sigma^{\delta}_{hk} \right)^2 \mid \mathcal{F}_{hk} \right\} \\ = & \lim_{h\downarrow 0} \left(\frac{\alpha_h}{\sqrt{h}} \right)^2 \left\{ E\left((1-\gamma s_k)^{2\delta\eta} \left| \frac{{}_h u_{hk}}{\sqrt{h}} \right|^{2\delta\eta} \right) + 4n^2_{\delta\eta,\upsilon} E\left((1-\gamma s_k)^{2\delta\eta} \right) \right. \\ & \qquad \left. -4n_{\delta\eta,\upsilon} E\left((1-\gamma s_k)^{2\delta\eta} \left| \frac{{}_h u_{hk}}{\sqrt{h}} \right|^{\delta\eta} \right) \right\} {}_h\sigma^{2\delta\eta}_{hk} \\ = & \lim_{h\downarrow 0} \left(\frac{\alpha_h}{\sqrt{h}} \right)^2 \left\{ (1+\gamma)^{2\delta\eta} + (1-\gamma)^{2\delta\eta} \right\} (m_{\delta\eta,\upsilon} - 2n^2_{\delta\eta,\upsilon})\, {}_h\sigma^{2\delta\eta}_{hk}, \end{aligned}$$

which gives the diffusion function of volatility in (1).

As regards correlation issues, the proof is very similar to that of theorem 3.1:

$$\lim_{h\downarrow 0} h^{-1} E\left\{ \left({}_h r_{hk} - {}_h r_{h(k-1)}\right) \left({}_h \sigma^{\delta}_{h(k+1)} - {}_h \sigma^{\delta}_{hk}\right) \mid \mathcal{F}_{hk} \right\}$$

$$= \lim_{h\downarrow 0} \frac{\alpha_h}{\sqrt{h}} \cdot E\left\{ \frac{{}_h u_{hk}}{\sqrt{h}} (1-\gamma s_k)^{\delta\eta} \left(\left| \frac{{}_h u_{hk}}{\sqrt{h}} \right|^{\delta\eta} - 2n_{\delta\eta,\upsilon} \right) \mid \mathcal{F}_{hk} \right\} \cdot {}_h \sigma^{\delta\eta+1}_{hk} \cdot \sqrt{{}_h r_{h(k-1)}},$$

where

$$\lim_{h\downarrow 0} E\left\{ \frac{{}_h u_{hk}}{\sqrt{h}} (1-\gamma s_k)^{\delta\eta} \left(\left| \frac{{}_h u_{hk}}{\sqrt{h}} \right|^{\delta\eta} - 2n_{\delta\eta,\upsilon} \right) \right\}$$

$$= E\left\{ \widetilde{u} \left(1-\gamma \cdot \text{sign}(\widetilde{u})\right)^{\delta\eta} \left(|\widetilde{u}|^{\delta\eta} - 2n_{\delta\eta,\upsilon} \right) \right\}$$

$$= \left\{ (1-\gamma)^{\delta\eta} - (1+\gamma)^{\delta\eta} \right\} \frac{2^{\frac{\delta\eta-\upsilon+1}{2}} \nabla_\upsilon^{\delta\eta+1}}{\Gamma(\frac{1}{\upsilon})^2} \left\{ \Gamma\left(\frac{\delta\eta+2}{\upsilon}\right) \Gamma\left(\frac{1}{\upsilon}\right) - \Gamma\left(\frac{\delta\eta+1}{\upsilon}\right) \Gamma\left(\frac{2}{\upsilon}\right) \right\},$$

and $\widetilde{u}$ is ged_υ.

Using an identification device as in the proof of theorem 3.1, we find that:

$$\rho = \frac{\left\{ (1-\gamma)^{\delta\eta} - (1+\gamma)^{\delta\eta} \right\} \frac{2^{\frac{\delta\eta-\upsilon+1}{2}} \nabla_\upsilon^{\delta\eta+1}}{\Gamma(\frac{1}{\upsilon})^2} \left\{ \Gamma\left(\frac{\delta\eta+2}{\upsilon}\right) \Gamma\left(\frac{1}{\upsilon}\right) - \Gamma\left(\frac{\delta\eta+1}{\upsilon}\right) \Gamma\left(\frac{2}{\upsilon}\right) \right\}}{\left\{ (1+\gamma)^{2\delta\eta} + (1-\gamma)^{2\delta\eta} \right\} \left(m_{\delta\eta,\upsilon} - 2n^2_{\delta\eta,\upsilon} \right)}.$$

In correspondence with reasonable values of δ, η and υ, the term $\{\Gamma\left(\frac{\delta\eta+2}{\upsilon}\right)\Gamma\left(\frac{1}{\upsilon}\right) - \Gamma\left(\frac{\delta\eta+1}{\upsilon}\right)\Gamma\left(\frac{2}{\upsilon}\right)\}$ is strictly positive, thus restricting sign(ρ) to be minus sign(γ), as in theorems. 3.1 and 3.2.

Appendix B: Standard regularity conditions and the convergence of the criterion

ASSUMPTION B1.

-plim$_N$ $\mathfrak{L}_N({}_\Delta r; b) = \mathfrak{L}_\infty(a_0; b)$, say, uniformly in $b \in B \subset \mathbb{R}^5$.
-plim$_N$ $\frac{\partial^2 \mathfrak{L}_N}{\partial b \partial b'}({}_\Delta r; b) = \ddot{\mathfrak{L}}_\infty\ (a_0; b)$, say, uniformly in $b \in B$. Further, $\ddot{\mathfrak{L}}_\infty\ (.)$ is invertible.
-$\left[\sqrt{N} \frac{\partial \mathfrak{L}_N}{\partial b}({}_\Delta r; b) \right]_{b=b_0(a_0)} \xrightarrow{d} N(0, J(a_0))$.

CONVERGENCE OF THE CRITERION (Sketch). We assume as in Broze et al. (1998) the continuity of the partial application $a \mapsto \widehat{b}_{N,s}^{(h)}(a)$, and for the case $S = 1$, we define $\widehat{b}_N^{(h)}(.) \equiv \widehat{b}_{N,1}^{(h)}(.)$ and ${}_{\Delta}\widetilde{r}(.) \equiv {}_{\Delta,h}\widetilde{r}^{(1)}(.)$. It is not hard to show that under conditions on $\mathfrak{L}_N({}_{\Delta}\widetilde{r}(a); b)$ that parallel those in assumption B1 stated above for the direct criterion $\mathfrak{L}_N({}_{\Delta}r; b)$, the simulated estimator is asymptotic normal:

$$\sqrt{N}\left(\widehat{b}_N^{(h)}(a) - b_0^{(h)}(a)\right) \xrightarrow{d} N\left(0, \ddot{\mathfrak{L}}_{\infty}^{(h)-1}\left(a; b_0^{(h)}(a)\right) \cdot J^{(h)}(a) \cdot \ddot{\mathfrak{L}}_{\infty}^{(h)-1}\left(a; b_0^{(h)}(a)\right)\right),$$

where $b_0^{(h)}(a) = \arg\max_b \mathfrak{L}_{\infty}^{(h)}(a; b)$, the limit simulation problem, and $\ddot{\mathfrak{L}}_{\infty}^{(h)}(.)$ and $J^{(h)}(.)$ are defined similarly as $\ddot{\mathfrak{L}}_{\infty}(.)$ and $J(.)$. Now, it follows from theorem 3.2 that the solution of (21): $\{{}_h r_{hk}, {}_h\sigma_{hk}^{\delta}\}_{k=0,1,\cdots} \Rightarrow \{r(\tau), \sigma(\tau)^{\delta}\}_{\tau \geq 0}$ (the solution of (1)). By this, an extension of a result cited in Fornari and Mele (2000*b*) (appendix B) that shows that the solution of (22) is unique, stationary and ergodic (for fixed h), and assuming the uniform continuity of the criterion $\mathfrak{L}_N(.; b)$, it follows that $\mathfrak{L}_N({}_{\Delta}\widetilde{r}(a_0); b) \Rightarrow \mathfrak{L}_N({}_{\Delta}r(a_0); b)$ as $h \downarrow 0$, and we suppose, as in Broze et al. (1998), that the convergence is uniform in b. Finally, because $\operatorname{plim}_N \mathfrak{L}_N({}_{\Delta}\widetilde{r}(a); b) = \mathfrak{L}_{\infty}^{(h)}(a; b)$ and $\operatorname{plim}_N \mathfrak{L}_N({}_{\Delta}r; b) = \mathfrak{L}_{\infty}(a_0; b)$, uniformly in $b \in B$ (both by assumption), one can easily verify that for small h, this implies $b_0^{(h)}(a_0) = \arg\max_b \mathfrak{L}_{\infty}^{(h)}(a_0; b) = \arg\max_b \mathfrak{L}_{\infty}(a_0; b) = b_0(a_0)$. This is:

$$\lim_{h \downarrow 0} b_0^{(h)}(a_0) = b_0(a_0),$$

while for fixed h, it is assumed that there exists only one solution to the system $b_0^{(h)}(a) = b_0(a_0)$: this has the form $\mathcal{A}^{(h)}(a_0)$, with $\lim_{h \downarrow 0} \mathcal{A}^{(h)}(a_0) = a_0$. Now by proposition 6 in Broze et al. (1998), one has that $\sqrt{N}\left({}_h\widehat{a}_N(a_0) - \mathcal{A}^{(h)}(a_0)\right) \xrightarrow{d} \mathrm{N}\left(0, \Sigma^{(h)}\right)$ (for fixed h), where $\Sigma^{(h)}$ is such that $\lim_{h \downarrow 0} \Sigma^{(h)} = 2V_0^{-1}\Gamma_0 V_0'$, and (28) follows for $S = 1$. In the preceding expressions, Γ_0 is defined as the limit of $\Gamma_0^{(h)}$ as $h \downarrow 0$, V_0 is defined similarly, and $\Gamma_0^{(h)}$ is the limit of $\ddot{\mathfrak{L}}_{\infty}^{(h)-1}\left(a; b_0^{(h)}(a)\right) \cdot J^{(h)}(a) \cdot \ddot{\mathfrak{L}}_{\infty}^{(h)-1}\left(a; b_0^{(h)}(a)\right)$ as $N \uparrow \infty$, whereas $V_0^{(h)}$ is the limit of $\left[\frac{\partial \widehat{b}_N^{(h)}}{\partial a}(a)\right]_{a=\mathcal{A}^{(h)}(a_0)}$ as $N \uparrow \infty$. The case $S > 1$ is similar.

Appendix C: How to rescale volatility for diffusions?

Here we provide details on how we rescaled ARCH-filtered volatility for diffusions. Let us rewrite the first equation of the Euler-Maruyama discrete approximation of (1) in (2) as:

$$r_n = \iota h + (1-\theta h) r_{n-1} + \sqrt{h}\sigma_{n-1}\sqrt{r_{n-1}}u_n,\ n = 1, \cdots, \widetilde{N}, \tag{C1}$$

where $\widetilde{N}$ denotes the total number of points generated by the simulations and u_n is $NID(0,1)$. Simulated data are sampled every ℓ points. Iterating (C1) leaves:

$$\begin{aligned} r_n &= \frac{\iota}{\theta}\left\{1-(1-\theta h)^{\ell}\right\} + (1-\theta h)^{\ell} r_{n-\ell} \\ &\quad + \sqrt{h}\left\{\sigma_{n-1}\sqrt{r_{n-1}}u_n + (1-\theta h)\,\sigma_{n-2}\sqrt{r_{n-2}}u_{n-1} + (1-\theta h)^2\,\sigma_{n-3}\sqrt{r_{n-3}}u_{n-2}\right. \\ &\quad \left. + \cdots + (1-\theta h)^{\ell-1}\,\sigma_{n-\ell}\sqrt{r_{n-\ell}}u_{n-(\ell-1)}\right\}. \end{aligned}$$

Because a diffusion is continuous with locally bounded paths, when h is low enough r and σ do not move too much within the unsampled ℓ subintervals. Let us denote with $\overline{r}^{\ell}_{n-1}$ and $\overline{\sigma}^{\ell}_{n-1}$ the (random) representative values of r and σ within the unsampled intervals that are such that the previous equation can be written approximately as:

$$\begin{aligned} r_n &= \frac{\iota}{\theta}\left\{1-(1-\theta h)^{\ell}\right\} + (1-\theta h)^{\ell} r_{n-\ell} \\ &\quad + \overline{\sigma}^{\ell}_{n-1}\cdot\sqrt{h}\left\{u_n + (1-\theta h)\,u_{n-1} + \cdots + (1-\theta h)^{\ell-1}\,u_{n-(\ell-1)}\right\}\sqrt{\overline{r}^{\ell}_{n-1}}. \end{aligned}$$

Our objective is to estimate each point of the sequence $\left\{\overline{\sigma}^{\ell}_{j}\right\}_{j=\ell,2\ell,\cdots,\widetilde{N}/\ell}$ in order to use it to filter the actual (discretely sampled) volatility path generated by the second equation of the Euler-Maruyama discrete approximation of (1) in (2): $\{\sigma_j\}_{j=1}^{\widetilde{N}/\ell} = \{\sigma(\ell\cdot j)\}_{j=1}^{\widetilde{N}/\ell}$.

Rewrite the previous equation as:

$$\begin{aligned} r_n &= \frac{\iota}{\theta}\left\{1-(1-\theta h)^{\ell}\right\}+(1-\theta h)^{\ell} r_{n-\ell} \\ &\quad +\overline{\sigma}_{n-1}^{\ell}\cdot\sqrt{h}\sqrt{1+(1-\theta h)^{2}+(1-\theta h)^{4}+\cdots+(1-\theta h)^{2(\ell-1)}}\cdot\sqrt{\overline{r}_{n-1}^{\ell}}\cdot\widetilde{u}_{n-1}^{\ell} \\ &= \frac{\iota}{\theta}\left\{1-(1-\theta h)^{\ell}\right\}+(1-\theta h)^{\ell} r_{n-\ell}+\overline{\sigma}_{n-1}^{\ell}\cdot\sqrt{\frac{h\left(1-(1-\theta h)^{2\ell}\right)}{1-(1-\theta h)^{2}}}\cdot\sqrt{\overline{r}_{n-1}^{\ell}}\cdot\widetilde{u}_{n}^{\ell}, \end{aligned}$$

where $\widetilde{u}^{\ell}$ is a standard Gaussian variate.

Now all the models we used in this paper actually deliver an estimate of

$$v_j \equiv \overline{\sigma}_j^{\ell}\cdot\sqrt{\frac{h\left(1-(1-\theta h)^{2\ell}\right)}{1-(1-\theta h)^{2}}},\ j=\ell,2\ell,\cdots,\widetilde{N}/\ell\,. \tag{C2}$$

Therefore, an estimate of each point of the sequence $\left\{\overline{\sigma}_j^{\ell}\right\}_{j=\ell,2\ell,\cdots,\widetilde{N}/\ell}$ is obtained by inverting formula (C2) to form the desired sequence:

$$\overline{\sigma}_j^{\ell}=\sqrt{\frac{1-(1-\theta h)^{2}}{h\left(1-(1-\theta h)^{2\ell}\right)}}\cdot v_j,\ j=\ell,2\ell,\cdots,\widetilde{N}/\ell\,. \tag{C3}$$

In this paper, we used:

$$h=\frac{1}{\Delta\cdot\ell},\ \Delta=52,\ \ell=25,$$

and the estimates of θ reported in Table 6 of the main text are such that $\sqrt{\frac{1-(1-\theta h)^{2}}{h\left(1-(1-\theta h)^{2\ell}\right)}}$ is always close to 7.218.

The filtered series of volatility reported throughout the paper are based on formula (C3) (see, however, below for numerical improvements of this formula). To relate the number found before to the correction given in formula (19) for the intercept of the volatility equation, note that:

$$\omega_{\text{q-aml}}=\Delta^{-1}\cdot\Delta^{-1/2}\cdot\widehat{w}_{\Delta}.$$

Here the correcting term is $\Delta^{-1/2} = 7.211$, which in practice is very close to the conversion factor given above.

In addition to being based on the stability of volatility within unsampled periods, the conversion formula (C3) is based on the assumption that the (small) changes of $\sigma\sqrt{r}$ are not autocorrelated. Relaxing such an assumption requires a much more complicated approach with continuous updatings. A reliable alternative consists in finding numerically a conversion formula similar to (C3). In this paper, we proceeded in the following way. We simulated 5000 times the continuous time system (1) in correspondence of the parameter estimates found in section 5. Then we defined:

$$\aleph \equiv \frac{1}{5000 \cdot \left(\widetilde{N}\ /\ell\right)} \sum_{i=1}^{5000} \sum_{j=1}^{\widetilde{N}/\ell} \frac{\sigma_i(\ell \cdot j)}{v_{ij}},$$

where $\sigma_i(\ell \cdot j)$ and v_{ij} are simulated volatility and filtered volatility as of time j obtained in the ith simulation. We found that $\aleph \approx 6.928$.

References

Aït-Sahalia, Y., 2000. Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach. Econometrica, forthcoming.

Altissimo, F., Fornari, F., Mele, A., 2001. Simulated Nonparametric Estimation of Continuous Time Models for Asset Prices and Returns. Mimeo, Bank of Italy and Queen Mary College.

Andersen, T.G., Lund, J., 1997a. Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate. Journal of Econometrics 77, 343–77.

Andersen, T.G., Lund, J., 1997b. Stochastic Volatility and Mean Drift in the Short Rate Diffusion: Sources of Steepness, Level, and Curvature in the Yield Curve. Mimeo, Northwestern University.

Ball, C.A., Roma, A., 1994. Stochastic Volatility Option Pricing. Journal of Financial and Quantitative Analysis 29, 589-607.

Bates, D. S., 1991. The Crash of '87: Was it Expected? The Evidence from Options Markets. Journal of Finance 46, 1009-44.

Black, F., 1976. Studies of Stock Price Volatility Changes. Proceedings of the 1976 Meeting of the American Statistical Association, 177-81.

Black, F., Scholes, M., 1973. The Pricing of Options and Corporate Liabilities. Journal of Political Economy 81, 637-59.

Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroskedasticity. Journal of Econometrics 31, 307-27.

Bollerslev, T., Rossi, P.E., 1996, Introduction, in: P.E. Rossi (ed.), Modeling Stock Market Volatility — Bridging the Gap to Continuous Time. San Diego, Academic Press, xi-xviii.

Bollerslev, T., Wooldridge, J.M., 1992. Quasi-maximum Likelihood Estimation of Dynamic Models with Time Varying Covariances. Econometric Reviews 11, 143-72.

Brandt, M., Santa-Clara, P., 2001. Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets. Journal of Financial Economics, forthcoming.

Broze, L., Scaillet, O., Zakoïan, J., 1995. Testing for Continuous Time Models of the Short Term Interest Rate. Journal of Empirical Finance 2, 199-223.

Broze, L., Scaillet, O., Zakoïan, J., 1998. Quasi-Indirect Inference for Diffusion Processes. Econometric Theory 14, 161-86.

Campbell, J. Y., Lo, A.W., MacKinlay A.C., 1997. The Econometrics of Financial Markets. Princeton, Princeton University Press.

Chapman, D.A., Long J.B., Pearson, N.D., 1999. Using Proxies for the Short Rate: When Are Three Months Like an Instant? Review of Financial Studies 12, 763-806.

Corradi, V., 2000. Reconsidering the Diffusion Limit of the GARCH(1,1) Process. Journal of Econometrics 96, 145-53.

Ding, Z., Granger, C., Engle, R.F., 1993. A Long Memory Property of Stock Market Returns and a New Model. Journal of Empirical Finance 1, 83-106.

Drost, F.C., Nijman, T.E., 1993. Temporal Aggregation of GARCH Processes. Econometrica 53, 385-407.

Drost, F.C., Werker, B.J.M., 1996. Closing the GARCH Gap: Continuous Time GARCH Modeling. Journal of Econometrics 74, 31-57.

Duan, J.C., 1997. Augmented GARCH(p, q) Process and its Diffusion Limit. Journal of Econometrics 79, 97-127.

Durham, G. B., 2001, Likelihood-Based Specification Analysis of Continuous-Time Models of the Short-Term Interest Rate. Mimeo, University of North Carolina.

Durham, G. B., Gallant, A.R., 2001. Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes. Mimeo, University of North Carolina.

Engle, R.F., Lee, G.G.J., 1996. Estimating Diffusion Models of Stochastic Volatility, in: Rossi, P.E. (ed.), Modeling Stock Market Volatility—Bridging the Gap to Continuous Time, Academic Press, San Diego, 333-55.

Fornari, F., Mele, A., 1997a. Weak Convergence and Distributional Assumptions for a General Class of Nonlinear Arch Models. Econometric Reviews 16, 205-29.

Fornari, F., Mele, A., 1997b. Sign- and Volatility- Switching ARCH Models: Theory and Applications to International Stock Markets. Journal of Applied Econometrics 12, 49-65.

Fornari, F., Mele, A., 2000. Stochastic Volatility in Financial Markets—Crossing the Bridge to Continuous Time. Kluwer Academic Publishers, Boston.

Fornari, F., Mele, A., 2001. Recovering the Probability Density Function of Asset Prices using GARCH Models as Diffusion Approximations. Journal of Empirical Finance 8, 83-110.

Gallant, A. R., Tauchen, G., 1996. Which Moments to Match? Econometric Theory 12, 657-81.

Gallant, A. R., Tauchen, G., 1998. Reprojecting Partially Observed Systems with Applications to Interest Rate Diffusions. Journal of American Statistical Association 93, 10-24.

Gouriéroux, C., Monfort, A., 1996. Simulation-Based Econometric Methods. Oxford, Oxford University Press.

Gouriéroux, C., Monfort, A., Renault, E:, 1993. Indirect Inference. Journal of Applied Econometrics 8, S85-S118.

Heston, S., 1993. A Closed Form Solution for Options with Stochastic Volatility with Application to Bond and Currency Options. Review of Financial Studies 6, 327-44.

Hull, J., White, A., 1987. The Pricing of Options on Assets with Stochastic Volatility. Journal of Finance 42, 281-300.

Johnson, H., Shanno, D., 1987. Option Pricing when the Variance is Changing. Journal of Financial and Quantitative Analysis 22, 143-51.

Lo, A.W., 1988. Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data. Econometric Theory 4, 231-47.

Longstaff, F., Schwartz, E., 1992. Interest-Rate Volatility and the Term Structure: A two Factor General Equilibrium Model. Journal of Finance 47, 1259-82.

Merton, R., 1973. Theory of Rational Option Pricing. Bell Journal of Economics and Management Science 4, 637-54.

Nelson, D.B., 1990. ARCH Models as Diffusion Approximation. Journal of Econometrics 45, 7-38.

Nelson, D.B., 1992. Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model. Journal of Econometrics 52, 61-90.

Nelson, D.B., 1996. Asymptotic Filtering Theory for Multivariate ARCH Models. Journal of Econometrics 71, 1-47.

Nelson, D.B., Foster, D.P., 1994. Asymptotic Filtering Theory for Univariate ARCH Models. Econometrica 62, 1-41.

Nicolau, J., 1999. Modeling the DEM/USD Exchange rate as a Stationary Continuous Time Process. Mimeo, Instituto Superior de Economia e Gestão.

Pedersen, A.R., 1995. A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. Scandinavian Journal of Statistics 22, 55-71.

Rossi, P.E., 1996. Modeling Stock Market Volatility—Bridging the Gap to Continuous Time. San Diego, Academic Press.

Schwartz, B.A., Nelson, D.B., Foster, D.P., 1993. Variance Filtering with ARCH Models: a Monte-Carlo Investigation. Mimeo, Tel Aviv University, Tel Aviv.

Schwert, W., 1989. Why Does Stock Market Volatility Change Over Time? Journal of Finance 44, 1115-54.

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

Stroock, D., Varadhan, S., 1979. Multidimensional Diffusion Processes. Berlin, Springer-Verlag.

Taylor, S., 1986. Modeling Financial Time Series. Chichester, UK, Wiley.

Taylor, S., 1994. Modeling Stochastic Volatility: a Review and Comparative Study. Mathematical Finance 4, 183-204.

Vasicek, O., 1977. An Equilibrium Characterization of the Term Structure. Journal of Financial Economics 5, 177-88.

Wiggins, J., 1987. Option Values Under Stochastic Volatility. Theory and Empirical Estimates. Journal of Financial Economics 19, 351-72.