Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER Journal of Economic Dynamics & Control 30 (2006) 931–966
JOURNAL OF Economic Dynamics & Control
www.elsevier.com/locate/jedc

Approximating volatility diffusions with CEV-ARCH models ☆

Fabio Fornari[a], Antonio Mele[b,*]

[a]*Bank for International Settlements, Basel, Switzerland*
[b]*London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK*

Received 13 August 2003; accepted 14 March 2005
Avalilable online 22 June 2005

Abstract

This article develops a new model of the ARCH class which allows volatility to react nonlinearly to past shocks as a function of the past volatility level. We show that this model approximates any CEV-diffusion model for stochastic volatility, and we judge its empirical performance as a diffusion approximation to models of the short-term rate with stochastic volatility and as a filter of the unobserved volatility. We show that the estimation of the continuous time scheme to which the discrete time ARCH model converges can be safely based on simple moment conditions linking the discrete time to the continuous time parameters. 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. A Monte-Carlo study reveals that the filtering performances of this model are remarkably good, even in the presence of serious misspecification.

JEL classification: C15; E43; G12

Keywords: Stochastic volatility; ARCH filtering; Indirect inference

☆A first version of this paper circulated as Bank of Italy Discussion Paper No. 397: "A Simple Approach to the Estimation of Continuous Time C.E.V. Stochastic Volatility Models of the Short-Term Rate" (February 2001).

*Corresponding author. Tel.: +44 0 20 7955 6203.
E-mail address: a.mele@lse.ac.uk (A. Mele).

0165-1889/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jedc.2005.03.008

1. Introduction

The estimation of continuous time models has recently received increasing attention by both financial economists and macroeconomists. From methods extending the original Hansen's (1982) GMM approach (see, for example, Duffie and Singleton, 1993; Gouriéroux et al., 1993; Gallant and Tauchen, 1996), this literature has evolved towards approximating the maximum likelihood (ML) estimator. Aït-Sahalia (2002, 2003) developed a closed-form approximation to the ML estimator for scalar and multivariate observable diffusions whereby the unknown transitional density of the model can be approximated, in closed-form, with great accuracy. Brandt and Santa-Clara (2002) and Durham and Gallant (2002), 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, to recover the unknown transition densities. Empirical results on short-term rate models are concordant, independently of the methodologies employed: traditional univariate diffusions perform poorly relative to bivariate continuous time models in which the interest rate dynamics is coupled with its conditional volatility dynamics.[1] These empirical findings are not unexpected. They simply represent the continuous time counterpart of the universal finding of ARCH-type effects in time series of financial price changes.[2] Their theoretical underpinnings may be found in the initial contribution of Nelson (1990), who showed that the basic ARCH models are reasonable approximations to the typical diffusion processes used by theorists in financial economics.[3] For alternative (non-parametric) approaches to volatility measurements, see Andersen et al. (2002).

In this paper, we study the effectiveness of ARCH-type models as auxiliary devices in a variety of interesting experiments of statistical inference. We pursue two fundamental objectives. We study whether ARCH-type models can be safely used to obtain (i) approximations to diffusions *as well as* (ii) filters in continuous time models with unobservable state variables. Our agenda is to take as given a continuous time data-generating process; and then test whether ARCH-type models can (i) approximate and/or support the estimation of its parameters and (ii) recover the dynamics of the unobservable volatility. This is a research topic started by Nelson (1990) and subsequently, for no apparent reason, abandoned. As an example of the lack of results in this field, Campbell et al. (1997) emphasized in their authoritative textbook that 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" (p. 381). This conjecture was

[1]See, e.g., Andersen and Lund (1997a,b) for results obtained through simulated methods of moments and Durham (2003) for results obtained through simulated ML methods.

[2]The unanimous finding of Arch-type effects in financial data has led researchers (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.

[3]The major contribution of Nelson to this literature can be found in Part II of the book edited by Rossi (1996).

indeed natural because ARCH models are relatively simple to implement by users not trained in the nuts and bolts of continuous time financial econometrics. And even for those of us who are, filtering continuous time volatility of a time series from a discrete time realization does represent a formidable challenge. Yet the literature started by Nelson (1990) *theoretically* demonstrated that ARCH models are natural devices to perform this task. Should ARCH models really exhibit good approximating properties *in practice*, they could be of substantial help to investment bankers or risk regulators concerned with estimation and volatility filtering of continuous time models. In this article, we investigate such empirical properties of ARCH models and show that they are very encouraging in practice.

Our second contribution is that we study these properties within a new class of ARCH models. This new class steps into the largely unexplored issue of how volatility reacts to shocks as a function of the volatility level. We assume that this reaction takes place in a non-linear way and we call the resulting model constant elasticity of variance (CEV)-ARCH because we show that it approximates any diffusion model in which volatility follows a CEV process. Given the general empirical success of CEV models for asset prices in finance, our contribution can thus be viewed as a natural attempt to introduce CEV features into the empirical modeling of volatility. We apply these ideas to continuous time models of the short-term rate with stochastic volatility. This choice is in part motivated by an attempt to bridge a gap with theoretical (yield curve) models of the short-term rate volatility. For example, a popular class of yield curve models is the affine one (see Duffie and Kan, 1996), in which the variance of volatility is an affine function of the volatility level. Such a property is not shared by any available ARCH model, but it may emerge as a special case of our general CEV-ARCH specification of the volatility process. Finally, the issue of how interest rate volatility changes with the level of the short-term rate has received considerable attention in the literature since the seminal contribution of Chan et al. (1992). However, there are no analyses investigating the shape of the variance of volatility. Our model paves the way to investigating this issue in a relatively simple way.

On a more detailed level, our contribution in this paper is to show (i) that our CEV-ARCH scheme converges to a continuous time model as the sampling interval shrinks to zero; (ii) that one can safely use the likelihood function of the CEV-ARCH in lieu of the true likelihood function (which cannot be computed analytically); (iii) that our approximating model makes no significant error in recovering the true model's volatility – even in the presence of important misspecifications; and finally, (iv) that in an empirical application, our discrete time ARCH approximation does capture the salient statistical features predicted by the corresponding continuous time model. On a strictly theoretical standpoint, the main difficulty we face lies in the previous items (ii) and (iv). Indeed, we provide closed-form moment conditions linking the discrete time to the continuous time parameters. These conditions do guarantee the weak convergence of the CEV-ARCH model towards the continuous time reference model. Yet ARCH models (and our CEV-ARCH model) are typically not closed under temporal aggregation (Drost and

Nijman, 1993; Drost and Werker, 1996). Therefore, we need to control for (and correct) potential "disaggregation" biases. To implement such an additional step, we use ARCH models as *auxiliary* devices in simulation-based (indirect inference) schemes.[4] To anticipate the results, we find that when applied to real data, the correction made by indirect inference is not statistically significant – a result we obtain through a global specification test for just-identified models originally suggested by Gouriéroux et al. (1993).[5]

As we mentioned in the previous item (iii), not only do we care about the ability of CEV-ARCH models in correctly estimating the parameters of its continuous time limit; we also want to make sure that the model constitutes a good volatility filtering device. This is an important step of our analysis. Suppose for example that the short-term rate is generated by a continuous time two-factor model (such as Eq. (1) below). Standard evaluation models would then predict that in the absence of arbitrage opportunities, the rational bond price function depends on both the short-term rate *and* its instantaneous volatility. As a result, volatility becomes an essential ingredient in the practical implementation of a term structure model in this context. We already know, from a theoretical standpoint, that appropriate sequences of ARCH models are able to give a consistent estimate of 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). In this paper, we provide evidence that the desired 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).

The article is organized as follows. Section 2 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 Appendix A and B.

[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 diffusion approximation properties of Arch models in a financial 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 asymptotically achieves the same efficiency as the true (intractable) ML estimator when the auxiliary model generates a density that 'smoothly embeds' the true likelihood function of the discretely sampled diffusion. Our estimation strategy has the aim of ascertaining whether our auxiliary model is a reasonable approximation to the continuous time model.

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

2. The reference continuous time model and the CEV-ARCH process

2.1. The CEV-ARCH model

We take as given a data-generating process in which the short-term-rate volatility is solution to a CEV process

$$\begin{aligned} \mathrm{d}r(\tau) &= (\iota - \theta r(\tau))\,\mathrm{d}\tau + \sigma(\tau)\sqrt{r(\tau)}\,\mathrm{d}W^{(1)}(\tau), \\ \mathrm{d}\sigma(\tau)^{\delta} &= (\omega - \varphi\sigma(\tau)^{\delta})\,\mathrm{d}\tau + \psi\sigma(\tau)^{\delta\cdot\eta}\,\mathrm{d}(\rho W^{(1)}(\tau) + \sqrt{1-\rho^2}\,W^{(2)}(\tau)), \end{aligned} \tag{1}$$

where $a = (\iota, \theta, \delta, \omega, \varphi, \psi, \eta, \rho)$ is a vector of parameters, and $W^{(i)}$, $i = 1, 2$, are standard Brownian motions. The $\sqrt{r}$-term included in the diffusion term of the short rate equation captures an empirical regularity known as the 'level effect', i.e., *ceteris 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 $\sigma|r|^{d}$ $(d \geqslant 0.5)$ is possible, though it would not change dramatically our empirical results.[7]

The objective of this paper is to use ARCH-type models to implement (i) estimation of the above continuous time parameters and (ii) extraction of the unobserved short-term rate volatility process $\sigma(\cdot)$.

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

$$\begin{aligned} {}_{h}r_{h(k+1)} - {}_{h}r_{hk} &= (\iota - \theta \cdot {}_{h}r_{hk})h + {}_{h}\sigma_{hk}\sqrt{{}_{h}r_{hk}} \cdot {}_{h}u_{h(k+1)}, \\ {}_{h}\sigma^{\delta}_{h(k+1)} - {}_{h}\sigma^{\delta}_{hk} &= (\omega - \varphi \cdot {}_{h}\sigma^{\delta}_{hk})h + \psi \cdot {}_{h}\sigma^{\delta\cdot\eta}_{hk}\sqrt{h}\,{}_{h}\widetilde{\xi}_{h(k+1)}, \end{aligned} \tag{2}$$

where h denotes the discretization step,

$$\begin{pmatrix} {}_{h}u_{hk} \\ {}_{h}\widetilde{\xi}_{hk} \end{pmatrix} \sim NID\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} h & \sqrt{h}\rho \\ \sqrt{h}\rho & 1 \end{pmatrix} \right)$$

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

It is well known that as $h \downarrow 0$, (2) converges weakly (or in distribution) to (1).[8] Hence, the higher the sampling frequency, the higher should be the accuracy of, say, ML estimates of a obtained with (2). Unfortunately, (2) represents a discrete time stochastic variance model for which standard ML methods are practically

[7] If $d = 1$, then $x(\tau) \equiv \mathrm{e}^{\theta\tau} r(\tau) - \iota/\theta(\mathrm{e}^{\theta\tau} - 1)$ admits the representation $x(\tau) = x(0) + \int_0^{\tau} f(u, x(u))\,\mathrm{d}Z(u)$, where $Z(\tau) \equiv \int_0^{\tau} \sigma(s)\,\mathrm{d}W^{(1)}(s)$ and $f(t, x) \equiv x + \iota/\theta(\mathrm{e}^{\theta t} - 1)$. Under the parameter restrictions developed by Aït-Sahalia (1996, Appendix), σ^{δ} in (1) has a strong solution. By Theorem 2.1, p. 375 in Revuz and Yor (1999), there exists then a strong solution X which is $F^{Z}(\tau)$-adapted where $F^{Z}(\tau) \equiv \sigma(Z(u) : u \leqslant \tau)$. Since Z is a martingale, the case $d = \frac{1}{2}$ is treated by modifying the proof in Revuz and Yor (pp. 376–377) through the usual Yamada–Watanabe device as described for example in Karatzas and Shreve (1991) (pp. 291-292).

[8] If (1) has a unique strong solution denoted as $\{r(\tau), \sigma(\tau)^{\delta}\}_{\tau \geqslant 0}$, *weak convergence* of $\{{}_{h}r_{hk}, {}_{h}\sigma^{\delta}_{hk}\}_{k=1,2,\ldots}$ in (2) to $\{r(\tau), \sigma(\tau)^{\delta}\}_{\tau \geqslant 0}$ means that the finite-dimensional distributions of $\{{}_{h}r_{hk}, {}_{h}\sigma^{\delta}_{hk}\}_{k=1,2,\ldots}$ converge to those of $\{r(\tau), \sigma(\tau)^{\delta}\}_{\tau \geqslant 0}$ 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).

unfeasible. Furthermore, there are no simple techniques to filter the actual volatility path out of (1).

Our approach in this paper is very simple. We use ARCH models as diffusion approximations. To gain some intuition of how the approximation works, consider the standard Garch(1,1) of Bollerslev (1986)

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

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

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

and suppose that $u \sim \mathrm{N}(0, 1)$. Discretize time so as to make $n : hk \leqslant n \leqslant h(k+1)$, $k = 1, 2, \ldots$ and let the parameters w, β, α vary with h by introducing sequences w_h, β_h, α_h. Next let $h \downarrow 0$; the resulting volatility process converges in distribution to:[9]

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

where

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

Eq. (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:

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

also converges in distribution to the following diffusion limit:

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

Eq. (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. This property

[9]To obtain some intuition for this result, notice that the sequence $(\xi_n)_{n=1}^{\infty} \equiv (u_n^2 - \mathrm{E}(u^2))_{n=1}^{\infty}$ 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 $\mathrm{d}W^{(2)}(\cdot)$. Furthermore, the re-normalizing $\sqrt{2}$-term in the last equation of (5) is explained by the fact that $\xi = u^2 - \mathrm{E}(u^2) = u^2 - 1$ 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.

may significantly restrict the use of ARCH models in practice. In the term structure literature, for example, many modelers are interested in the so-called affine class of models (Duffie and Kan, 1996), in which the variance of the variance is proportional to the variance level,

$$\mathrm{d}\sigma(\tau)^2 = (\omega - \varphi\sigma(\tau)^2)\,\mathrm{d}\tau + \psi\sigma(\tau)\,\mathrm{d}W^{(2)}(\tau).$$

None of the previous ARCH models converges in distribution to such diffusion limit. We now introduce an ARCH scheme that does not force the elasticity of variance to one.[10] Consider, for instance, the following model:

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

which can also be written as

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

This model also 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 distribution to

$$\mathrm{d}\sigma(\tau)^2 = (\omega - \varphi\sigma(\tau)^2)\,\mathrm{d}\tau + \psi\sigma(\tau)^{2\eta}\,\mathrm{d}W^{(2)}(\tau).$$

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

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

As before, we will show that as the sampling frequency gets higher and higher, the volatility process in (8) converges in distribution to

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

which corresponds 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).[11]

[10] 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).

[11] In the same vein, one may introduce non-linear volatility dynamics into discrete time models that match any desidered feature of the resulting diffusion limit. Consider, for instance, the following model:

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

Using the methods of Section 3, it can then be shown that this model converges in distribution toward the solution to

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

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).

2.2. Filtering and invariance properties of the CEV-ARCH: Preliminary Monte-Carlo evidence

The practical relevance of the previous approximation schemes can be grasped very simply. Fig. 1 depicts the typical filtering behavior of an ARCH model applied to a simplified version of (1). The unbroken line is one weekly sampled trajectory of the volatility, $\sigma(\tau)$, simulated within the following model:

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

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). Naturally, for each simulated estimation of the ARCH model, we considered ourselves constrained to *only* knowing the realization of the simulated $r(\tau)$. Fig. 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. More precisely, when we compare the volatility trajectories filtered with Eq. (8) (with parameters set at their estimates we obtain in our empirical section (see Section 4, Table 5)) to those directly simulated from (1), we find that their patterns are very similar and similar also to those of Fig. 1. Table 1 reports precise results assessing the performance of this volatility filtering based on (8). 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. Note also that to compare the simulated volatility to the filtered volatility, the latter has to be

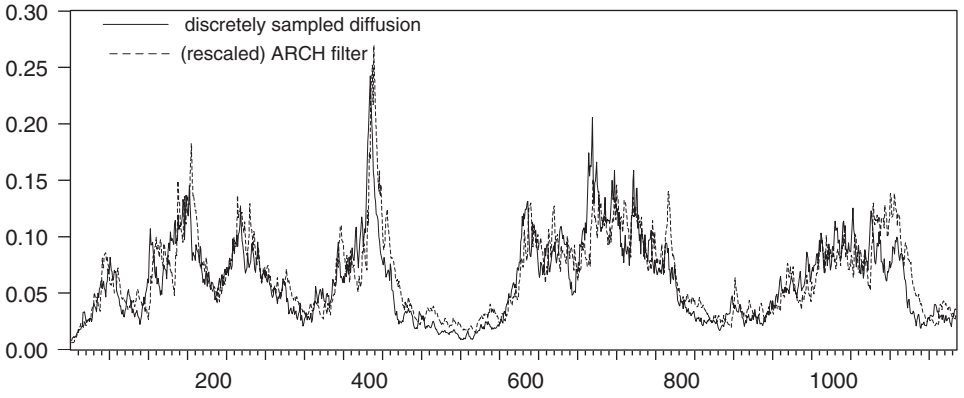

Fig. 1. Filtered weekly volatility of $\sigma(\tau)$ in (1) by means of an ARCH model.

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} RMSE: 1.8609×10^{-2} $\omega/\varphi = 7.895 \times 10^{-2}$	-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} RMSE: 1.6692×10^{-2} $\sqrt{\frac{1}{\varphi}(\omega - \frac{1}{4}\psi^2)} = 6.1985 \times 10^{-2}$	-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. (8) to 1000 simulated weekly sampled trajectories from system (1). In these simulations, $\rho = 0$, $\iota = 8 \times 10^{-3}$, $\theta = 0.11$, $\varphi = 0.38$ 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) $\omega = 2.36 \times 10^{-3}$, $\psi = 0.06$ 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 Monte-Carlo average (with the RMSE and the steady-state expectation of σ in parentheses), median and standard deviation of the volatility filtering error is also reported.

"rescaled for diffusions"; techniques for dealing with this issue are introduced and explained in Appendix B.

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. An important simplification is obtained by fixing δ and η at their ML estimates obtained after fitting model (8). To ascertain whether such a 'time-scale invariance' is sensible in practice, we implemented a Monte-Carlo experiment. We considered model (1) and fixed $\rho = 0$ (consistently with subsequent empirical evidence reported in Section 4) and the other coefficients at the values in Table 1. We then simulated (1) 1000 times with 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 1135 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

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

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 estimating (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.

2.3. 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 we already mentioned, 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}, $\delta \geqslant 1$, has a steady-state distribution that is an inverted Gamma with mean $\overline{\omega}/\varphi$ (e.g., Lemma 3.1, p. 217 in Fornari and Mele (1997a)); the stationary distribution of σ is consequently given by

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

where $\Gamma(\cdot)$ is the Gamma function (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, Fornari and Mele (2000) 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 $\omega \equiv 0$. Indeed, $\mathscr{V} \equiv \log \sigma^2$ is solution to

$$\begin{aligned} \mathrm{d}\mathscr{V}(\tau) = &\left(-\frac{2\varphi+\psi^2}{\delta} + 2\frac{\omega}{\delta}\exp\left(-\frac{\delta}{2}\mathscr{V}(\tau)\right)\right)\mathrm{d}\tau \\ &+ \frac{2\psi}{\delta}\,\mathrm{d}(\rho W^{(1)} + \sqrt{1-\rho^2}\,W^{(2)}). \end{aligned} \tag{11}$$

[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.

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 the stochastic volatility option pricing domain and in Andersen and Lund (1997a) or Gallant and Tauchen (1998) in the interest rate framework. To see this in more detail, consider the linear mean reverting model utilized in Andersen and Lund,

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

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 to

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

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.

Fig. 2 (panel A) depicts a comparison between the stationary densities that are associated to (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

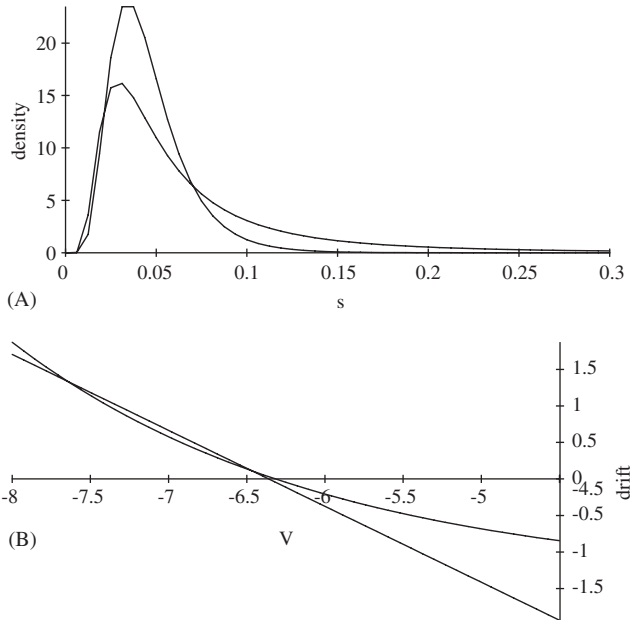

Fig. 2. Panel A (top): Density comparison and Panel B (Bottom): Drift comparison.

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 Fig. 2 (panel B), the two drift functions have the same order of magnitude whenever volatility is low. But as soon as volatility visits higher regions, 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 we mentioned in the Introduction, various methods have been recently proposed to estimate the parameters of a discretely sampled diffusion. In this situation, ML methods are in practice unfeasible because the likelihood function implied by the measure induced by a discretely sampled diffusion is typically unavailable in explicit form. In this paper, we consider a simple alternative and make use of a (tractable) exact likelihood function of a class of approximating models. The main idea is to resort to a suitably chosen class of ARCH models converging in distribution to the solution to (1) as the sampling frequency becomes infinite – as we heuristically illustrated in the previous section. Because the resulting likelihood function is associated with a model converging in distribution to the solution to (1) 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 effort. 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.[13] To quantify such potential drawbacks of the procedure, we show how to construct a 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), which can be viewed as the natural substitute of a global specification test in just-identified problems.

[13]On a strictly theoretical standpoint, such a correspondence may exist as soon as one weakens the concept of an Arch model (Drost and Nijman, 1993; Drost and Werker, 1996). Moreover, Corradi (2000) criticized the conditions Nelson (1990) imposed to show the convergence of the basic Garch(1,1) to a diffusion (see our footnote 15 for details on how we adapt Corradi's critique to our setup).

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 to (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):

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

where the indexing $n = 0, 1, \ldots$ 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 $w^{(\Delta)} > 0, \alpha_{\Delta}, \beta_{\Delta} \geqslant 0, \gamma \in (-1, 1), \delta > 0$. Finally, γ allows for the 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 to (1), discretize time as $hk \leqslant n \leqslant h(k+1)$

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

(with $s_k = \mathrm{sign}({}_{h}u_{hk})$ 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 our empirical results suggest a simplification of (1) in which $\delta = 1$ and $\rho = 0$ (see Section 4), here we provide more general results potentially useful to deal with 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 zero in (1).[14]

[14]Moreover, we slightly complicate the theoretical analysis and allow standardized residuals to be general error distributed. This complication is not considered in the empirical section of the paper.

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, 2000, 2001) (see also Duan (1997) for related work). We only introduce notation for the filtration generated by $\{{}_hr_{h(i-1)}, {}_h\sigma^{\delta}_{hi}\}_{i=1}^{k}$, which is $\mathscr{F}_{hk}$, 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

$$\frac{\upsilon \exp(-\frac{1}{2}\nabla_{\upsilon}^{-\upsilon}|x|^{\upsilon})}{2^{1+\upsilon^{-1}}\nabla_{\upsilon}\Gamma(\upsilon^{-1})}, \quad \text{where } \nabla_{\upsilon}^{2} \equiv \frac{\Gamma(\upsilon^{-1})}{2^{2/\upsilon}\Gamma(3\upsilon^{-1})} \ (\upsilon > 0).$$

The following convergence result is an extension of Theorem 2.3 p. 211 in Fornari and Mele (1997a). It allows for the presence of the instantaneous correlation between $\{{}_hr_{hk}\}_{k=0,1,\ldots}$ and $\{{}_h\sigma^{\delta}_{hk}\}_{k=0,1,\ldots}$ as h shrinks to zero

Theorem 3.1. *Let*

$$m_{\delta,\upsilon} = \frac{2^{(2\delta/\upsilon)-1}\nabla_{\upsilon}^{2\delta}\Gamma((2\delta+1)/\upsilon)}{\Gamma(\upsilon^{-1})}, \quad n_{\delta,\upsilon} = \frac{2^{(\delta/\upsilon)-1}\nabla_{\upsilon}^{\delta}\Gamma((\delta+1)/\upsilon)}{\Gamma(\upsilon^{-1})},$$

and let ${}_hu_{hk}/\sqrt{h}$ *be general error distributed. Let*

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

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

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

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

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

Then, $\{{}_hr_{h(k-1)}, {}_h\sigma^{\delta}_{hk}\}_{k=0,1,\ldots} \Rightarrow \{r(\tau), \sigma(\tau)^{\delta}\}_{\tau\geqslant 0}$ *as* $h \downarrow 0$, *where* $\{r(\tau), \sigma(\tau)^{\delta}\}_{\tau\geqslant 0}$ *are solutions to* (1) *when* $\eta \equiv 1$.

Let, in addition

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

The preceding approximation result then 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,

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

gets 'closer and closer' to the distribution generated by the sample paths generated by (1), with ρ given by (16). Comparing (13)–(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

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

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).

As is clear, the convergence results in Theorem 3.1 are related to the rate at which discrete-time parameters converge to their continuous time counterparts. Interestingly, in the empirical section we find that not only does the parametrization in (14) provide a sensible picture of the volatility dynamics, it even passes the global consistency test that checks, ex post, the accuracy of the approximation in (15)–(16) and that we present below (see Section 3.2).[15]

We now generalize both (7) and (8) and consider the following approximating model:

$$
\begin{aligned}
{}_{\Delta} r_{n+1} &= {}_{\Delta} r_n + \iota_{\Delta} - \theta_{\Delta} \cdot {}_{\Delta} r_n + {}_{\Delta}\sigma_{n+1}\sqrt{{}_{\Delta} r_n} \cdot {}_{\Delta} u_{n+1},\\
{}_{\Delta}\varepsilon_n &= {}_{\Delta} u_n \cdot {}_{\Delta}\sigma_n, \quad \frac{{}_{\Delta} u_n}{\sqrt{\Delta}} \sim ged_{\upsilon},
\end{aligned}
\tag{20}
$$

[15]The estimators in (19) are based on the moment conditions (17). As we noted earlier, they may be affected by a disaggregation bias. Furthermore, Corradi (2000) questioned the realism of the moment conditions Nelson (1990) originally developed in his seminal paper. Her reasoning can be generalized here as follows. In the third equation of (14), the term generating the diffusion terms is proportional to $(h^{-\delta/2}\alpha_h) \cdot |{}_h u_{hk}|^{\delta}$. Under the third moment condition in (17), this term is $o_p(\sqrt{h})$. Hence, a well-defined diffusion limit can be obtained by scaling 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 $\alpha_h \propto h^q$, $q \in \mathbb{R}$. This leaves three alternatives

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

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.

$$
\begin{aligned}
{}_{\Delta}\sigma_{n+1}^{\delta} &= w_{\Delta} + \alpha_{\Delta}(|{}_{\Delta}\varepsilon_n| - \gamma \cdot {}_{\Delta}\varepsilon_n)^{\delta\eta} + \beta_{\Delta} \cdot {}_{\Delta}\sigma_n^{\delta} \\
&\quad + \alpha_{\Delta} \cdot \mathrm{E}\{(|{}_{\Delta}u_n| - \gamma \cdot {}_{\Delta}u_n)^{\delta\eta}\} \cdot \{{}_{\Delta}\sigma_n^{\delta} - {}_{\Delta}\sigma_n^{\delta\eta}\}.
\end{aligned}
$$

Discretizing time in (20) as in (13), and rearranging, yields

$$
{}_{h}r_{h(k+1)} = {}_{h}r_{hk} + \iota_h - \theta_h \cdot {}_{h}r_{hk} + {}_{h}\sigma_{h(k+1)}\sqrt{{}_{h}r_{hk}} \cdot {}_{h}u_{h(k+1)},
$$

$$
{}_{h}\varepsilon_{hk} = {}_{h}u_{hk} \cdot {}_{h}\sigma_{hk}, \quad \frac{{}_{h}u_{hk}}{\sqrt{h}} \sim ged_{\upsilon},
$$

$$
\begin{aligned}
{}_{h}\sigma_{h(k+1)}^{\delta} - {}_{h}\sigma_{hk}^{\delta} &= w_h - (1 - h^{-\frac{\delta\eta}{2}}\mathrm{E}\{|{}_{h}u_{hk}|^{\delta\eta}(1 - \gamma s_k)^{\delta\eta}\}\alpha_h - \beta_h){}_{h}\sigma_{hk}^{\delta} \\
&\quad + \alpha_h \cdot (|{}_{h}u_{hk}|^{\delta\eta}(1 - \gamma s_k)^{\delta\eta} - \mathrm{E}\{|{}_{h}u_{hk}|^{\delta\eta}(1 - \gamma s_k)^{\delta\eta}\})h^{-\frac{\delta\eta}{2}}\sigma_{hk}^{\delta\eta}.
\end{aligned} \tag{21}
$$

We have

Theorem 3.2. *Let*

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

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

and

$$
\rho \equiv \frac{2^{\frac{\delta\eta - \upsilon + 1}{\upsilon}}\nabla_{\upsilon}^{\delta\eta+1}\Gamma\left(\dfrac{\delta\eta + 2}{\upsilon}\right) \cdot ((1 - \gamma)^{\delta\eta} - (1 + \gamma)^{\delta\eta})}{\Gamma(\upsilon^{-1})\sqrt{(m_{\delta\eta,\upsilon} - n_{\delta\eta,\upsilon}^2)((1 - \gamma)^{2\delta\eta} + (1 + \gamma)^{2\delta\eta}) - 2n_{\delta\eta,\upsilon}^2(1 - \gamma)^{\delta\eta}(1 + \gamma)^{\delta\eta}}}. \tag{24}
$$

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

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

Then, $\{{}_{h}r_{h(k-1)}, {}_{h}\sigma_{hk}^{\delta}\}_{k=0,1,\ldots} \Rightarrow \{r(\tau), \sigma(\tau)^{\delta}\}_{\tau \geqslant 0}$ *as* $h \downarrow 0$, *where* $\{r(\tau), \sigma(\tau)^{\delta}\}_{\tau \geqslant 0}$ *are solutions to* (1) *and* $\{{}_{h}r_{h(k-1)}, {}_{h}\sigma_{hk}^{\delta}\}_{k=0,1,\ldots}$ *are solution to* (21).

In the same vein, 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 via 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 ged assumption for the innovation process u in (21) (see Section 3.1), the q-aml estimator of $b = (\Delta^{-1}\iota_\Delta, \Delta^{-1}\theta_\Delta, \Delta^{-3/2}w_\Delta, \Delta^{-1}\varphi_\Delta, \Delta^{-1/2}\psi_\Delta, \gamma, \delta, \eta, \upsilon)^\top$ 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 $\upsilon \equiv 2$. Motivated by the Monte-Carlo findings reported in Section 2.2, we impose time-scale invariance on δ and η, and assume the same for γ. Accordingly, we re-interpret b as a vector in $\mathbb{R}^5$ (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 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 $\mathbb{R}^5$, with coordinates $\iota, \theta, \overline{\omega}, \varphi, \psi$.

Let

$$p\lim_N \frac{\partial^2}{\partial b \partial b^\top} \mathfrak{L}_N \equiv \ddot{\mathfrak{L}}_\infty,$$

where $\ddot{\mathfrak{L}}_\infty$ is invertible and $\mathfrak{L}_\infty$ is the limit of $\mathfrak{L}_N$, and let $\mathfrak{L}_\infty \equiv \mathfrak{L}_\infty(a_0, b)$ to emphasize the dependence of $\mathfrak{L}_\infty$ on the true parameter vector. Clearly, asymptotic normality of the pseudo-ML estimator holds

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

where J is the (N-scaled) asymptotic variance of the pseudo-score, 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 generate 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 discretize 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 $\{{}_hr_{hk}, {}_h\sigma^{\delta}_{hk}\}_{k=0,1,\ldots}$ solution to

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

then (20) is embedded in $\{{}_hr_{hk}, {}_h\sigma^{\delta}_{hk}\}_{k=0,1,\ldots}$ (namely for $h \equiv \Delta$), yet $\{{}_hr_{hk}, {}_h\sigma^{\delta}_{hk}\}_{k=0,1,\ldots}$ converges weakly to the solution to (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 ${}_hu_{hk}/\sqrt{h}$ from the normal distribution; one obtains ${}_{h,h}\widetilde{r}^{(s)}(a) = \{{}_h\widetilde{r}^{(s)}_{hk}(a)\}_{k=0}^{N/h}$, $s = 1, \ldots, S$, where S is the number of simulations. For each simulation retain the (N) numbers ${}_h\widetilde{r}^{(s)}_{hk}(a)$ that correspond to integer indices of time, and estimate the auxiliary model on each series of simulated data

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

where ${}_{\Delta,h}\widetilde{r}^{(s)}(\cdot)$ denotes the set of the simulated short-term rate with integer indices 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) to the following five-dimensional system:

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

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

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

For large N, the preceding is in fact an equality in distribution, and the covariance matrix of

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

is the covariance matrix of

$$
\widehat{b}_N - \frac{1}{S}\sum_{s=1}^{S} \widehat{b}^{(h)}_{N,s}(a_0), \text{ i.e. } \left(1 + \frac{1}{S}\right)\mathrm{cov}(\widehat{b}^{(h)}_{N,s}(a_0)),
$$

and one has

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

where Γ_0 is the asymptotic covariance matrix of the pseudo-ML 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 multidimensional diffusion) 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 $d > \frac{1}{2}$.

Notice also that (26) does *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 to (1). For comparison purposes, 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

$$\sqrt{N}\left(\widehat{b}_N - \frac{1}{S}\sum_{s=1}^{S} \widehat{b}_{N,s}^{(h)}(\widehat{b}_N)\right) \overset{d}{\rightarrow} \mathrm{N}\left(0, \left(I - \frac{\partial b}{\partial a}(a_0)\right) \ddot{\Omega}_\infty^{-1} J \ddot{\Omega}_\infty^{-1} \left(I - \frac{\partial b}{\partial a}(a_0)^{\top}\right) + \frac{1}{S}\, \ddot{\Omega}_\infty^{-1} J \ddot{\Omega}_\infty^{-1}\right). \tag{28}$$

The previous result can thus be used to gauge the practical importance of disaggregation biases arising from the use of an approximating model.

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.[16] The motivation for using weekly data lies in an attempt to avoid problems raised by market microstructure effects. This is the same data

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

Table 2

Mean	Median	Maximum	Minimum	Std. dev.	Skewness	Kurtosis		
Summary statistics of r								
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. It is a sample of 1135 observation of the US 3-month T-Bill rate. It is observed between May 30, 1973 and February 22, 1995.

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. Non-stationarity 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 non-stationarity with a 90% probability; more generally, one rejects non-stationarity at the 85–90% to the extent of the 15th 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. As mentioned in Sections 2 and 3, we now simplify the representation in (1) and fix $\delta = \eta = 1$. 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. That is, under the restrictions $(\delta, \eta, \gamma) \equiv (1, 1, 0)$, the model was cast as

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

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), $\Delta_0 = (0\ \Delta^{-1}\ 0\ \Delta^{-1}\ 0)^{\top}$, 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 5 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

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

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 $\delta = \upsilon = 2$.[17]

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.[18] Hence, we are left with a specification in which $(\delta, \eta, \upsilon) = (1, 1, 2)$. 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

[17]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)).

[18]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.

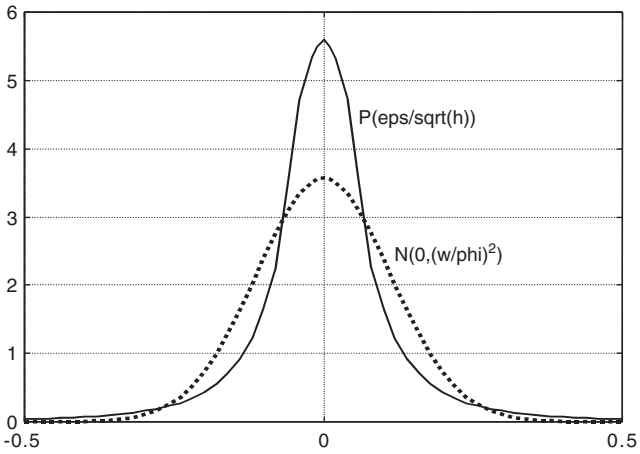

Fig. 3. Stationary distributions of errors.

Table 3
QML estimates of (29)[a]

Parameter	Estimate	t-stat[b]
c_0	1.555×10^{-4}	2.09 (2.58)
c_1	0.9979	6.79×10^2 (9.86×10^2)
w	1.110×10^{-4}	4.36 (3.41)
α	0.1504	13.05 (13.76)
β	0.8728	93.97 (139.18)

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

residuals in Eq. (21) is given by

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

as $h \downarrow 0$. Fig. 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

Table 4
Summary statistics of the conditional volatility σ as filtered by (29)[a]

Mean	Median	Maximum	Minimum	Std. dev.	Skewness	Kurtosis
7.102×10^{-3}	5.483×10^{-3}	2.805×10^{-2}	2.042×10^{-3}	4.306×10^{-3}	1.761	6.048

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

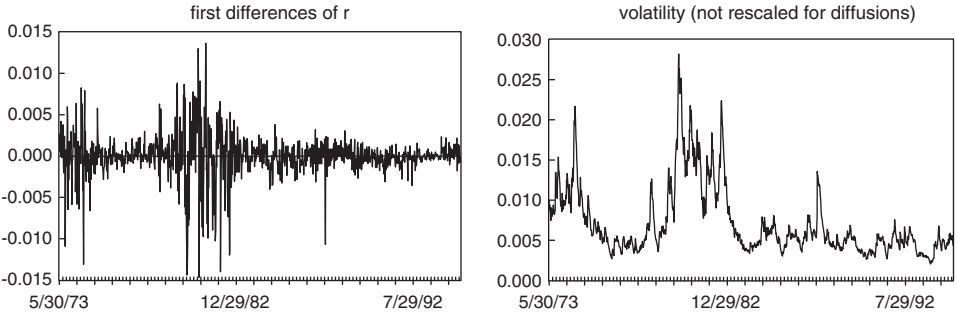

Fig. 4. First difference of the interest rate and estimated volatility.

column of Table 5; its shape suggests that it should capture the usual stylized facts of the unpredictable parts of the vast majority of financial time series.

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.

Table 4 presents summary statistics of the volatility filtered by the model (not yet 'rescaled for diffusions'). Fig. 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 1980s. 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

$$\frac{w}{1 - 0.798 \cdot \alpha - \beta} = 15.458 \times 10^{-3},$$

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.

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 through the moment conditions in Theorem 3.2. 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 (28) and using (30) as the Jacobian of the binding function. The joint test of adequacy of the full vecto $\widehat{b}_N$ is asymptotically $\chi^2(5)$ (see the main text) and here it takes a value equal to 6.53 (the 95% critical level of a $\chi^2(5)$ is 11.05).

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

The second column of Table 5 reports the q-aml estimates of a. To correct their potential disaggregation bias we implement the indirect inference setup by simulating system (1) with the Euler–Maruyama approximation (2) with $h^{-1} \equiv 1300$, which corresponds to generating 25 sub-intervals within a week. 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% probability bands around the corresponding indirect inference estimates. Perhaps more importantly, we formally checked the adequacy of the auxiliary model through the consistency test described in Section 3.2, and found that the adjustment speed of the short-term rate is the only parameter that does not pass the test at the standard 95% level. Finally, we performed a joint test of adequacy of the full vector $\widehat{b}_N$. Under the null that $T_N \equiv R_{5\times5}\Delta\widehat{b}_N \equiv R_{5\times5}[\widehat{b}_N - S^{-1}\sum_{s=1}^{S}\widehat{b}_{N,s}^{(h)}(\widehat{b}_N)] = \mathbf{0}_{5\times1}$ $(R_{5\times5} = I_{5\times5})$, $W_N \equiv \Delta\widehat{b}_N^{\top}\cdot[var(T_N)]^{-1}\cdot\Delta\widehat{b}_N$ is asymptotically chi-square distributed with 5 degrees of freedom. We find that $W_N = 6.5304$, which is well beyond the 95% critical level 11.05.

These findings are of special interest here because Drost and Nijman (1993) showed that only a weakened concept of ARCH model (the so-called weak-ARCH model) does aggregate. In the same vein, Drost and Werker (1996) generalized the Drost and Nijman setup and introduced 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. This is naturally so because

Table 6
Monte-Carlo study[a]

Parameter	Average	Median	Std. dev.
c_0	1.640×10^{-4}	1.610×10^{-4}	3.340×10^{-5}
c_1	0.9974	0.9976	1.764×10^{-3}
w	1.210×10^{-4}	1.130×10^{-4}	4.420×10^{-5}
α	0.1548	0.1544	2.405×10^{-2}
β	0.8665	0.8669	2.056×10^{-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.

standard ARCH models do simply represent an excellent proxy to the corresponding (discrete time) weak-ARCH models. Naturally, these are issues deserving a deep theoretical investigation that lies far beyond the scope of this paper.

Next, we wish to check that the previous estimation results do not depend on the dimension of the simulation experiment ($S = 50$). We thus implement a sort of reverse engineering exercise in which we look for the ARCH model one may expect to estimate should the true data-generating mechanism be (1). Precisely, 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, and remove 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). Note 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. Fig. 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% 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 ith replication and sampled at n. Let $\widehat{\sigma}_{i,n}$ be the corresponding (rescaled) ARCH estimate. We are interested in evaluating the average filtering error in all the simulations, $\{\mathscr{E}_i\}_{i=1}^{5000}$, where $\mathscr{E}_i \equiv \frac{1}{1135}\sum_{n=1}^{1135}(\sigma_{i,n} - \widehat{\sigma}_{i,n})$. Fig. 6 displays the Monte-Carlo distribution of the average filtering error. It has an average value of 9.610×10^{-5} and a standard deviation of 3.275×10^{-3} The RMSE, defined as

$$\sqrt{\frac{1}{5000}\sum_{i=1}^{5000}\left(\frac{1}{1135}\sum_{n=1}^{1135}(\sigma_{i,n} - \widehat{\sigma}_{i,n})^2\right)},$$

is equal to 0.0209.

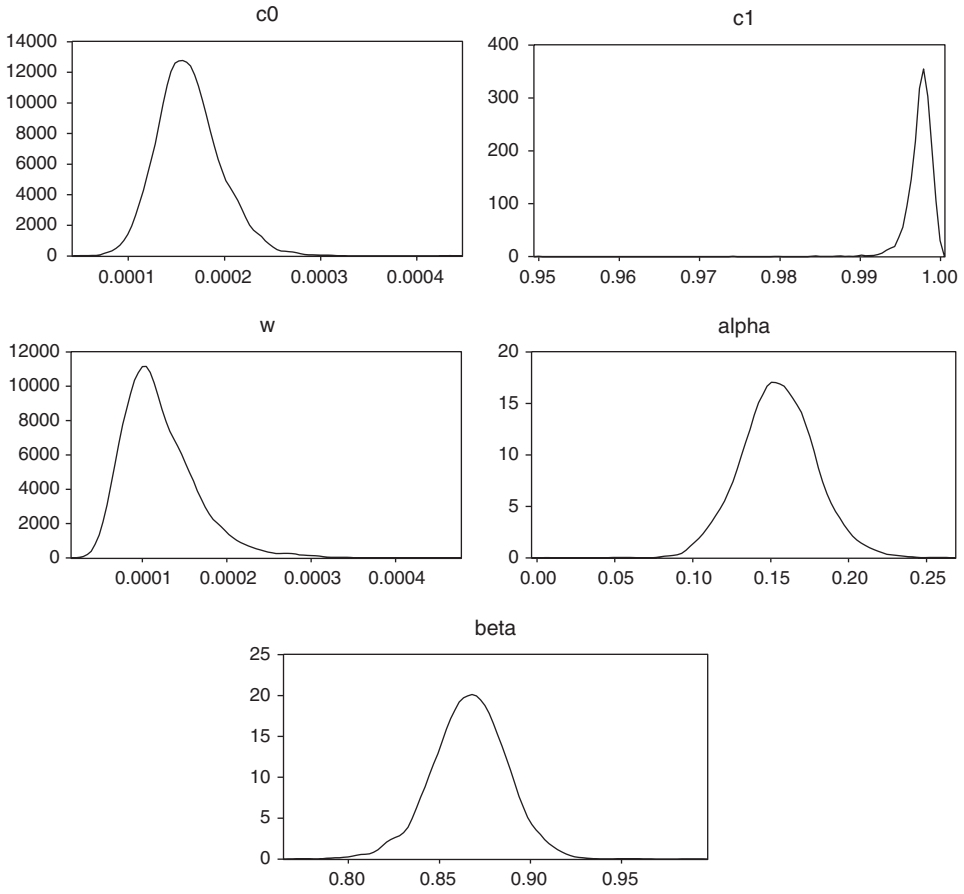

Fig. 5. Monte-Carlo densitites of the ARCH parameters estimates.

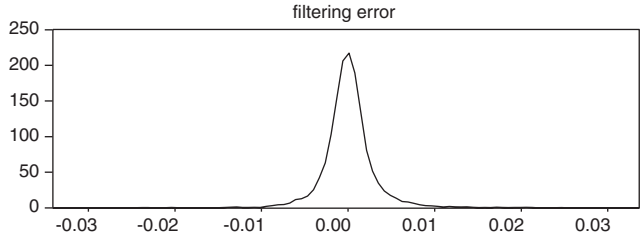

Fig. 6. Monte-Carlo filtering error.

The last objective of this section is to show that even in the presence of an important kind of misspecification which one could be easily faced with in many applications involving the modeling of interest rates, the kind of models considered

Table 7
Monte-Carlo study[a]

	Average	Median	Std. dev.
Volatility filtering error	-3.6815×10^{-5} RMSE: 0.0285 $\omega/\varphi = 7.895 \times 10^{-2}$	-6.0461×10^{-5}	3.9666×10^{-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:

$$\begin{aligned} \mathrm{d}r(\tau) &= \theta(l(\tau) - r(\tau))\,\mathrm{d}\tau + \sqrt{r(\tau)}\sigma(\tau)\,\mathrm{d}W^{(1)}(\tau), \\ \mathrm{d}\sigma(\tau) &= (\omega - \varphi\sigma(\tau))\,\mathrm{d}\tau + \psi\sigma(\tau)\,\mathrm{d}W^{(2)}(\tau), \\ \mathrm{d}l(\tau) &= (b_1 - b_2 l(\tau))\,\mathrm{d}\tau + b_3\sqrt{l(\tau)}\,\mathrm{d}W^{(3)}, \end{aligned}$$

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.

in this paper still remain a valid reference, at least insofar as one considers volatility filtering issues. Precisely, suppose that the data generating process is a three-factor model including the short-term rate, stochastic volatility, and a stochastic central tendency factor. The question we want to answer is: are the filtering results of this paper still valid when we attempt to extract 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). The main message of their work 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 1000 simulated trajectories of a three-factor model that extends in a natural way model (9) as

$$\begin{aligned} \mathrm{d}r(\tau) &= \theta(l(\tau) - r(\tau))\,\mathrm{d}\tau + \sqrt{r(\tau)}\sigma(\tau)\,\mathrm{d}W^{(1)}(\tau), \\ \mathrm{d}\sigma(\tau) &= (\omega - \varphi\sigma(\tau))\,\mathrm{d}\tau + \psi\sigma(\tau)\,\mathrm{d}W^{(2)}(\tau), \\ \mathrm{d}l(\tau) &= (b_1 - b_2 l(\tau))\,\mathrm{d}\tau + b_3\sqrt{l(\tau)}\,\mathrm{d}W^{(3)}, \end{aligned} \tag{32}$$

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 (1997b), i.e. $b_1 = 0.0078$, $b_2 = 0.1257$ and $b_3 = 0.0493$, and repeat the same experiment 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 has 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 Fig. 1. Considered as a (stochastic) volatility filter, model (29) would be hardly rejected as a valid tool of analysis, even in the presence of neglected factors.

5. Conclusion

The intent of this paper was to explore to what extent ARCH models can be practically used for the purpose of providing parameter estimates and volatility filtering for diffusion processes. Since the *standard* ARCH models that have traditionally been used in the empirical literature converge in distribution towards a restrictive class of 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. Our searching strategy in this paper was to find sequences of ARCH models approximating CEV-diffusions with *linear* drift functions. This strategy may be extended quite easily to diffusions with non-linear drift functions. Also, our results rely in practice on using high-frequency data which could be contaminated by market microstructure noise. A very challenging topic for future research is to model such a noise, thus extending the results in Aït-Sahalia et al. (2004) to the heteroskedastic case.

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 indirect inference methods. Finally, the volatility filtering performances of the models are excellent. Even if the data generating mechanism is a *three*-factor model, the volatility filtering errors of a *two*-factor model have the same order of magnitude as in the absence of misspecification. These findings suggest very simple and yet efficient tools to extract the (unobserved) volatility of many multidimensional diffusion processes that are of interest in financial economics and macroeconomics.

Acknowledgements

We thank Yacine Aït-Sahalia, Filippo Altissimo, Ron Gallant, Steven Satchell, José Scheinkman, seminar participants at Princeton University and Cambridge University, and above all, two anonymous referees and the Editor Carl Chiarella for valuable comments on previous drafts of this paper. We acknowledge that after producing the latest revisions of this article, we came across a working paper by Ishida, I. and R.F. Engle: Modeling Variance of Variance: The Square Root, the

Affine, and the CEV-GARCH Models (NYU-Stern, November 2002), in which the authors studied the same CEV-ARCH model we introduced in our 2001 working paper and which we further develop here. As usual, responsibility for any views or errors in this article rests with the authors.

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 to the following stochastic differential equations:

$$\begin{aligned} \mathrm{d}r(\tau) &= (\iota - \theta r(\tau))\,\mathrm{d}\tau + \sigma(\tau)\sqrt{r(\tau)}\,\mathrm{d}W^{(1)}(\tau), \\ \mathrm{d}\sigma(\tau)^{\delta} &= (\omega - \varphi\sigma(\tau)^{\delta})\,\mathrm{d}\tau + \psi\sigma(\tau)^{\delta}\,\mathrm{d}W^{(3)}(\tau), \end{aligned}$$

where $\{W^{(j)}(\tau)\}_{\tau\geqslant 0}, j = 1$ and 3, are $\mathscr{F}(\tau)$-Brownian motions. This has been shown in Theorem 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 a similar convergence result.

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

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

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

$$\lim_{h\downarrow 0} h^{-1}\mathrm{E}\{({}_{h}r_{hk} - {}_{h}r_{h(k-1)})({}_{h}\sigma^{\delta}_{h(k+1)} - {}_{h}\sigma^{\delta}_{hk})|\mathscr{F}_{hk}\}$$

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

By (18), and the fact that ${}_{h}u_{hk}/\sqrt{h}$ is $ged_{(\upsilon)}$ for each h,

$$\begin{aligned} &\lim_{h\downarrow 0} h^{-1}\mathrm{E}\{({}_{h}r_{hk} - {}_{h}r_{h(k-1)})({}_{h}\sigma^{\delta}_{h(k+1)} - {}_{h}\sigma^{\delta}_{hk})|\mathscr{F}_{hk}\} \\ &\quad = \lim_{h\downarrow 0} h^{-1}\mathrm{E}\{(\iota_h - \theta_h\cdot {}_{h}r_{h(k-1)} + {}_{h}\sigma_{hk}\sqrt{{}_{h}r_{h(k-1)}}\cdot {}_{h}u_{hk}) \\ &\qquad \times(w_h + (\alpha_h|{}_{h}u_{hk}|^{\delta}(1-\gamma s_k)^{\delta}h^{-\frac{\delta}{2}} + \beta_h - 1){}_{h}\sigma^{\delta}_{hk})|\mathscr{F}_{hk}\} \\ &\quad = \lim_{h\downarrow 0} h^{-1}\mathrm{E}\{{}_{h}u_{hk}(\alpha_h|{}_{h}u_{hk}|^{\delta}(1-\gamma s_k)^{\delta}h^{-\frac{\delta}{2}} + \beta_h - 1)\cdot {}_{h}\sigma^{\delta+1}_{hk}|\mathscr{F}_{hk}\}\cdot\sqrt{{}_{h}r_{h(k-1)}} \\ &\quad = \lim_{h\downarrow 0} h^{-1-\frac{\delta}{2}}\alpha_h\cdot\mathrm{E}\{{}_{h}u_{hk}|{}_{h}u_{hk}|^{\delta}(1-\gamma s_k)^{\delta}\cdot {}_{h}\sigma^{\delta+1}_{hk}|\mathscr{F}_{hk}\}\cdot\sqrt{{}_{h}r_{h(k-1)}} \\ &\quad = \lim_{h\downarrow 0}\frac{\alpha_h}{\sqrt{h}}\{((1-\gamma)^{\delta} - (1+\gamma)^{\delta})\int_{\mathbb{R}_+} x^{\delta+1}p(\mathrm{d}x)\}\cdot {}_{h}\sigma^{\delta+1}_{hk}\cdot\sqrt{{}_{h}r_{h(k-1)}}, \end{aligned}$$

where $p(\cdot)$ denotes the $ged_{(\upsilon)}$ density, or

$$\lim_{h\downarrow 0} h^{-1}\mathrm{E}\{({}_hr_{hk} - {}_hr_{h(k-1)})({}_h\sigma^{\delta}_{h(k+1)} - {}_h\sigma^{\delta}_{hk})|\mathscr{F}_{hk}\}$$
$$= \lim_{h\downarrow 0} \frac{\alpha_h}{\sqrt{h}} \{(1-\gamma)^{\delta} - (1+\gamma)^{\delta}\} K \cdot {}_h\sigma^{\delta+1}_{hk} \cdot \sqrt{{}_hr_{h(k-1)}};$$

here,

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

By using (18),

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

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

Hence,

$$\lim_{h\downarrow 0} h^{-1}\mathrm{E}\{({}_hr_{hk} - {}_hr_{h(k-1)})({}_h\sigma^{\delta}_{h(k+1)} - {}_h\sigma^{\delta}_{hk})|\mathscr{F}_{hk}\} = \frac{\psi\overline{K}}{\sqrt{Z}}\,\sigma^{\delta+1} \cdot \sqrt{r},$$

where

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

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. □

A.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:

$$\sigma^{\delta}_{n+1} = w + \beta\sigma^{\delta}_n + \alpha(1-\gamma s_n)^{\delta\eta}(|u_n|^{\delta\eta} - \mathrm{E}(|u_n|^{\delta\eta}))\sigma^{\delta\eta}_n, \quad \gamma \in (-1,1). \tag{A.1}$$

The main difference between model (20) and model (A.1) is the way how asymmetries in volatility are modeled. Suppose for instance that $\gamma > 0$ in model (A.1). 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. In this respect, model (A.1) represents a variant 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 (A.1) converges in distribution to (1) as the sampling frequency gets higher and higher. Consider the following approximating scheme:

$$
\begin{aligned}
{}_h\sigma^{\delta}_{h(k+1)} - {}_h\sigma^{\delta}_{hk} = w_h - (1-\beta_h){}_h\sigma^{\delta}_{hk} + \alpha_h(1-\gamma s_k)^{\delta\eta} \\
\times\{|{}_hu_{hk}|^{\delta\eta} - \mathrm{E}(|{}_hu_{hk}|^{\delta\eta})\}h^{-\frac{\delta\eta}{2}}{}_h\sigma^{\delta}_{hk}
\end{aligned}
$$

and introduce the following moment conditions:

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

For each h, we have that

$$
\mathrm{E}\{(1-\gamma s_k)^{\delta\eta}(|{}_hu_{hk}|^{\delta\eta} - \mathrm{E}(|{}_hu_{hk}|^{\delta\eta}))h^{-\frac{\delta\eta}{2}}|\mathscr{F}_{hk}\} = 0
$$

and so the drift per unit of time is

$$
h^{-1}\mathrm{E}({}_h\sigma^{\delta}_{h(k+1)} - {}_h\sigma^{\delta}_{hk}|\mathscr{F}_{hk}) = h^{-1}w_h - h^{-1}(1-\beta_h){}_h\sigma^{\delta}_{hk}.
$$

By taking limits for $h \downarrow 0$, and using the moment conditions (A.2), we obtain the drift function of volatility in (1).

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

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

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

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

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

where

$$\lim_{h\downarrow 0} \mathrm{E}\left\{\frac{{}_h u_{hk}}{\sqrt{h}}(1-\gamma s_k)^{\delta\eta}\left(\left|\frac{{}_h u_{hk}}{\sqrt{h}}\right|^{\delta\eta} - 2n_{\delta\eta,\upsilon}\right)\right\}$$
$$= \mathrm{E}\{\widetilde{u}(1-\gamma\cdot\mathrm{sign}(\widetilde{u}))^{\delta\eta}(|\widetilde{u}|^{\delta\eta} - 2n_{\delta\eta,\upsilon})\}$$
$$= \{(1-\gamma)^{\delta\eta} - (1+\gamma)^{\delta\eta}\}\frac{2^{\frac{\delta\eta-\upsilon+1}{2}}\nabla_\upsilon^{\delta\eta+1}}{\Gamma\left(\frac{1}{\upsilon}\right)^2}$$
$$\times\left\{\Gamma\left(\frac{\delta\eta+2}{\upsilon}\right)\Gamma\left(\frac{1}{\upsilon}\right) - \Gamma\left(\frac{\delta\eta+1}{\upsilon}\right)\Gamma\left(\frac{2}{\upsilon}\right)\right\}$$

and $\widetilde{u}$ is ged_υ.

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

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

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

Appendix B. 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

$$r_n = \iota h + (1-\theta h)r_{n-1} + \sqrt{h}\sigma_{n-1}\sqrt{r_{n-1}}u_n, \quad n = 1,\ldots,\widetilde{N}, \tag{B.1}$$

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 (B.1) yields

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

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

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

Our objective is to estimate each point of the sequence $\left\{\overline{\sigma}^{\ell}_{j}\right\}_{j=\ell,2\ell,\ldots,\widetilde{N}/\ell}$ 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

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

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

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

$$v_j \equiv \overline{\sigma}^{\ell}_{j}\cdot\sqrt{\frac{h(1-(1-\theta h)^{2\ell})}{1-(1-\theta h)^2}}, \quad j=\ell,2\ell,\ldots,\widetilde{N}/\ell. \tag{B.2}$$

Therefore, an estimate of each point of the sequence $\{\overline{\sigma}^{\ell}_{j}\}_{j=\ell,2\ell,\ldots,\widetilde{N}/\ell}$ is obtained by inverting formula (B.2) to form the desired sequence

$$\overline{\sigma}^{\ell}_{j} = \sqrt{\frac{1-(1-\theta h)^2}{h(1-(1-\theta h)^{2\ell})}}v_j, \quad j=\ell,2\ell,\ldots,\widetilde{N}/\ell. \tag{B.3}$$

In this paper, we used

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

and the estimates of θ reported in Table 6 of the main text are such that

$$\sqrt{\frac{1-(1-\theta h)^2}{h(1-(1-\theta h)^{2\ell})}}$$

is close to 7.218.

The filtered series of volatility reported throughout the paper are based on formula (B.3) (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 (B.3) is based on the assumption that the (small) changes of $\sigma\sqrt{r}$ are not autocorrelated. Relaxing such an assumption requires a more complicated approach with continuous updatings. A reliable alternative is to numerically search for a conversion factor similar to (B.3). 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 4. Then we defined

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

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 $C \approx 6.928$.

References

Aït-Sahalia, Y., 1996. Testing continuous-time models of the spot interest rate. Review of Financial Studies 9, 385–426.

Aït-Sahalia, Y., 2002. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrica 70, 223–262.

Aït-Sahalia, Y., 2003. Closed-form likelihood expansions for multivariate diffusions. Working Paper.

Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2004. How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies 18, 351–416.

Andersen, T.G., Lund, J., 1997a. Estimating continuous-time stochastic volatility models of the short-term interest rate. Journal of Econometrics 77, 343–377.

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.

Andersen, T.G., Bollerslev, T., Diebold, F.X., 2002. Parametric and nonparametric volatility measurement. In: Aït-Sahalia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics. North-Holland, Amsterdam, forthcoming.

Black, F., 1976. Studies of stock price volatility changes. In: Proceedings of the 1976 Meeting of the American Statistical Association, pp. 177–181.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–659.

Bollerslev, T., 1986. Generalized autoregressive conditional teteroskedasticity. Journal of Econometrics 31, 307–327.

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

Bollerslev, T., Wooldridge, J.M., 1992. Quasi-maximum likelihood estimation of dynamic models with time varying covariances. Econometric Reviews 11, 143–172.

Brandt, M., Santa-Clara, P., 2002. Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. Journal of Financial Economics 63, 161–210.

Brenner, R.J., Harjes, R.H., Kroner, K.F., 1996. Another look at alternative models of the short-term interest rate. Journal of Financial and Quantitative Analysis 31, 85–108.

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, 161–186.

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

Chan, K.C., Karolyi, G.A., Longstaff, F.A., Sanders, A.B., 1992. An empirical comparison of alternative models of short-term interest rate. Journal of Finance 47, 1209–1228.

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, 145–153.

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.

Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. Mathematical Finance 6, 379–406.

Duffie, D., Singleton, K.J., 1993. Simulated moments estimation of Markov models of asset prices. Econometrica 61, 929–952.

Durham, G.B., 2003. Likelihood-based specification analysis of continuous-time models of the short-term interest rate. Journal of Financial Economics 70, 463–487.

Durham, G.B., Gallant, A.R., 2002. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes (with discussion). Journal of Business and Economic Statistics 20, 297–338.

Fornari, F., Mele, A., 1997a. Weak convergence and distributional assumptions for a general class of nonlinear ARCH models. Econometric Reviews 16, 205–229.

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, 657–681.

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 University Press, Oxford.

Gouriéroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8, S85–S118.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Heston, S., 1993. A closed form solution for options with stochastic volatility with application to bond and currency options. Review of Financial Studies 6, 327–344.

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, 143–151.

Karatzas, I., Shreve, S., 1991. Brownian Motion and Stochastic Calculus. Springer, Berlin.

Longstaff, F., Schwartz, E., 1992. Interest-rate volatility and the term structure: a two factor general equilibrium model. Journal of Finance 47, 1259–1282.

Merton, R., 1973. Theory of rational option pricing. Bell Journal of Economics and Management Science 4, 637–654.

Nelson, D.B., 1990. ARCH models as diffusion approximation. Journal of Econometrics 45, 7–38.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347–370.

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.

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.

Revuz, D., Yor, M., 1999. Continuous Martingales and Brownian Motion. Springer, Berlin.

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

Schwert, W., 1989. Why does stock market volatility change over time? Journal of Finance 44, 1115–1154.

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

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

Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177–188.

Wiggins, J., 1987. Option values under stochastic volatility. Theory and empirical estimates. Journal of Financial Economics 19, 351–372.