

# Modeling Variance of Variance: The Square-Root, the Affine, and the CEV GARCH Models\*

Isao Ishida<sup>†</sup>

Department of Economics  
University of California, San Diego

Robert F. Engle

Stern School of Business  
New York University

and

Department of Economics  
University of California, San Diego

First Version: August 20, 2001

This Version: November 20, 2002

## Abstract

This paper develops a new econometric framework for investigating how the sensitivity of the financial market volatility to shocks varies with the volatility level. For this purpose, the paper first introduces the square-root (SQ) GARCH model for financial time series. It is an ARCH analogue of the continuous-time square-root stochastic volatility model popularly used in derivatives pricing and hedging. The variance of variance is a linear function of the conditional variance in the SQGARCH and of the square of it in the GARCH. After showing some implications of this difference, the paper introduces the constant-elasticity-of-variance (CEV) GARCH model, which allows more flexible fitting of variance-of-variance dynamics. The paper develops conditions for stationarity, the existence of finite moments,  $\beta$ -mixing, and other properties of the conditional variance process via the general state-space Markov chains approach. In particular, the paper generalizes the strict stationarity condition for the GARCH(1,1) and gives attention to a discrete-time analogue of the phenomenon known in the continuous-time finance literature as “volatility-induced stationarity,” which may occur with the integrated or mildly explosive CEVGARCH, shedding light on the stabilizing effect of the variance of variance. A diffusion limit for this model is also established. Several alternative models including the affine and the exponential CEV models are explored. The empirical estimates of the CEVGARCH model for the S&P 500 index and DM/US\$ exchange rate time series suggest that the variance of variance grows faster than linearly with the conditional variance.

*Key words: Variance of variance; square-root GARCH; CEVGARCH; continuous-time stochastic volatility model; general state-space Markov chains approach; diffusion approximation.*

---

\*The authors wish to thank James Hamilton, Bruce Lehmann, Piers Redmore, and Mark Van De Vyver for their helpful comments and suggestions on earlier versions of the paper, and Patrick Fitzsimmons and Howell Tong for their valuable advice on the use of the Markov chains approach. This research was supported in part by NSF Grant SES-9730062 and by the NBER Asset Pricing Group.

<sup>†</sup>Corresponding author. Mailing address: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508. Email: iishida@econ.ucsd.edu

# 1 Introduction

The intensive research of the past two decades has uncovered many features of the way in which financial market volatility changes through time. One important empirical question that has remained relatively neglected in the volatility literature is how the sensitivity of volatility to shocks varies with the volatility level. This is directly related, for example, to the questions of how rapidly and to what extent the market becomes turbulent when consecutively positive volatility shocks hit the market and how it may revert to quiescence. In this paper, we propose a new discrete-time econometric framework for empirically addressing these questions and apply it to the S&P 500 index and deutsche mark/US dollar exchange rate time series data.

Time-varying volatility violates a key assumption of the constancy of volatility in the Black-Scholes option pricing theory (Black and Scholes [11], Merton [70]) and manifests itself in option prices as Black-Scholes implied volatility “smiles” and “smirks.” In the continuous-time option pricing literature, one of the most popular models designed to capture this phenomenon is Heston’s [61] bivariate diffusion model in which the volatility dynamics are driven by a second source of uncertainty modeled by a Brownian motion, possibly correlated with the first directly driving the asset price dynamics. One feature that distinguishes this model from other bivariate continuous-time stochastic volatility (SV) models (see, e.g., Wiggins [89] and Hull and White [63]) is the square-root specification of the diffusion function in the volatility equation. In these SV models, how the sensitivity of volatility to shocks changes with the level of volatility is determined by the diffusion function. The square-root SV (SQSV) diffusion model owes much of its popularity to the analytical tractability afforded by the square-root diffusion function specification (See Feller [48], Cox, Ingersoll, and Ross [31]) leading to closed or nearly closed-form pricing formulas for options written on the underlying following such a process (see Heston [61])<sup>1</sup>, rather than to economic or empirical justifications. There has not been much discussion on the specification of the diffusion function of the volatility equation in spite of the significant amount of attention that the same issue of the univariate short-term interest rate models has attracted<sup>2</sup> (e.g., Chan et al. [21], Brenner, Harjes, and Kroner [16], Ait-Sahalia [1], [2], [3], and Conley et al. [29]). Empirical rejection of the SQSV diffusion model (see, e.g., Chernov and Ghysels [28] and Andersen, Benzoni, and Lund [4]) has led to the recent rise in popularity of a new generation of models that superimpose jump components in the asset price equation while still retaining the square-root specification of the diffusion function of the volatility equation (e.g., Bates [7], [8] and Pan [77]), a notable exception to this trend being the constant-elasticity-of-variance stochastic volatility (CEVSV) diffusion model of

---

<sup>1</sup>Recently, Jian and Kinght [64] and Singleton [81] introduced characteristic-function based estimators that exploit the very properties that lead to closed-form option pricing formulas.

<sup>2</sup>Note that the debate in the short-term interest rate literature that parallels ours has been about the diffusion function of the univariate interest rate equation per se rather than that of the volatility equation of the bivariate stochastic volatility interest rate process.

Engle and Lee [44] and Jones [65]. The choice of this extension of the SQSV over others that have the potential of improving the fit to data is at least partially for retaining analytical convenience, particularly in option pricing applications.

On the discrete-time, empirical side, the success of the ARCH framework introduced by Engle [43] has led to explosive growth of another strand of the time-varying volatility literature<sup>3</sup>. While much of the stochastic volatility option pricing theory follows the tradition of continuous-time financial economics for analytical tractability, much of the empirical research has been done within the discrete-time econometric framework for ease of estimation. Nelson [72], however, has shown that some of the most successful discrete-time ARCH models including Bollerslev’s [12] GARCH(1,1) specification are closely related to their continuous-time counterparts through weak convergence. A comprehensive analysis by Duan [38], [39] that extends the results of Nelson [72] has shown that most of the popularly used continuous-time bivariate SV diffusion models, in fact, can be approximated by members of the “augmented GARCH” family.

To our knowledge, however, no ARCH counterpart of Heston’s continuous-time square-root volatility specification has been studied in published work. Given the plethora of the ARCH family of models and the popularity of the SQSV diffusion, this absence of the square-root specification in the ARCH literature is somewhat surprising<sup>4</sup>. To fill this gap, we propose the *square-root GARCH* (or *SQGARCH*) model as a discrete-time ARCH analogue to the SQSV diffusion model. The SQGARCH conditional variance equation has the *volatility exponent* of one half, which is defined to be the power to which the conditional variance in the variance shock coefficient is raised, and consequently the conditional variance of conditional variance (or *variance of variance* for short) grows linearly with the current level of conditional variance whereas in the standard GARCH(1,1) model, which has the volatility exponent of one, it grows linearly with the square of the current level of conditional variance. Hence, the conditional variance increases from average levels to high levels in a less pronounced manner in response to consecutively positive variance shocks of a given size under the SQGARCH than the GARCH. We then make the natural extension of freeing the volatility exponent to take values other than one or one-half. In this new flexible volatility exponent model, which subsumes the standard GARCH and the SQGARCH models as special cases, the elasticity of variance of variance w.r.t. the current level of conditional variance

---

<sup>3</sup>For surveys, see Bollerslev, Chou, and Kroner [14], Bollerslev, Engle, and Nelson [15], Hamilton [56, Chapter 21], Bera and Higgins [9], Palm [78], or Gouriéroux [53]. Also, see Li, Ling, and McAleer [68] for a survey of recent findings on the probabilistic properties of ARCH-type models and statistical estimation issues.

<sup>4</sup>Heston and Nandi [62] showed that their ARCH-type model that admits a closed-form description of the multi-period return distribution (an attractive feature for the purpose of option pricing) has a bonus of converging weakly under a suitable reparameterization scheme to a special case of the continuous-time SQSV process. This special case is not truly a bivariate diffusion process since the two Brownian motions driving the asset price and its volatility are perfectly correlated. We note, however, that the augmented GARCH model of Duan [39] nests a standard deviation based model that is linked through weak convergence to the bivariate diffusion model of Stein and Stein [82], which in turn is closely related to Heston’s SQSV model.

is constant at twice the value of the volatility exponent. For this reason, we call it the *constant-elasticity-of-variance GARCH* (or *CEVGARCH*) model. In this model, the variance of variance is a direct measure of the level-dependent sensitivity of the conditional variance to the shocks. The CEVGARCH model can be considered as a discrete-time ARCH analogue of the CEVSV diffusion model of Engle and Lee [44] and Jones [65]. Note that, although we retain the familiar acronym “CEV” in naming our new model, the object on the LHS of the equation is the conditional variance so that “variance” in the term *constant-elasticity-of-variance* refers to the variance of variance.

When we cast the models in the CEV framework, a curious fact that we immediately notice is that in the discrete-time volatility literature the models with the volatility exponent of one, namely the GARCH(1,1) model and its asymmetric variants, are the standard while in the continuous-time SV literature the model with the exponent of a half, namely the SQSV diffusion model, is the standard even when the actual time series being modeled are the same. However, the value of the elasticity of variance of variance is a key determinant of the temporal evolution of the conditional distribution of financial time series and may have important implications for financial applications. Our CEVGARCH model provides a framework in which this variance-of-variance elasticity issue may be investigated empirically. Mimicking the family of parametric one-factor diffusion models in the interest rate literature, from which the SQSV diffusion model originated, we suggest a few other possibilities for generalization that make the ARCH econometric toolbox more complete. The framework of the CEVGARCH model also enables us to understand the stabilizing effect of the variance of variance that may induce stationarity even when the deterministic autoregressive component of the conditional variance equation is explosive. We derive the conditions for “variance-of-variance-induced stationarity” using the powerful tools of the general state-space Markov chains approach.

The rest of the paper is organized as follows. In section 2, we illustrate how the variance-of-variance dynamics is constrained in the standard GARCH(1,1) model. In Section 3, we first study some properties of the SQGARCH model and then propose the CEVGARCH model that nests the GARCH and the SQGARCH as special cases, and explore possibilities for extending it in several directions. In Section 4, we further study the statistical properties of the CEVGARCH(1,1) model in a more formal setting. In Section 5, we present empirical examples using the Standard and Poor’s 500 index and the deutsche mark/US dollar exchange rate data. In the last section, we give a brief conclusion.

## 2 Variance of variance in the standard GARCH(1,1) model

Throughout this paper, we work within the setting of a fixed, complete, filtered probability space  $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$  and use the notation  $E_t[Z] := E[Z | \mathcal{F}_t]$  and  $Var_t(Z) := E[(Z - E_t[Z])^2 | \mathcal{F}_t]$  where

$Z \in \mathcal{F}$  is any random variable<sup>5</sup>. To see the effect of introducing alternative specifications for the variance shock coefficient, we first consider the following version of the standard GARCH(1,1) model of Bollerslev [12] for a time series  $\{R_t : t \in \mathbb{Z}_+ := \{0, 1, \dots\}\}$  under study, e.g., the period  $t$  return on some financial asset:

$$R_t = \mu_t + \varepsilon_t \quad (2.1)$$

$$\varepsilon_t = \sqrt{h_t} z_t \quad (2.2)$$

$$h_t = \omega + \beta h_{t-1} + \alpha \varepsilon_{t-1}^2 \quad (2.3)$$

where  $\mu_t := E_{t-1}[R_t]$ ,  $\{z_t \in \mathcal{F}_t : t \in \mathbb{Z} := \{\dots, -1, 0, 1, \dots\}\}$  is a given sequence of random shocks with  $E_{t-1}[z_t] = 0$  and  $Var_{t-1}[z_t] = 1$ ,  $\{h_t : t \in \mathbb{Z}_+\}$  is initialized by a random variable  $h_0$  with some initial probability distribution  $\pi_0$  so that  $h_t \in \mathcal{F}_{t-1}$  is the conditional variance of  $R_t$ , and  $\omega$ ,  $\beta$ , and  $\alpha$  are positive constants. In this expository section and the next, we simply assume that  $\{z_t\}$ ,  $\pi_0$ , and the constants satisfy conditions, including  $\alpha + \beta < 1$ , for the covariance stationarity of  $\{\varepsilon_t\}$ . By rearranging the conditional variance equation in (2.3), we obtain:

$$\begin{aligned} h_t &= \omega + (\alpha + \beta) h_{t-1} + \alpha (\varepsilon_{t-1}^2 - h_{t-1}) \\ &= \omega + \gamma h_{t-1} + \alpha h_{t-1} \eta_{t-1} \end{aligned} \quad (2.4)$$

where  $\gamma := \alpha + \beta$  and  $\eta_t := z_t^2 - 1$ . Written in this form, the conditional variance  $h_t$  is the sum of a constant term  $\omega$ , an autoregressive term  $\gamma h_{t-1}$ , and a zero-mean “variance shock” term  $\alpha h_{t-1} \eta_{t-1}$ .

Although the value of the conditional variance  $h_1$  for  $\varepsilon_1$  is known today (time  $t = 0$ ), i.e.,  $Var_0(h_1) = 0$ , those of  $h_t$ 's with  $t > 1$  are not, and therefore, conditional on the information available today, have nonzero variances  $Var_0(h_t) > 0$ .  $Var_0(h_t)$  indicates how variable the future one-period-ahead conditional variances themselves are expected to be. For example, the variance (conditional on  $\mathcal{F}_0$ ) of the variance (conditional on  $\mathcal{F}_1$ ) of  $\varepsilon_2$  is:

$$Var_0(h_2) = \alpha^2 h_1^2 E_0[\eta_1^2] = (\kappa_z - 1) \alpha^2 h_1^2 \quad (2.5)$$

where  $\kappa_z$  denotes the conditional kurtosis of  $z_t$ , which we assume to be a finite constant in this section and the next. If the distribution of  $z_t$  is standard normal, then  $\kappa_z - 1 = 2$ . Even when  $\kappa_z$  is infinite (and consequently the variance of variance is infinite),  $\alpha h_1$  is a useful measure of dispersion of the conditional distribution of time-varying variance. (2.5) implies that the conditional variance itself becomes more volatile as the current level of the conditional variance increases. As Campbell and Hentschel [19] and Nelson and Foster [75] noted, the variance of variance increases more than proportionately with the level of conditional variance. However, this feature of the dynamics of conditional variance implied by the GARCH model may or may not fit the data well.

---

<sup>5</sup>Of course, for the weak convergence discussion of Subsection 4.4, it does not matter on what particular probability spaces the processes and random variables involved are defined.

An interesting aspect of this growth rate is that it is fast enough to cause what we call the “variance-of-variance-induced strict stationarity” for the mildly explosive as well as the integrated GARCH(1,1) processes with  $\gamma \geq 1$  and  $E[\ln(\beta + \alpha z_t^2)] < 0$ , which may be considered a discrete-time analogue of the “volatility-induced stationarity” studied by Conley et al. [29] in the context of diffusion models of interest rates (See also Hansen and Scheinkman [59]). We will elaborate more on this in Section 4.

If we further rearrange the terms in (2.4), the conditional variance equation becomes:

$$h_t - h_{t-1} = \varphi(\theta - h_{t-1}) + \alpha h_{t-1} \eta_{t-1} \quad (2.6)$$

where  $\varphi := 1 - \gamma$  determines the speed at which the conditional variance reverts to its “long-run” mean  $\theta := E[h_t] = \omega(1 - \gamma)^{-1}$  if  $\gamma < 1$  as we are assuming in this section. When the conditional variance equation is written in this form, the resemblance of the GARCH(1,1) model to the following bivariate SV diffusion model studied by Nelson [72] becomes apparent:

$$d \ln Y_t = \mu dt + \sqrt{v_t} dW_{1,t} \quad (2.7)$$

$$dv_t = \varphi(\theta - v_t) dt + \sigma_v v_t dW_{2,t} \quad (2.8)$$

where  $\mu, \varphi, \theta \in \mathbb{R}$ , and  $\sigma_v > 0$  are constants satisfying a certain set of conditions. Here,  $Y_t$  is the stochastic process of the researcher’s interest, e.g., an asset price process, defined in terms of a bivariate stochastic differential equations system driven by some two-dimensional Brownian motion  $(W_{1,t}, W_{2,t})$  in which the square of the diffusion function of the asset price equation,  $v_t$ , itself follows another diffusion process. Nelson [72] derived a rigorous description of the link between the discrete-time GARCH(1,1) process and its continuous-time diffusion counterpart. We will return to this topic in Subsection 4.4.

### 3 The square-root GARCH model and further extensions

#### 3.1 The square-root GARCH(1,1) model

Although the GARCH(1,1) has become the standard model for volatility dynamics among the ARCH family of models in much of the empirical research in macroeconomics and finance, many of the researchers and practitioners working with continuous-time financial modeling have preferred the following square-root specification of Heston’s [61] (the continuous-time SQSV model) to (2.8):

$$dv_t = \varphi(\theta - v_t) dt + \sigma_v \sqrt{v_t} dW_t \quad (3.1)$$

Sharing the same form, the volatility specification (3.1) inherits the analytical tractability of the CIR interest rate model (see Cox, Ingersoll, and Ross [31]). Its popularity relative to similar models such as (2.8) in financial applications is primarily due to this fact rather than to empirical or theoretical justification.

(3.1) suggests the following ARCH analogue:

$$h_t - h_{t-1} = \varphi(\theta - h_{t-1}) + \alpha\sqrt{h_{t-1}}\eta_{t-1} \quad (3.2)$$

which replaces  $h_{t-1}$  in the variance shock term in (2.6) by the conditional standard deviation  $\sqrt{h_{t-1}}$ , or written in a more conventional way:

$$h_t = \omega + \gamma h_{t-1} + \alpha\sqrt{h_{t-1}}\eta_{t-1} \quad (3.3)$$

which we call the *square-root GARCH(1,1)* model, or *SQGARCH(1,1)* model for short. The constants and  $\{\eta_t\}$  are as in the last section. As we will see in the next section, an additional restriction  $0 < \alpha < 2\omega^{1/2}\gamma^{1/2}$  with  $\omega, \gamma > 0$  is sufficient for ensuring the positivity a.s. of the conditional variance  $h_t$ .

The sensitivity of the one-period-ahead conditional variance to the shock  $\eta_t$  in the SQGARCH model still grows with the current level of the conditional variance without a bound, but at a slower rate than in the GARCH model. Also, the entire variance shock term,  $\alpha\sqrt{h_t}\eta_t$ , may still realize an arbitrarily large value as long as  $P(\eta_t > c) > 0$  for any  $c > 0$  unlike in the case of the modified GARCH (MARCH) model of Friedman and Laibson [50] that places an upper bound on the corresponding component.

Before introducing a more general model that subsumes both the GARCH and the SQGARCH as special cases, it is useful to take advantage of the relative analytical ease afforded by these two models and see the effects of different powers to which the current conditional variance in the shock component is raised. Also, the SQGARCH(1,1) being a discrete-time ARCH-type counterpart of a very popular continuous-time SQSV diffusion model, its properties are interesting in its own right. If the SQGARCH(1,1) process generates the data, the unconditional variance of  $\varepsilon_t$  is  $\omega(1-\gamma)^{-1}$  if  $\gamma < 1$  as in the standard GARCH(1,1) case. Assuming that the constants and the distribution of  $\{z_t\}$  are such that  $E[h_t^2]$  is time-invariant (possibly infinite), the unconditional kurtosis of  $\varepsilon_t$  in the SQGARCH(1,1) case is:

$$\kappa_\varepsilon := \frac{E[\varepsilon_t^4]}{\{E[\varepsilon_t^2]\}^2} = \frac{\kappa_z E[h_t^2]}{\{E[h_t]\}^2} = \kappa_z \left\{ 1 + \frac{(\kappa_z - 1)\alpha^2}{\omega(1 + \gamma)} \right\} \quad (3.4)$$

whereas in the GARCH(1,1) case it is

$$\kappa_\varepsilon = \frac{\kappa_z(1 - \gamma^2)}{1 - \gamma^2 - (\kappa_z - 1)\alpha^2} \quad (3.5)$$

Under the SQGARCH(1,1), the range of possible finite values of the unconditional kurtosis  $\kappa_\varepsilon$  for a given finite value of  $\kappa_z$  is bounded from above by  $\kappa_z(2\kappa_z - 1)$ , which is 15 if the distribution of  $z_t$  is standard normal, because of the inequality constraint  $\alpha < 2\omega^{1/2}\gamma^{1/2}$  for positivity. The GARCH(1,1) model, while not placing any upper bound on the range of possible finite values of  $\kappa_\varepsilon$ , requires  $\gamma^2 + (\kappa_z - 1)\alpha^2 < 1$  for  $\kappa_\varepsilon < \infty$ .

The variance of variance is dependent on the current level of conditional variance under the SQGARCH(1,1) as well:

$$Var_{t-1}(h_{t+1}) = \alpha^2 h_t E_{t-1} \left[ (z_t^2 - 1)^2 \right] = (\kappa_z - 1) \alpha^2 h_t \quad (3.6)$$

In the SQGARCH(1,1) case, it grows *linearly* with the current level of conditional variance as opposed to the faster-than-linear growth case of the GARCH(1,1) model. To see one implication of this, consider the GARCH(1,1)  $\{\varepsilon_{G,t}\}$  and the SQGARCH(1,1)  $\{\varepsilon_{SQG,t}\}$  processes driven by the same standardized shock series  $\{z_t\}$  with kurtosis  $\kappa_z < \infty$ . To make the comparison of the two processes meaningful, suppose that they share the same values of  $\omega$  and  $\gamma$  and that their  $\alpha$ 's ( $\alpha_G$  for the GARCH and  $\alpha_{SQG}$  for the SQGARCH) are related by the following equation:

$$\alpha_{SQG} = \left\{ \frac{\omega(1+\gamma)}{1-\gamma^2 - (\kappa_z - 1)\alpha_G^2} \right\}^{1/2} \alpha_G \quad (3.7)$$

Then the two processes have the same unconditional variance and kurtosis, assuming that the conditions for their finiteness and constancy, including  $\gamma < 1$ , are met. Suppose furthermore that, at time  $t = 0$ , the conditional variances of both  $\varepsilon_{G,1}$  and  $\varepsilon_{SQG,1}$  are the same at  $h_1$ .

Note also that the term structures of the conditional variances are the same since

$$h_{i,t|0} = \omega \sum_{s=0}^{t-2} \gamma^s + \gamma^{t-1} h_1, \quad t \geq 2, \quad i = G \text{ or } SQG \quad (3.8)$$

where  $h_{i,t|0} := E_0[h_{i,t}] = E_0[\varepsilon_{i,t}^2]$ . However, the conditional variances of the multiperiod-ahead conditional variances,  $Var_0(h_{i,t})$ ,  $t \geq 2$ , and hence the multiperiod-ahead conditional kurtoses can be very different across the two processes depending on the current level  $h_t$  (shared by the two processes) of the one-period-ahead conditional variance. Roughly speaking, if  $h_1$  is higher than the threshold value that makes  $Var_0(h_{G,2})$  equal to  $Var_0(h_{SQG,2})$ , then the one-period-ahead conditional variance has a higher chance of getting extremely high or going back to the average level under the GARCH than under the SQGARCH while it is more likely to remain near the current high level under the SQGARCH than under the GARCH. Looking further into the future, the term structure of the variance of variance is given by:

$$Var_0(h_{i,t}) = \alpha_i^2 \sigma_\eta^2 \sum_{j=0}^{t-2} \gamma^{2j} E_0 \left[ h_{i,t-1-j}^{2p_i} \right], \quad i = G, \quad SQG \quad (3.9)$$

where  $p_G = 1$  (the GARCH) and  $p_{SQG} = 0.5$  (the SQGARCH), since

$$h_{i,t} - h_{i,t|0} = \alpha_i \sum_{j=0}^{t-2} \gamma^j h_{i,t-1-j}^{p_i} \eta_{t-1-j}. \quad (3.10)$$

Evidently, the conditional variance persistence parameter  $\gamma$  is also the determinant of the persistence in the conditional variance of multiperiod-ahead conditional variance. By repeated substitutions, we obtain the following expression for the SQGARCH case:

$$Var_0(h_{SQG,t}) = \frac{\alpha_{SQG}^2 \sigma_\eta^2}{1-\gamma} \left[ \sigma^2 \left\{ \frac{1-\gamma^{2t-2}}{1+\gamma} - \gamma^{t-2} (1-\gamma^{t-1}) \right\} + h_1 \gamma^{t-2} (1-\gamma^{t-1}) \right], \quad (3.11)$$

where  $\sigma_\varepsilon^2 := E[\varepsilon_{G,t}^2] = E[\varepsilon_{SQG,t}^2] = \omega(1-\gamma)^{-1}$ , which converges to the unconditional variance of conditional variance  $Var(h_{SQG,t}) = \alpha_{SQG}^2 \sigma_\eta^2 \sigma^2 (1-\gamma^2)^{-1}$  as  $t$  becomes large. For the GARCH(1,1) case, it is:

$$Var_0(h_{G,t}) = \alpha_G^2 \sigma_\eta^2 \sum_{j=0}^{t-2} \gamma^{2j} E_0[h_{G,t-1-j}^2], \quad (3.12)$$

which also converges a.s. to the same unconditional variance of conditional variance as in the SQGARCH(1,1) case.

For an illustration, consider an example with  $\sigma_\varepsilon^2 = 0.933$ ,  $\alpha_G = 0.052$ ,  $\gamma = 0.990$ ,  $\kappa_z = 4.702$ ,  $\omega = (1-\gamma)\sigma_\varepsilon^2$ , which previews the GARCH(1,1) estimation result for the 30-year sample of daily returns on the S&P 500 index, and  $\alpha_{SQG} = 0.070$  satisfying (3.7). Then, by construction, the two processes share the same unconditional variance  $\sigma_\varepsilon^2 = 0.933$  (if  $\{R_t\}$  is a series of daily percentage returns, then the annualized standard deviation is roughly 15% in this case) and unconditional kurtosis  $\kappa_\varepsilon = 9.193$ . Figure 1 depicts the term structures of the conditional standard deviations (instead of variances for intuitive ease) of multi-period-ahead conditional variances for this numerical example. The three term structure curves for either the GARCH or the SQGARCH case correspond to, from the top, the high ( $h_1 = 2\sigma_\varepsilon^2$ ), the medium ( $h_1 = \sigma_\varepsilon^2$ ), and the low ( $h_1 = 0.5\sigma_\varepsilon^2$ ) current conditional variance cases. Note that the curves for the medium current variance under the GARCH and the low current variance under the SQGARCH are not visually distinguishable.

Through the differences in the dynamics of higher conditional moments, the prices of an option, for example, are likely to be different depending on whether the underlying process is SQGARCH or GARCH even if the two processes share some common unconditional moments and current level of the conditional variance. Furthermore, if we are interested in one particular underlying process, at least one (and probably both) of the two models is misspecified, leading to different estimates of the parameter values (not necessarily satisfying the relationships such as (3.7) across two sets of parameter estimates) as well as different estimates of the current level and predictions of the future levels of the conditional variance. Most importantly, the differences in the conditional variance estimates and predictions are likely to be large after the market is hit by a series of volatility shocks of the same sign.

### 3.2 Various extensions of the GARCH and SQGARCH models

The GARCH(1,1) and SQGARCH(1,1) conditional variance equations can be unified under the following specification:

$$h_t = \omega + \gamma h_{t-1} + \alpha h_{t-1}^p \eta_{t-1} \quad (3.13)$$

where the volatility exponent  $p = 1$  or  $p = 1/2$  controls the way the variance of variance grows with the current level of variance. Once we extend the standard GARCH(1,1) model in this direction, it is imperative that we go further by allowing  $p$  to take any arbitrary positive value since there

is no known theoretical reason to restrict  $p$  to take either 1 or  $1/2$  in, e.g., financial applications. Equation (3.13) is a discrete-time, conditional variance analogue of the scalar interest rate diffusion model of Chan et al. [21]:

$$dv_t = \varphi(\theta - v_t) dt + \sigma_v v_t^p dW_t \quad (3.14)$$

which nests both the original CEV model of Cox [30] and the CIR model as special cases. More directly, it is an ARCH analogue of the continuous-time CEVSV model used by Engle and Lee [44] and Jones [65] in which  $\{v_t\}$ , the solution of (3.14), is the volatility process rather than the interest rate process. Adopting the terminology of the interest rate literature, we call the flexible volatility exponent case the *constant-elasticity-of-variance GARCH(1,1)* or *CEVGARCH(1,1)* model. When working with the continuous-time CEV model, the kind of analytical tractability afforded by  $p = 1/2$  is an attractive feature particularly in option or interest-rate product pricing applications. When using the discrete-time CEVGARCH model, however, we have much less reason to confine ourselves to SQGARCH ( $p = 1/2$ ) or GARCH ( $p = 1$ ).

We will call the case of  $p \leq 1$  the *sub-GARCH*. Going to the other side of the GARCH,  $p > 1$  makes (3.13) ill-defined as a model for conditional variance dynamics since no set of values for the constants  $(\omega, \gamma, \alpha)$  with  $\alpha > 0$  guarantees  $h_t > 0$  with probability one when  $p > 1$ . Although the support of  $\eta_{t-1}$  is bounded from below by  $-1$ , the shock term  $\alpha h_{t-1}^p \eta_{t-1}$ , with the coefficient  $\alpha h_{t-1}^p$  growing faster than linearly in  $h_{t-1}$ , may dominate  $\omega + \gamma h_{t-1}$  when the level of conditional variance  $h_{t-1}$  is large. In the next section, we propose one way of modifying the model with  $p > 1$  to ensure positivity of  $h_t$ . We will call the case of  $p \geq 1$  the *super-GARCH*. Fitting the variance-exponent-unconstrained version of the CEVGARCH model to the S&P 500 stock index and the deutsche mark(DM)/US dollar exchange rate return series, we find that the estimated  $p$  is significantly greater than one, which is in line with the findings of Jones [65]. We report the details of our empirical findings in Section 5.

Some of the alternatives to the standard GARCH volatility specification that have proven particularly successful in improving the fit to stock returns data are those designed to account for the empirical observation of Black [10], Christie [27], and others that the volatility of stock returns reacts asymmetrically to good and bad news. Various ways to capture this so-called “leverage effect” have been proposed in the literature (e.g., the exponential GARCH (EGARCH) of Nelson [74], the GJR-GARCH of Glosten, Jagannathan, and Runkle [52], and the threshold GARCH (TGARCH) of Zakořan [90]. See Hentschel [60] and Duan [39] for flexible models that nest many of the asymmetric variants). Using intraday price data of IBM stock, Hansen and Lunde [58] find evidence that asymmetric models also outperform the standard GARCH(1,1) in terms of the out-of-sample predictive accuracy. To introduce asymmetry to the CEVGARCH model, we may use, for example,  $\eta_t := z_t^2 + \phi z_t - 1$  with  $\phi \in \mathbb{R}$  in place of  $\eta_t = z_t^2 - 1$ . The addition of the term  $\phi z_t$  to the variance shock is in the spirit of the nonlinear asymmetric GARCH (NAGARCH) model of Engle and Ng [46]. The mean of the variance shock  $\eta_t$  is still zero, but now the impact of a negative

return shock on the variance is greater than a positive return shock of equal absolute magnitude when  $\phi < 0$ . As we will see,  $\phi \neq 0$  leads to  $dW_{1,t}dW_{2,t} = \rho dt$  with  $\rho \neq 0$  in the continuous-time limit where  $W_{1,t}$  and  $W_{2,t}$  are the Brownian motions driving the limit bivariate stochastic volatility diffusion process. As Heston [61] theoretically demonstrated, the asymmetry introduced by  $\phi \neq 0$  (or  $\rho \neq 0$ ) has important implications in option pricing applications. The empirical results of Nandi [71] employing Heston's SQSV option pricing formula suggest that allowing the correlation to be nonzero reduces errors in pricing and hedging of options, particularly the out-of-the-money ones; see also Andersen, Benzoni, and Lund [4] for a lucid illustration of the impacts of nonzero correlation  $\rho \neq 0$  on the prices of options written on such an underlying SQSV diffusion process, which manifests itself as Black-Scholes implied volatility "smirks" similar to the patterns given by the observed stock index option prices. For the remainder of this section, we regard  $\eta_t$  to be an arbitrary measurable function of  $z_t$  with  $E[\eta_t] = 0$ .

An obvious extension to the CEVGARCH(1,1) model is the CEVGARCH ( $P, Q$ ) model in which higher-order lags are included in the conditional variance equation:

$$h_t = \omega + \sum_{i=1}^P \gamma_i h_{t-i} + \sum_{i=1}^Q \alpha_i h_{t-i}^{p_i} \eta_{t-i} \quad (3.15)$$

where  $p_i$ 's may take different values in general.

In analogy to the continuous-time affine term-structure model of Brown and Schaefer [17] and Duffie and Kan [41], the following *affine GARCH* model generalizes the SQGARCH(1,1) model in a different direction:

$$h_t = \omega + \gamma h_{t-1} + \sqrt{\varrho + \alpha h_{t-1}} \eta_{t-1} \quad (3.16)$$

where  $\varrho \geq 0$ . The class of affine diffusion models and their extensions that include certain types of price jump components are known to preserve the analytical tractability of the CIR and SQSV models (see Dai and Singleton [32] and Duffie, Pan, and Singleton [42] for discussions on details of this class of models); hence their popularity both among continuous-time finance researchers and practitioners. In the affine GARCH model, the sensitivity of the conditional variance to variance shocks does not approach zero as the conditional variance level goes to zero. Again, this model may be further generalized to:

$$h_t = \omega + \gamma h_{t-1} + (\varrho + \alpha h_{t-1})^p \eta_{t-1} \quad (3.17)$$

with  $p > 0$ . Other functional forms for the variance shock coefficient may be considered.

Another possibility is the following specification for conditional variance dynamics, which is free from the problem of  $h_t$  turning negative even when  $p > 1$ .

$$\ln h_t = \ln h_{t-1} + \varphi \left( \frac{\theta}{h_{t-1}} - 1 \right) + \alpha h_{t-1}^{p-1} \eta_{t-1} \quad (3.18)$$

(3.18) is heuristically motivated as an ARCH-discretization to the log variance version of (3.14), obtained by applying Ito's formula and then dropping the term involving a quadratic variation. The second term in (3.18) gives rise to a mean-reverting property for  $\{\ln h_t\}$  if  $\varphi$  is positive. Simplifying (3.18) by replacing the first two terms with an affine function of  $\ln h_{t-1}$ , we obtain:

$$\ln h_t = \omega + \gamma \ln h_{t-1} + \alpha h_{t-1}^{p-1} \eta_{t-1} \quad (3.19)$$

which nets versions of the EGARCH(1,1) model as a special case ( $p = 1$ ).

In the rest of the paper, we focus on the flexible volatility-exponent CEVGARCH(1,1) model (3.13) and study its probabilistic properties in more detail, setting aside further investigation of the other extensions as a topic for future research.

## 4 Properties of the CEVGARCH(1,1) model

In this section, we investigate the probabilistic properties of the CEVGARCH(1,1) model in a more formal setting. First, assume throughout this section that  $\{z_t : t \in \mathbb{Z}\}$  is some given i.i.d. sequence of zero-mean, unit-variance random variables  $z_t \in \mathcal{F}_t, \forall t \in \mathbb{Z}$ . Define the notation  $\tilde{\mathcal{F}}_t := \sigma\{z_t, z_{t-1}, \dots\} \subset \mathcal{F}_t$ . Let  $\{\eta_t = \eta(z_t) : t \in \mathbb{Z}\}$  for all  $t \in \mathbb{N}$  be the conditional variance shock sequence where  $\eta$  is a measurable function such that  $\eta(z_t) > \eta_L$  a.s. for some constant  $\eta_L$  whose allowed range is to be defined below. Conventionally, the term ‘‘an ARCH-type process’’ (or a more specific name) is used to refer to  $\{R_t = \mu_t + \varepsilon_t : t \in \mathbb{Z}_+\}$  (or to  $\{\varepsilon_t : t \in \mathbb{Z}_+\}$ ) with  $\mu_t \in \mathcal{F}_{t-1}$  and  $\varepsilon_t = \sqrt{h_t} z_t$  as in (2.1) and (2.2) where  $\{h_t : t \in \mathbb{Z}_+\}$  with  $h_t \in \tilde{\mathcal{F}}_{t-1}$  (or more generally  $h_t \in \mathcal{F}_{t-1}$ ) is defined as a solution to some stochastic difference equation<sup>6</sup>. However, since we focus on  $\{h_t\}$  rather than  $\{\varepsilon_t\}$  in this section, we call  $\{h_t\}$  itself a GARCH process for convenience. We also assume that  $\{z_t\}$  is independent of the starting random variable  $h_0 \sim \pi_0$  where  $\pi_0$  is the initial probability distribution on  $(G, \mathcal{G})$ ,  $G := [h_L, \infty)$  and  $\mathcal{G} := \mathcal{B}(G)$ . Throughout, we denote the Borel  $\sigma$ -algebra on a set  $A$  w.r.t. the appropriate metric by  $\mathcal{B}(A)$ . Restrictions on  $h_L$  will be given later. Various properties of  $\{\varepsilon_t\}$  follow from those of  $\{h_t\}$  since we are assuming  $\{z_t\}$  to be i.i.d.; See Carrasco and Chen [20].

### 4.1 The definition of the CEVGARCH(1,1) model

Some probabilistic properties of the CEVGARCH(1,1) model are sharply different depending on whether the volatility exponent  $p$  is greater than, equal to, or smaller than one. We therefore consider the three cases separately. We shall call it the *sub-GARCH* (resp. *strictly sub-GARCH*) model if  $p \leq 1$  (resp.  $p < 1$ ), and the *super-GARCH* (resp. *strictly super-GARCH*) model if  $p \geq 1$

<sup>6</sup>In defining a time series, the doubly infinite time index set  $\mathbb{Z}$ , rather than  $\mathbb{Z}_+$ , is often used to allow for infinite backward substitution.

(resp.  $p > 1$ ). The conditions in the following formal definitions are for ensuring a.s. positivity. Further conditions will be added later for showing other properties.

**Definition 4.1 (The strictly sub-GARCH(1,1) process):** An  $\tilde{\mathcal{F}}_{t-1}$ -adapted time series  $\{h_t\}$  that satisfies any member of the following class of stochastic difference equations is called a strictly sub-GARCH(1,1) process:

$$h_t = \omega + \gamma h_{t-1} + \alpha h_{t-1}^p \eta_{t-1} \quad \text{for all } t \in \mathbb{N} \quad (4.1)$$

initialized by  $h_0$ , and the constants satisfy  $p \in (0, 1)$ ,  $\omega > 0$ ,  $\alpha > 0$ ,  $\gamma > 0$ ,  $h_L := \omega + \gamma^{-p/(1-p)} (-\alpha \eta_L p)^{1/(1-p)} (1 - p^{-1})$ , and  $\eta_L > -(1 - p)^{p-1} p^{-p} \gamma^p \omega^{1-p} \alpha^{-1}$ .

An important feature of  $\{h_t\}$  in Definition 4.1 is that it is a temporally homogeneous Markov chain, which is due to  $\{\eta_t\}$  being an i.i.d. sequence. Since  $\{h_t\}$  starts with an initial value in  $G$ , which is an absorbing set for  $\{h_t\}$ , i.e.,  $P(h_{t+1} \in G \mid h_t = h) = 1$  for all  $h \in G$ , we consider  $(G, \mathcal{G})$  as the state space. Since we also have  $P(h_{t+1} \in G \mid h_t = h) = 1$  for all  $h \in \mathbb{R}_+ \setminus G$ , this restriction entails little loss of generality. The minimum value of  $h_{t+1}$  for a given negative value of  $\eta_t$  is finite, unique, and increasing in  $\eta_t$  for the sub-GARCH model. The restriction on the support of  $\eta_t$  ensures that this minimum value is bounded away above zero; hence  $h_t \geq c > 0$  a.s. for some constant  $c$ .

The SQGARCH(1,1) ( $p = 1/2$ ) model is a special case of the sub-GARCH model. If we set  $\eta_t = z_t^2 - 1$ , then it reduces to the SQGARCH analogue of the standard GARCH(1,1). The a.s. positivity constraint becomes  $\alpha < 2\sqrt{\omega\gamma}$  ( $\alpha \leq 2\sqrt{\omega\gamma}$  if  $P(z_t = 0) = 0$ ). The sub-GARCH analogue of the NAGARCH(1,1) model is obtained by further setting  $\eta_t = z_t^2 + \phi z_t - 1$ .

**Definition 4.2 (The GARCH(1,1) process):** An  $\tilde{\mathcal{F}}_{t-1}$ -adapted time series  $\{h_t\}$  that satisfies any member of the following class of stochastic difference equations is called a GARCH(1,1) process:

$$h_t = \omega + \gamma h_{t-1} + \alpha h_{t-1} \eta_{t-1} \quad \text{for all } t \in \mathbb{N} \quad (4.2)$$

initialized by  $h_0$ , and the constants satisfy  $\omega > 0$ ,  $\alpha > 0$ ,  $\gamma > 0$ ,  $h_L > 0$ , and  $\eta_L \geq -\gamma\alpha^{-1}$ .

Note that the GARCH(1,1) defined above is more general than the standard GARCH(1,1) with  $\eta_t := z_t^2 - 1$ . This shock  $\eta_t$  in the standard GARCH satisfies the a.s. positivity condition since  $\eta_t \geq -1 > -1 - \beta\alpha^{-1} = -\gamma\alpha^{-1}$ . The NAGARCH(1,1) model is obtained as a special case by setting  $\eta_t = z_t^2 + \phi z_t - 1$ . Since our focus is on the volatility exponent  $p$ , it is convenient to call all CEVGARCH models with different specifications for  $\eta(\cdot)$  simply the GARCH model as long as their volatility exponent is one. We will also call all CEVGARCH models with  $p = 1/2$  the SQGARCH. This will keep the proliferation of new ARCH acronyms under control.

If  $p > 1$  and  $P(\eta_t < 0) > 0$ , there is no set of parameter values that ensures a.s. positivity of  $\{h_t\}$  defined by the stochastic difference equation (3.13) except for the uninteresting degenerate case of  $\alpha = 0$ , which we shall rule out in this paper. This is so because the shock coefficient  $\alpha h_t^p$  rises faster than linearly when  $p > 1$  so that when the current level of conditional variance  $h_t$  is

large and the shock  $\eta_t$  is negative, then  $\alpha h_t^p \eta_t$  may dominate the deterministic component  $\omega + \gamma h_t$  to make the next period conditional variance  $h_{t+1}$  negative. One way to avoid the conditional variance from taking a negative value with a positive probability is to restrict the growth of the shock coefficient on  $\eta_t$  in the following way.

**Definition 4.3 (The strictly super-GARCH(1,1) process):** An  $\tilde{\mathcal{F}}_{t-1}$ -adapted time series  $\{h_t\}$  that satisfies any member of the following class of stochastic difference equations is called a strictly super-GARCH(1,1) process:

$$h_t = \omega + \gamma h_{t-1} + \alpha H(h_{t-1}) \eta_{t-1} \text{ for all } t \in \mathbb{N} \quad (4.3)$$

initialized by  $h_0$ , where  $H(h) := \min\left(h^p, \frac{h_L - \omega - \gamma h}{\alpha \eta_L}\right)$ , and the constants satisfy  $p > 1$ ,  $\omega > h_L > 0$ ,  $\alpha > 0$ ,  $\gamma > 0$ , and  $\eta_L > -\infty$ .

The growth of the variance shock coefficient with the current level of conditional variance is reduced to be linear after a threshold is reached, but no upper bound on the impact of the whole variance shock term is imposed unlike in the case of the MARCH model. To be precise, the elasticity of variance of variance in the strictly super-GARCH case is only locally constant for values of  $h_{t-1}$  below the threshold. This damper has a side effect of preventing the conditional variance from running up extremely high in response to consecutively positive variance shocks. We emphasize, however, that the sole purpose in designing the linearization in (4.3) is to set the constant-elasticity domain of  $H(h)$  as large as possible while ensuring  $h_t > h_L$ . The true DGP may dictate a different relaxation of the constant-elasticity restriction.

Noting that  $H(h_{t-1}) = h_{t-1}^p$  always hold for the strictly sub-GARCH(1,1) and the GARCH(1,1) models as long as the lower bound on the conditional variance implied by the other constants is at least as high as  $h_L$ , we may use (4.3) in defining the CEVGARCH(1,1) model for an arbitrary  $p > 0$ . For this reason, in the sequel, we shall refer to the model defined by the difference equation (4.3) by the term *CEVGARCH(1,1)* regardless of the value of  $p$ . We may replace  $\min\left(h^p, \frac{h_L - \omega - \gamma h}{\alpha \eta_L}\right)$  with a smoothed version. Another alternative specification that ensures positivity of conditional variance ( $h_t \geq h_L > 0$ ) is:

$$h_t = \max\left(\omega + \gamma h_{t-1} + \alpha h_{t-1}^p \eta_{t-1}, h_L\right) \quad (4.4)$$

or a smoothed version of it.

## 4.2 Geometric ergodicity, existence of unconditional moments, and the mixing property

Many time series including ARCH-type processes are defined as a solution, adapted to a given filtration, of a system of nonlinear stochastic difference equations. It is not always easy to verify even the existence of an adapted, strictly stationary solution, let alone such properties as the finiteness of higher moments and the mixing property of the solution, by directly attempting to

obtain a solution through recursive backward substitution. The stochastic difference equation for the flexible volatility model proposed by Hentschel [60] is rather special in that it has a bilinear structure in terms of the Box-Cox transformation of the conditional standard deviation and the driving shock. Hentschel [60] showed that this structure is exactly the same as that of the conditional variance equation of the standard GARCH model so that Nelson’s derivation of the necessary and sufficient conditions for the strict stationarity and ergodicity of the solution of the standard GARCH(1,1) stochastic difference equation is also applicable to his model. His model nests the standard GARCH, EGARCH, NAGARCH, GJR-GARCH, TGARCH, the absolute value GARCH (TS-GARCH) of Taylor [83] and Schwert [80], and the asymmetric power GARCH (APARCH) of Ding, Engle and Granger [34].

However, the difficulty with the CEVGARCH model is that the volatility equation is a highly nonlinear function of the conditional variance unlike in the case of the Hentschel’s model or its generalization, the *augmented GARCH* model of Duan [39]. It is not easy to disentangle the Russian-matryoshka-doll structure that results when one tries to solve nonlinear stochastic difference equations such as that of the CEVGARCH model by recursive backward substitution. We therefore rely on the general state-space Markov chains approach. The use of this powerful approach in nonlinear time series analysis has long been advocated by, e.g., Tong [85], Chan [22], [24], Tong and Chan [25], [26], and its potential in the context of ARCH-type models was suggested by Bollerslev, Engle and Nelson [15]. Carrasco and Chen [20] applied it in a comprehensive manner to investigate the properties of various ARCH-type and other stochastic volatility models. Once we note that the CEVGARCH(1,1) process of Definitions 4.1-4.3 is a Markov chain on the state space  $(G, \mathcal{G})$  with temporarily homogenous transition probabilities, we may use the general state-space Markov chains approach<sup>7</sup>, in particular the so-called “Lyapunov-Foster drift conditions” checking methods. For the undefined Markov chains terminology used in this section and Appendix, see Meyn and Tweedie [69]. Nummelin [76] and Meyn and Tweedie [69] are two of the standard references for the modern general state-space Markov chains techniques. Chan [24] and Chan and Tong [26] provide concise and lucid introductions for time series analysts. We shall rely on the versions that assume the properties called *irreducibility* and *aperiodicity* (the general state-space counterparts to irreducibility and aperiodicity for countable state-space Markov chains; see Appendix for the definitions) since it is straightforward to verify that these properties hold for the CEVGARCH(1,1) model under the following conditions on the distribution of the shocks.

**Assumption 4.1:** *The distribution of  $\eta_t = \eta(z_t)$  has an absolutely continuous component w.r.t. the Lebesgue measure with a positive probability density function over  $(\underline{\eta}, \bar{\eta})$  for some constants*

---

<sup>7</sup>An obvious shortcoming of this approach is the requirement that the process be Markov, which in general precludes dependent driving shock sequences. However, the frontier of the literature on the Markov chains and related areas has been rapidly moving forward in recent years. For example, the stochastically recursive sequences approach allows for non-i.i.d. driving shocks and still delivers some results similar to those known for Markov chains (see Borovkov [18], Chapter 3, and the references therein).

$\eta_L \leq \underline{\eta} < 0 \leq \bar{\eta}$ , and  $E[\eta_t] = 0$ . Additionally,  $\gamma \leq 1$  holds if  $p < 1$  (the strictly sub-GARCH case), and  $\gamma + \alpha\underline{\eta} < 1$  holds if  $p = 1$  (the GARCH case).

The conditions of Assumption 4.1 are enough to ensure irreducibility and aperiodicity, for the proof of which it is possible to weaken the assumption  $\bar{\eta} > 0$ . For the GARCH case under Assumption 4.1, we may use  $h_L := \omega(1 - \gamma - \alpha\eta_L)^{-1} > \omega > 0$  as the lower bound for  $h_t$ .

Throughout the rest of the paper and Appendix, let  $X := \{X_t : t \in \mathbb{Z}_+\}$  be a Markov chain on state space  $(S, \mathcal{S})$  where  $S$  is some set and  $\mathcal{S}$  is a countably generated  $\sigma$ -algebra of subsets of  $S$ , with time-homogenous transition densities  $P^t(x, A) := P(X_t \in A \mid X_0 = x)$ ,  $x \in A$ ,  $A \in \mathcal{S}$ , and  $P(x, A) := P^1(x, A)$ . Let  $\|\lambda\|$  be the total variation norm for a bounded signed measure  $\lambda$  on  $(S, \mathcal{S})$ , i.e.,  $\|\lambda\| := \sup_{A \in \mathcal{S}} \lambda(A) - \inf_{A \in \mathcal{S}} \lambda(A)$ .

**Definition 4.4 (Harris ergodicity and geometric ergodicity):**  *$X$  is said to be Harris ergodic if there exists a probability measure  $\pi$  on  $(S, \mathcal{S})$  such that  $\lim_{t \rightarrow \infty} \|P^t(x, \cdot) - \pi(\cdot)\| = 0$  for all  $x \in S$ . If  $X$  is Harris ergodic and there exist a nonnegative measurable function  $M(x)$  with  $\int_S M(x) \pi(dx) < \infty$  and a constant  $c \in (0, 1)$  such that  $\|P^t(x, \cdot) - \pi(\cdot)\| \leq M(x) c^t$  for all  $t \in \mathbb{Z}_+$ , then  $X$  is said to be geometrically ergodic.*

Note that  $\|P^t(x, \cdot) - \pi(\cdot)\| = 2 \sup_{A \in \mathcal{S}} |P^t(x, A) - \pi(A)|$  holds for all  $x \in S$ . Harris ergodicity implies that there exists a unique stationary probability measure  $\pi$ , and that the distribution of  $X_t$  conditional on any starting value converges not only weakly but in total variation norm to the stationary one, with the rate of convergence being exponential in the case of geometric ergodicity. If  $X$  is started with the stationary probability measure  $\pi$ , then the unconditional distribution of  $X_t$  is  $\pi$  for all  $t \in \mathbb{Z}_+$  since  $\pi(A) = \int_S P(x, A) \pi(dx)$  holds for all  $A \in \mathcal{S}$  by definition, and clearly  $X$  is strictly stationary and can be extended to a strictly stationary process starting arbitrarily far in the past. Harris ergodicity also implies that  $\lim_{t \rightarrow \infty} \|\int_S P^t(x, \cdot) \lambda(dx) - \pi(\cdot)\| = 0$  for any probability measure  $\lambda$  (see, e.g., Nummelin [76, p.114]) so that the unconditional distribution of  $X_t$  converges to the stationary one regardless of the probability measure that starts the process. Drawing on Davydov [33, Proposition 1], Carrasco and Chen [20] noted that  $\beta(t) = \int_S \|P^t(x, \cdot) - \pi(\cdot)\| \pi(dx)$ , where  $\beta(t)$  is the  $\beta$ -mixing coefficient, holds for  $X$  initialized with  $\pi$  on the state space satisfying some technical conditions<sup>8</sup>, so that  $X$  being geometrically ergodic implies that  $X$  initialized with  $\pi$  is  $\beta$ -mixing (hence strongly mixing) with an exponential decay rate since  $M(x)$  is  $\pi$ -integrable. Below, we derive sufficient sets of additional conditions for the geometric ergodicity of the CEVGARCH process and for the finiteness of unconditional moments of the strictly stationary CEVGARCH process. In the next subsection, we provide a weaker set of conditions for strict stationarity. When  $X$  is strictly stationary with the stationary probability measure  $\pi$ , we use the

<sup>8</sup>These conditions are satisfied by  $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$  where  $N \in \mathbb{N}$  or by  $(G, \mathcal{G})$  for the CEVGARCH model defined in the previous subsection. Under the same conditions on the state space, Harris ergodicity is sufficient for  $X$  to be  $\beta$ -mixing (hence strongly mixing as well) since Harris ergodicity implies that  $X$  is aperiodic and ‘‘positive Harris recurrent’’ (see, e.g., Nummelin [76, p.114, Proposition 6.3]), which in turn is sufficient for  $X$  to be  $\beta$ -mixing by Corollary in Davydov [33].

notation  $E_\pi [g(X_t)] = \int_S g(x) \pi(dx)$  where  $g : S \rightarrow \mathbb{R}$  is a measurable function.

**Proposition 4.1:** *Suppose that the conditions in Assumption 4.1 are satisfied and that  $\gamma < 1$ , then the CEVGARCH process  $\{h_t\}$  is geometrically ergodic,  $E_\pi [h_t] < \infty$ , and there exist constants  $c_1 < 1$ ,  $c_2$ , and  $c_3$  such that  $\|P^t(h, \cdot) - \pi(\cdot)\| \leq c_1^t (c_2 + c_3 h)$  for all  $t$ . Furthermore,  $E_\pi [h_t^q] < \infty$  for  $q \in \mathbb{R}$  if any of the following conditions (i)-(iii) as well as Assumption 4.1 are satisfied:*

(i) *The strictly sub-GARCH(1,1) case ( $0 < p < 1$ ):  $\gamma < 1$  and  $E[|\eta_t|^q] < \infty$ .*

(ii) *The GARCH(1,1) case ( $p = 1$ ):  $E[(\gamma + \alpha\eta_t)^q] < 1$ .*

(iii) *The strictly super-GARCH(1,1) case ( $p > 1$ ):  $\gamma^q E[(1 - \eta_L^{-1}\eta_t)^q] < 1$ .*

Proof: See Appendix.

Carrasco and Chen [20] exploited the fact that many of the existing ARCH-type and stochastic volatility models are special cases of the so-called generalized polynomial random coefficient vector autoregressive model and used general state-space Markov chains tools specialized for this model. Since the CEVGARCH conditional variance equation is highly nonlinear, more specifically only one of the three  $h_t$ 's appearing in both sides of the equation is raised to the  $p$ th power and even a switching behavior is involved in the case of the strictly super-GARCH model, our CEVGARCH model is not a special case of the generalized polynomial random coefficient vector autoregressive model except when  $p = 1$ . However, Proposition 4.1 shows that geometric ergodicity and its accompanying properties similarly obtain under some conditions, exemplifying the usefulness of the general state-space Markov chains techniques for time series defined by nonlinear stochastic difference equations. The properties such as  $\beta$ -mixing may be useful for investigating further probabilistic properties of the CEVGARCH model and the asymptotic behavior of statistical estimators.

Our condition (ii) in  $E[(\gamma + \alpha\eta_t)^q] < 1$  for the GARCH(1,1) model when applied to the standard GARCH(1,1) or the NAGARCH(1,1) with an integer  $q \geq 1$  is the same as the condition (a) in Corollary 6 of Carrasco and Chen [20]. The sufficient condition for the existence of finite moments of the GARCH(1,1) model above corresponds to the necessary and sufficient condition derived by Nelson [73] for the standard GARCH(1,1) case. Note that  $\gamma < 1$  is implicit by Jensen's inequality in the conditions (ii) and (iii) of Proposition 4.1 for the GARCH and the strictly super-GARCH cases as well if  $q > 1$ . On the other hand, if  $q \in (0, 1)$ , then  $\gamma < 1$  is not implied by either (ii) or (iii).

### 4.3 Variance-of-variance-induced strict stationarity

Conley et al. [29] introduced the notion of "volatility-induced stationarity" in the context of the scalar diffusion models for the short-term interest rate. They analyzed the nonlinear-drift CEV stochastic differential equation, which nests the linear-drift case (3.14), where  $\{v_t\}$  is interpreted as the interest rate rather than volatility, and showed that a high volatility elasticity may induce a unique weak solution that is stationary even when the drift is positive for large levels of the interest rate. In this subsection, we show using the Markov chains approach that the CEVGARCH

conditional variance equation, which is a discrete-time ARCH analogue to (3.14), exhibits similar properties. Since the LHS variable is itself the conditional variance in (4.1)-(4.3), it would be less confusing to call volatility-induced stationarity in the CEVGARCH context “variance-of-variance-induced stationarity.” In this paper, we define *variance-of-variance-induced stationary* to mean strict stationarity of an adapted solution to the CEVGARCH difference equation with  $\gamma \geq 1$ <sup>9</sup>. Strict stationarity of the adapted solution to the integrated and mildly explosive GARCH(1,1) conditional variance equation demonstrated by Nelson [73]<sup>10</sup> may be given a new interpretation of discrete-time variance-of-variance-induced stationarity.

To understand this phenomenon intuitively, consider an explosive AR(1) process defined by  $X_t = \omega + \gamma X_{t-1} + \varepsilon_t$  where  $\{\varepsilon_t\}$  is an i.i.d. shock sequence and  $|\gamma| > 1$ , starting at  $t = 0$  with some initial value. It is repulsed further and further away from “home” since the “antigravitational force” of the explosive AR coefficient is incessantly exerted on it and its ability to move around randomly in search of home is fixed so that it goes arbitrarily far away eventually. On the other hand, the CEVGARCH(1,1) process  $\{h_t\}$  has an “automatically controlled drive” that makes it more volatile when it is pushed far away from home and needs to randomly move around more in search of home and slows it down when it is close to home. This automatic drive fully counters the antigravitational forces of the mildly explosive autoregressive coefficient, not too far above one, when  $p \geq 1$ , but does not fully do so when  $p < 1$ . This crude intuition is made precise and complete in Corollary 4.1 below. First, we introduce Proposition 4.2 as a tool for analyzing strict stationarity of real-valued time series defined as a solution to a highly nonlinear stochastic difference equation.

For the following proposition, let  $X$  satisfy the stochastic difference equation  $X_{t+1} = \mu_X(X_t) + \sigma_X(X_t, e_{t+1})$  initialized by  $X_0 \sim \pi_0$ , where  $S \subset [x_L, \infty)$  with some constant  $x_L > 0$ ,  $\mathcal{S} = \mathcal{B}(S)$ ,  $\{e_t\}$  is an i.i.d. sequence of random variables independent of  $\pi_0$ ,  $\mu_X : \mathbb{R} \rightarrow \mathbb{R}$  and  $\sigma_X : (0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}$  with  $E[\sigma_X(x, e_{t+1})] = 0 \forall x \in S$  are continuous functions, and  $\pi_0$  is some initial probability measure on  $(S, \mathcal{S})$ . An implicit assumption is that  $\mu_X(\cdot)$ ,  $\sigma_X(\cdot, \cdot)$ , and  $\{e_t\}$  are such as to ensure  $\mu_X(X_t) + \sigma_X(X_t, e_{t+1}) \geq x_L$  a.s. for all  $t$ . As we show below, the CEVGARCH(1,1) model has these properties.

**Proposition 4.2:** *Let  $X$  be irreducible and aperiodic. If  $E \left[ \ln \left( \frac{\mu_X(x)}{x} + \frac{\sigma_X(x, e_{t+1})}{x} \right) \right] < 0$  holds for all  $x$  large enough, then  $X$  is Harris ergodic.*

Proof: See Appendix.

Some of the conditions of Proposition 4.2 may be relaxed, e.g., by using the recent results of

---

<sup>9</sup>This definition is not as close a discrete-time analogue of the definition of Conley et. al [29] as possible. It would be more in the spirit of Conley et al. [29] to call all cases of strictly super-GARCH stationarity variance-of-variance-induced.

<sup>10</sup>As Nelson [73] showed,  $E[\ln(\beta + \alpha z_t^2)] < 0$  is necessary as well as sufficient for strict stationarity of the standard GARCH(1,1) model under some conditions. Also, it is of historical interest to note that Quinn’s [79] proof of a similar result for a simple bilinear model of Tong [84] is essentially the same as that of Nelson [73].

Fonseca and Tweedie [49] and Tweedie [87], particularly if we only wish to show the existence of a stationary probability measure, but for our discussion of the CEVGARCH(1,1) model Proposition 4.2 suffices.

By setting  $X_t = h_t$ ,  $\mu_X(h_t) = \omega + \gamma h_t$ ,  $\sigma_X(h_t, e_{t+1}) = \alpha h_t^p e_{t+1}$ , and  $e_{t+1} = \eta_t$ , we obtain the sub-GARCH(1,1) model. For the strictly sub-GARCH(1,1) model, we have  $\lim_{h \rightarrow \infty} E[\ln(\omega h^{-1} + \gamma + \alpha h^{p-1} \eta_t)] \rightarrow \ln \gamma$  as  $h \rightarrow \infty$  by the bounded convergence theorem and a truncation argument (see Durrett [40, p.17, Theorem 3.8]) since  $\omega h^{-1} + \gamma + \alpha h^{p-1} \eta_t \rightarrow \gamma$  a.s. Hence, the strictly sub-GARCH(1,1) model with  $\gamma < 1$  admits a strictly stationary probability measure. However, for the case of the strictly sub-GARCH(1,1) model with  $\gamma \geq 1$ , i.e.,  $\ln \gamma \geq 0$ , we may not use Proposition 4.2 since the variance shock term does not grow fast enough relative to the deterministic part as the level of current conditional variance increases. For the GARCH(1,1) case, we similarly have  $\lim_{h \rightarrow \infty} E[\ln(\omega h^{-1} + \gamma + \alpha h^{p-1} \eta_t)] = E[\ln(\gamma + \alpha \eta_t)]$  so that Nelson's [73] condition  $E[\ln(\gamma + \alpha \eta_t)] < 0$  that does not rule out  $\gamma > 1$  is obtained. For the strictly super-GARCH case,  $\lim_{h \rightarrow \infty} E[\ln(\omega h^{-1} + \gamma + \alpha \min(h_t^{p-1}, \frac{h_L h^{-1} - \omega h^{-1} + \gamma}{-\alpha \eta_L}) \eta_t)] = \ln \gamma + E[\ln(1 - \frac{\eta_t}{\eta_L})]$ . Noting that  $E[\ln(1 - \frac{\eta_t}{\eta_L})] < \ln 1 = 0$  by Jensen's inequality, we see that the super-GARCH model exhibits the property of variance-of-variance-induced strict stationarity if  $1 \leq \gamma < \exp(-E[\ln(1 - \frac{\eta_t}{\eta_L})])$ . In the case of the continuous-time CEV diffusion with an affine drift,  $p > 1$  is sufficient for stationarity. The linearization of the sensitivity growth for positivity is responsible for the presence of the upper bound  $\exp(-E[\ln(1 - \frac{\eta_t}{\eta_L})])$  on  $\gamma$  for the strict stationarity in the strictly super-GARCH case. These results are summarized in the following corollary as a CEVGARCH(1,1) generalization of the sufficiency part of Nelson's strict stationarity condition for the standard GARCH(1,1) model.

**Corollary 4.1:** *The CEVGARCH process is Harris ergodic if the conditions of Assumption 4.1 and any of the following are satisfied:*

- (i) *The strictly sub-GARCH(1,1) case ( $0 < p < 1$ ):  $\gamma < 1$ ;*
- (ii) *The GARCH(1,1) case ( $p = 1$ ):  $E[\ln(\gamma + \alpha \eta_t)] < 0$ ;*
- (iii) *The strictly super GARCH(1,1) case ( $p > 1$ ):  $\gamma < \exp(-E[\ln(1 - \frac{\eta_t}{\eta_L})])$ .*

One of the implications of Harris ergodicity is that the CEVGARCH conditional variance equation has an adapted, strictly stationary solution. The sufficient conditions for Harris ergodicity obtained above are a consequence of Jensen's inequality and the conditional variance having a lower bound above zero. Notably, the inequality condition for the GARCH(1,1) case above when applied to the standard GARCH(1,1) model is identical to the necessary and sufficient condition derived by Nelson [73] through a direct solution approach. However, our new proof using the Markov chains approach makes the role of the variance of variance more transparent. Heuristically applying Taylor's approximation, we have:

$$E[\ln(\gamma + \alpha \eta_t)] \approx \ln \gamma - \frac{1}{2} \gamma^{-2} \alpha^2 \sigma_\eta^2$$

if  $\sigma_\eta^2 := \text{Var}(\eta_t) < \infty$ . The variance of the shock term  $\sigma_\eta^2$  counteract the force of  $\gamma \geq 1$ .

#### 4.4 Weak convergence of the CEVGARCH(1,1) process to the continuous-time CEVSV process

In this subsection, we justify the CEVGARCH(1,1) model as a discrete-time analogue of the continuous-time CEVSV diffusion by showing the weak convergence<sup>11</sup> of the former to the latter by the moment matching procedure of Nelson [72].

First, consider the scalar stochastic differential equation (3.14) with some initial distribution  $\zeta_{v,0}$ ,  $\zeta_{v,0}(\{v_0 > 0\}) = 1$ , where the constants  $\varphi$ ,  $\theta$ ,  $\sigma_v$ ,  $p$  are such as to ensure the existence of a unique weak solution  $\{v_t : t \geq 0\}$  with  $P(v_t \in (0, \infty) \text{ for all } t \in [0, \infty)) = 1$  (See Ait-Sahalia [1], Appendix 1 for a discussion of such conditions). Since  $v_t > 0$  for all  $t$ , the equation (3.13) may equivalently be written as:

$$dv_t = \varphi(\theta - v_t)dt + \sigma_v |v_t|^p dW_{2,t}$$

in which nondifferentiability of  $|v|$  at  $v = 0$  does not pose any problem. Next, we define a sequence of Markov chains by the following stochastic difference equation of the CEVGARCH form:

$$v_{\Delta, (k+1)\Delta} = \omega_{\Delta} + \gamma_{\Delta} v_{\Delta, k\Delta} + \alpha_{\Delta} |v_{\Delta, k\Delta}|^p \eta_{\Delta, k}, \quad k \in \mathbb{Z}_+ \quad (4.5)$$

with some initial distribution that converges to  $\zeta_{v,0}$  where, for each  $\Delta$ ,  $\{\eta_{\Delta, k}\}$  is an i.i.d. sequence of zero-mean random variables with variance  $\sigma_{\eta}^2 > 0$  and  $E\left[|\eta_{\Delta, k}|^{2+\delta}\right] = c < \infty$  ( $c$  does not depend on  $\Delta$ ) for some  $\delta > 0$ . For example,  $\eta_{\Delta, k} = z_{\Delta, k}^2 + \phi z_{\Delta, k} - 1$ , where  $\{z_{\Delta, k}\}$  is an i.i.d. sequence of standard normal random variables for each  $\Delta$ , satisfies  $E\left[|\eta_{\Delta, k}|^{2+\delta}\right] < \infty$ . Here, since the LHS of (3.13) for finite  $\Delta$ 's turning negative with a positive probability is not our concern, we simply use the absolute value operator before raising  $v_{\Delta, k\Delta}$  to the  $p$ th power on the RHS instead of a sensitivity growth damper of the type introduced earlier for the super-GARCH model. Any sequence of constants  $\{\omega_{\Delta}, \beta_{\Delta}, \alpha_{\Delta}\}$  satisfying  $\lim_{\Delta \downarrow 0} \Delta^{-1} \omega_{\Delta} = \omega$ ,  $\lim_{\Delta \downarrow 0} \Delta^{-1} (1 - \gamma_{\Delta}) = \varphi$ , and  $\lim_{\Delta \downarrow 0} \Delta^{-1} \alpha_{\Delta}^2 = \sigma_v^2 / \sigma_{\eta}^2$ , is sufficient for our purpose, but for simplicity let us adopt the following particular reparameterization scheme:

$$\begin{aligned} \omega_{\Delta} &= \omega \Delta, \\ \gamma_{\Delta} &= 1 - \varphi \Delta, \\ \alpha_{\Delta} &= \sigma_v \sigma_{\eta}^{-1} \Delta^{1/2}, \end{aligned} \quad (4.6)$$

which is essentially the same as that used as an example by Nelson [72] for a sequence of GARCH(1,1)-M processes. Substituting (4.6) into (4.5), we obtain:

$$v_{\Delta, (k+1)\Delta} - v_{\Delta, k\Delta} = (\omega - \varphi v_{\Delta, k\Delta}) \Delta + \sigma_v |v_{\Delta, k\Delta}|^p \Delta^{1/2} \xi_{\Delta, k}$$

---

<sup>11</sup>Here, as in Nelson [72], the discrete-time processes are turned into continuous-time processes whose sample paths are elements in  $D$  where  $D$  is the space of functions from  $[0, \infty)$  into  $\mathbb{R}^N$ ,  $N = 1$  or  $2$ , that are right continuous with left limits, equipped with the Skorohod metric. So weak convergence refers to weak convergence of a sequence of probability measures on  $(D, \mathcal{B}(D))$  generating the sample paths of the processes.

where  $\{\xi_{\Delta,k}\}$ ,  $\xi_{\Delta,k} := \sigma_\eta^{-1} \eta_{\Delta,k}$ , is the standardized shock sequence.

Then, the first two moments per unit time of  $v_{\Delta,(k+1)\Delta} - v_{\Delta,k\Delta}$  conditional on  $v_{\Delta,k\Delta} = v \in (0, \infty)$  are:

$$\Delta^{-1} E [v_{\Delta,(k+1)\Delta} - v_{\Delta,k\Delta} \mid v_{\Delta,k\Delta} = v] = \omega - \varphi v$$

and

$$\Delta^{-1} E \left[ (v_{\Delta,(k+1)\Delta} - v_{\Delta,k\Delta})^2 \mid v_{\Delta,k\Delta} = v \right] = \sigma_v^2 |v|^{2p} + O(\Delta),$$

where  $O(\Delta)$  vanishes uniformly on compact sets. We also have  $\Delta^{-1} E \left[ (v_{\Delta,(k+1)\Delta} - v_{\Delta,k\Delta})^{2+\delta} \mid v_{\Delta,k\Delta} = v \right] = 0$  vanishing uniformly on compact sets as  $\Delta \downarrow 0$ . Next, turn each  $\{v_{\Delta,k\Delta}\}$  into the continuous time process  $\{v_{\Delta,t}; t \geq 0\}$ , indexed by  $\Delta$ , whose sample paths are step functions with jumps at  $t \in \Delta\mathbb{N}$ , by defining  $v_{\Delta,t} := v_{\Delta,k\Delta}$  for  $t \in [k\Delta, (k+1)\Delta)$ . Then, Assumptions 2-5 of Theorem 2.2 in Nelson [72] are satisfied so that the sequence of processes  $\{v_{\Delta,t}; t \geq 0\}$  converges weakly to the process  $\{v_t; t \geq 0\}$ . Note that, although  $v_{\Delta,k\Delta}$  may become zero or negative depending on the values of the constants,  $\tau_\Delta := \inf \{t \geq 0 : v_{\Delta,t} \leq 0\} \rightarrow \infty$  a.s. as  $\Delta \downarrow 0$ .

It is possible to turn (4.5) into a system of equations by adding

$$y_{\Delta,k\Delta} - y_{\Delta,(k-1)\Delta} = \mu_\Delta (y_{\Delta,(k-1)\Delta}, v_{k\Delta}) \Delta + |v_{\Delta,k\Delta}|^{1/2} z_{\Delta,k} \Delta^{1/2}$$

where  $\{\mu_\Delta(\cdot, \cdot)\}$  is a sequence of continuous functions  $\mathbb{R}^2 \rightarrow \mathbb{R}$  satisfying conditions implied by Assumption 5 of Nelson [72] and, for each  $\Delta$ ,  $\{z_{\Delta,k}\}$  is such that  $\{z_{\Delta,k}, \xi_{\Delta,k}\}$  is an i.i.d. sequence of zero-mean random vectors with  $E[z_{\Delta,k} \xi_{\Delta,k}] = \rho$ ,  $E[|z_{\Delta,k}|^{2+\delta}] = c' < \infty$  ( $c'$  does not depend on  $\Delta$ ) for some  $\delta > 0$  and independent of the initial joint distribution of  $(y_{\Delta,0}, v_{\Delta,0})$ . Then,  $|v_{\Delta,k\Delta}|$  is the conditional variance of  $y_{\Delta,k\Delta} - y_{\Delta,(k-1)\Delta}$  per unit time. Suppose that  $(y_{\Delta,0}, v_{\Delta,0})$  converges in distribution to  $(y_0, v_0)$  with the joint distribution  $\zeta_0$  and the marginal distribution of  $v_0$  being  $\zeta_{0,v}$ . We can then show by a straightforward application of Nelson's [72] Theorem 2.2 again that the sequence of bivariate continuous-time processes  $\{y_{\Delta,t}, v_{\Delta,t}; t \geq 0\}$  where  $y_{\Delta,t} := y_{\Delta,k\Delta}$  for  $t \in [k\Delta, (k+1)\Delta)$  converges weakly to a bivariate diffusion process defined as a unique weak solution (if one exists) to

$$dy_t = \mu(y_t, v_t) dt + \sqrt{v_t} dW_{1,t}$$

and (3.14) with the initial distribution  $\zeta_0$  where  $\mu(\cdot, \cdot)$  is the limit of  $\mu_\Delta(\cdot, \cdot)$  in the sense of Assumption 2 in Nelson [72] and  $W_1$  is some Brownian with  $dW_1 dW_2 = \rho dt$ .

As an example, suppose that  $\{y_{\Delta,t} := \ln Y_{\Delta,t}\}$  is a sequence of log asset return processes and specify the conditional expected return per unit time  $\mu_\Delta(y, v) := r_f + (r_v - \frac{1}{2})v$  in the fashion of the GARCH-M model of Engle, Lilien, and Robins [45] where the constants  $r_f$  and  $r_v$  are respectively the risk free rate of return per unit time and the risk premium per unit conditional variance and per unit time. In derivative pricing applications, the risk premium specification affects the risk-neutralized version of the conditional variance process in a crucial way (see Duan [37]). The Jensen's inequality adjustment term  $-v/2$  placed in the drift makes the conditional

expectation of the gross return  $Y_{\Delta,(k+1)\Delta}/Y_{\Delta,k\Delta} = \exp(r_f + r_v v_{\Delta,(k+1)k})$ . Suppose further that  $p = 1/2$  (the SQGARCH(1,1) specification) and  $\eta_{\Delta,k} := z_{\Delta,k}^2 + \phi z_{\Delta,k} - 1$  with  $E[z_{\Delta,k}^3] = 0$ , then the limit diffusion process satisfies:

$$dy_t = \left[ r_f + \left( r_v - \frac{1}{2} \right) v_t \right] dt + \sqrt{v_t} dW_{1,t} \quad (4.7)$$

$$dv_t = \varphi(\theta - v_t) dt + \sigma_v \sqrt{v_t} dW_{2,t} \quad (4.8)$$

with  $\rho := \phi/\sqrt{\kappa_z - 1 + \phi^2}$ . This formula for  $\rho$  applies to the CEVGARCH/CEVSV diffusion in general as long as the specification  $\eta_{\Delta,k} = z_{\Delta,k}^2 + \phi z_{\Delta,k} - 1$  is kept, and is  $\phi/\sqrt{2 + \phi^2}$  under normality. One possible application of this result is option pricing: First, estimate the discrete-time SQGARCH model for the underlying asset price process, turn it into a “locally” risk-neutralized version by Duan’s [37] GARCH option pricing method and plug the resulting parameter values into Heston’s [61] closed-form option pricing formula. The advantages are that the estimation is performed using only discrete-time observations of the underlying asset price process so that the resulting option prices may be used to judge whether the currently observed market prices are too high or too low in contrast to the approaches that estimate the parameters of the model by fitting the observations of current option prices, and that, based on discretely-sampled time series data only, the SQGARCH model is easier to estimate than the SQSV diffusion model.

We note, however, that ARCH-type models in which the volatility process is driven only by the volatility shock sequence  $\{\xi_{\Delta,k}\}$  have a single distinct source of uncertainty in the sense that  $\xi_{\Delta,k}$  is measurable with respect to the  $\sigma$ -algebra generated by  $z_{\Delta,k}$ . It may well be that this property is built into the likelihood function of an ARCH-type model in such a way as to cause the naively constructed plug-in estimator for the diffusion based on the maximum likelihood estimator of the ARCH-type model to have a bias that does not vanish even as the time interval between observations shrinks<sup>12</sup> if the true process is a bivariate SV diffusion. Using Duan’s local risk-neutralization procedure for derivatives pricing and hedging has another potential drawback of the same origin: An implicit assumption of zero premium on the part of volatility risk  $W_2$  orthogonal to the first source of risk  $W_1$  (See Duan [38]).

An important question related to weak convergence, which we do not address in this paper, is how the continuous-time CEVSV or high-frequency CEVGARCH processes aggregate over time (see Drost and Nijman [35] and Drost and Werker [36]). Of particular interest is the effect of temporal aggregation on the variance of variance dynamics.

---

<sup>12</sup>Based on theoretical results for a closely related issue, Wang [88] warns against the plug-in procedure.

## 5 Empirical examples

As empirical examples, we apply the CEVGARCH model to the daily returns on two financial time series: the Standard and Poor's 500 composite index (S&P 500) and the DM/US dollar exchange rate. The stock market dataset spans the three decades from the end of 1969 through the end of 1999 while the exchange rate dataset is of the decade from the end of 1989 through the end of 1999. The source of all data is Datastream. The daily returns are defined as log differences of the daily index or exchange rate values, multiplied by one hundred so that they are measured in percent. We do not attempt to estimate the conditional mean return process beyond including a constant term for the mean returns.

In all of the estimated models, we specify the shock term for the conditional variance by  $\eta_t = z_t^2 + \phi z_t - 1$  in light of the substantial evidence of volatility asymmetry in equity returns documented in the empirical finance literature<sup>13</sup>. We use the maximum-likelihood (ML) estimation procedure with the Student's  $t$ -distribution as the specification for the conditional distribution of the standardized return shock  $z_t$ . The use of the Student's  $t$ -distribution in the GARCH context was first proposed by Bollerslev [13] to fit the conditional excess kurtosis that appears to remain in financial time series even after the GARCH-induced fat-tail effect is accounted for. Our choice follows the finding of Bollerslev, Engle, and Nelson [15] that the generalized  $t$ -distribution for  $z_t$  when applied to several daily US stock returns series including the S&P 500 index returns delivers a marked improvement in fit over the generalized error distribution nested in it but not over the Student's  $t$ -distribution also nested in it. Among others, Hamilton and Susmel [57] and Brenner, Harjes and Kroner [16] also used the Student's  $t$ -distribution. See, however, the finding of Hansen and Lunde [58] that ARCH-type models with the Student's  $t$  specification do not always outperform those with the Gaussian specification in out-of-sample predictions of the DM/US dollar and IBM stock return volatilities. The models estimated with the Gaussian shock specification are strongly outperformed in fitting the data by their  $t$ -distribution counterparts and the estimates of the parameters common to both are similar so that the Gaussian estimation results are not reported.

Since our primary motivation for introducing the CEVGARCH model is the recent debate about the empirical validity of the SQSV specification, which has been the baseline model in the continuous-time finance literature, we focus on the first-order (1,1) models and estimate both the CEVGARCH models with the restriction  $p = 1/2$  (the SQGARCH) as well as  $p = 1$  (the GARCH, the baseline model in the discrete-time literature) and without such restrictions on the volatility exponent. We use the version of the CEVGARCH defined by (4.3) that forces the growth rate to be linear when the conditional variance reaches a certain level. As noted previously, it nests the sub-GARCH model ( $p \leq 1$ ) without such a damper on the conditional variance growth when

---

<sup>13</sup>As mentioned earlier, adhering to the standard nomenclature of the ARCH literature, the CEVGARCH with  $p = 1$  and this specification for the volatility shock  $\eta_t$  should be called the NAGARCH.

we allow the exponent to take values less than or equal to one as well. The lower bound on the conditional variance  $h_L$  in (4.3) is set to  $10^{-8}$ .

## 5.1 Standard and Poor's 500 Composite Index

### 5.1.1 The overall fit and the volatility exponent estimates

The estimation results for the S&P 500 returns are summarized in Tables 1. The most striking feature is the estimate for the exponent  $p$  of the unconstrained CEVGARCH(1,1) model being 1.71, which is more than seven standard errors above one. The likelihood ratio (LR) test statistics for the SQGARCH restriction  $p = 1/2$  and the GARCH restriction  $p = 1$  are respectively 132.49 and 70.12 with the associated asymptotic  $p$ -values being approximately zero (We may alternatively use the  $t$ -ratio test since there is only one linear restriction involving one parameter here). In terms of the Akaike information criteria (AIC), and the Schwarz-Bayes information criteria (BIC), the unconstrained CEVGARCH provides the best fit while the GARCH fits the data better than the SQGARCH does. (The GARCH and the SQGARCH have the same number of estimated parameters so that the penalties for the model complexity are the same and do not matter for the order of these two.) These results are not surprising given that the SQGARCH's volatility exponent  $p = 1/2$  is farther away from the estimate 1.71 in the unconstrained CEVGARCH model than is the GARCH,  $p = 1$ . For the thirty-year daily S&P 500 returns series and the estimated values of the parameters, the super-GARCH volatility exponent, 1.71, estimated for the unconstrained CEV model does not make the conditional variance run up explosively high as frequently as one might first think. During the entire sample period, the sensitivity growth damper was in effect only on 5 days, all of which were within 7 trading days following the October 19, 1987 stock market crash.

The Ljung-Box statistics for the serial correlations of orders up to ten in squared residual returns from the SQGARCH, the GARCH, and the exponent-unconstrained CEVGARCH(1,1) are all approximately 504 (the slight differences are due to the differences in the mean return estimates) strongly indicating time-variation in conditional variances. When we standardize the squared residuals by dividing them by the conditional variance estimates of the respective models, the Ljung-Box statistics are dramatically reduced to 23.23 for the SQGARCH(1,1), 11.00 for the GARCH(1,1), and 5.68 for the unconstrained CEVGARCH. Assuming that these Ljung-Box statistics are distributed as  $\chi^2$  with the degrees of freedom 10 under the null of no serial correlation in the standardized squared residuals, the  $p$ -values are 0.010, 0.358, and 0.841 respectively so that the null is not rejected at usual levels. The order of the three models in terms of the Ljung-Box statistics of the standardized squared residuals is consistent with the likelihood-based comparisons discussed above.

These results favoring the GARCH over the SQGARCH are in agreement with the empirical findings of Engle and Lee [44], who estimated the bivariate continuous-time CEVSV diffusion model

for the daily log returns on the S&P 500 index (January 1971 to September 1990). They applied the indirect inference/EMM estimation method developed by Gouriéroux, Monfort, and Renault [51] and Gallant and Tauchen [51] with the choice of the score generator being the GARCH(1,1) with Gaussian and  $t$ -distributed standardized shocks, and found the estimated value of the exponent  $p$  of the continuous-time SV model to be close to one, along with other evidence against the SQSV. Our estimate of the volatility exponent of the CEVGARCH model being larger than one is also in agreement with Jones [65], who found the estimate of the exponent in his continuous-time CEVSV model for S&P 100 daily returns to be greater than one.

### 5.1.2 Persistence of the conditional variance

Turning our attention to other features of the models, the estimate of the conditional variance persistence parameter  $\gamma$  for the SQGARCH is close to one as in the well-documented case of the GARCH conditional variance persistence (recall that  $\gamma$  is usually expressed as  $\alpha + \beta$  in the GARCH case), although they are statistically significantly smaller than one at usual levels in both cases. The estimate of  $\gamma$  for the exponent-unconstrained CEVGARCH is 1.0065, which is more than two standard errors above one and explosive. Our Monte Carlo estimate,  $-1.242$ , for  $E[\ln(\gamma(1 - \eta_t \eta_L^{-1}))]$  using the estimated parameters and a simulated sample of one million independent realizations of  $t$ -distributed  $z_t$  and the sample analogue,  $-1.314$  (the standard error 0.025), calculated with the estimated parameters and the standardized residuals seem to indicate that the strict stationarity condition is not violated. However, we must note that this is just an informal diagnostic checking. The behavior of the ML estimator for the parameters involved in this calculation may be nonstandard when the DGP is not strictly stationary<sup>14</sup>.

Whether or not the S&P 500 index log returns series is an infinite variance process with  $\gamma > 1$  is also a very important question. One may conjecture from our estimation results that misspecification of the volatility exponent in the SQGARCH and GARCH models leads to inconsistently low estimates of  $\gamma$  and that the true  $\gamma$  is indeed slightly greater than one. However, in light of the empirical and simulation results of Lamoureux and Lastrapes [67] that the prevalence of apparently high volatility persistence in financial time series may be due to the failure of the standard GARCH model to account for the likely presence of structural breaks, we must be cautious in interpreting our results (See also Hamilton and Susmel [57], Granger and Hyung [55], and Andreou and Ghysels [5] on this and related issues). Further investigation of the behavior of the estimator for the persistence parameter in the simultaneous presence of various types of misspecification is clearly needed.

---

<sup>14</sup>Lee and Hansen [66] have established that the Gaussian quasi-ML estimator for the standard GARCH(1,1) model is locally consistent and asymptotically normal under some regularity conditions, which holds for the integrated or explosive GARCH(1,1) DGP as long as the strict stationarity condition  $E[\ln(\beta + \alpha z_t^2)] < 0$  is not violated. It is our hope that their results extend to our  $t$ -distribution-based ML estimator for the CEVGARCH(1,1) model as well.

### 5.1.3 Asymmetry

The estimates of the volatility asymmetry parameter  $\phi$  are slightly less than -1, which are significantly less than zero in all models. The stylized fact that future stock market volatility increases more in response to bad news than to good news appears to be robust to different specifications of the exponent. The role of  $\phi$  is analogous to that of  $\rho$ , the correlation between return and volatility shocks in the SV diffusion models that have attracted a considerable amount of attention in the continuous-time option pricing literature. A negative correlation may cause the Black-Scholes implied volatility curves to exhibit “smirk” patterns, which are observed in the options markets. If we are to consider the discrete-time CEVGARCH model as a good approximation to the CEVSV diffusion process generating the data, then our estimates of  $\phi$  and the  $t$ -distribution’s degrees-of-freedom parameter  $\nu$  can be transformed (subject to the caveat given in Section 4.4) into the “diffusion-limit” estimates of the instantaneous correlation coefficient  $\rho$  of the two Brownian motions driving the bivariate process by the approximation formula  $\rho = \phi/\sqrt{\kappa_z - 1 + \phi^2}$  of Section 4.4 where  $\kappa_z = 3 + 6/(\nu - 4)$ . The numbers thus obtained from the three models are similar and approximately -0.49, which is somewhat less negatively large than the estimates in the neighborhood of -0.6 for the SQSV diffusion models (with or without jumps) for the S&P 500 index, obtained in several studies including Bakshi, Cao, and Chen [6] using cross-sections of observed option prices, Andersen, Benzoni, and Lund [4] using only time series of returns on the index, Pan [77] using a panel of both. These numbers are in line with those in the SQSV and CEVSV diffusion models for the S&P 100 index, obtained by Jones [65]. When we use  $\kappa_z = 3$  and  $\phi$  obtained by either  $t$ -distribution-based or Gaussian ML estimation, the correlation estimates are approximately -0.6. We note, however, that the estimates of  $\rho$  reported in the empirical literature do not seem to be stable across different estimators and different sample periods even when the same S&P 500 index and similar models are employed and that the differences in the estimated values are not solely due a possible segmentation of the underlying and the options markets. The GMM and characteristic function-based estimates in the SQSV diffusion obtained by Jiang and Knight [64] and the indirect-inference and the GARCH diffusion-limit estimates in the CEVSV diffusion model ( $p=1$ ) obtained by Engle and Lee [44] are substantially smaller in absolute magnitude than -0.6. These two studies use only time series observations of index returns. The sample correlation of the daily changes in the implied volatility and the index returns obtained by Bakshi, Cao, and Chen [6] is also much closer to zero than -0.60 is, while the SQSV-implied correlation obtained Nandi [71] using a panel of returns on the index and the option prices is -0.79.

### 5.1.4 Excess kurtosis

The estimates for the  $t$ -distribution’s degrees-of-freedom parameter  $\nu$  is 7.74 for the unconstrained CEVGARCH model, which implies kurtosis of 4.61, well in excess of 3, but smaller than the sample

kurtosis 5.91 of the model’s standardized residuals. The kurtoses for the standardized return shocks implied by the SQGARCH and the GARCH estimates of  $\nu$  are similar to the CEVGARCH’s, but these values fall far short of the sample kurtoses for the SQGARCH and the GARCH standardized residuals (9.59 and 7.57) respectively. However, these large gaps are mostly due to a single extreme observation (-22.83%) of the stock market crash occurring on October 19, 1987. Removing this observation, the sample kurtoses of the standardized residuals are reduced to similar levels across the three models (5.73 for the SQGARCH, 5.76 for the GARCH, and 5.59 for the unconstrained CEVGARCH) while the ML estimates of  $\nu$  are little changed by this removal (the details of re-estimation results not reported here). The reason that the sample kurtosis reduction is not pronounced for the unconstrained CEVGARCH case is that, reacting more sensitively to the pre-crash jitteriness of the market, this model gives a huge prediction of 9.35% for the conditional variance of the October 19, 1987 return (relative to the SQGARCH’s 2.88%, the GARCH’s 4.19%, and the sample variance 0.93% of the S&P 500 returns data) so that the residual for the crash observation is reduced to -7.48% after standardization. Depending on how we constrain the volatility exponent, we reach very different inferences regarding how likely it was to observe a large absolute price change of the crash magnitude. We also note that all of our ML estimates for  $\nu$  are close to the estimate 8.1 obtained by Engle and Lee [44] in an asymmetric version of the GARCH(1,1) for daily S&P 500 returns data .

### 5.1.5 Subsample estimation results

There is a concern that the estimate of the exponent  $p$  being larger than one might also be an artifact of the single crash observation. Re-estimating the models using the dataset that excludes the crash observation, we obtained 1.64 as the exponent estimate (the standard error is 0.08; the details of the estimation results are not reported for brevity), only slightly down from the previous 1.71, so that the large estimate for the exponent is not solely due to the crash observation. The subsample estimation results using each of the 1970s, 1980s, and 1990s datasets are summarized in Tables 2-1, 2-2, and 2-3. A general picture that emerges from these results is that while the exponent estimates are always in the super-GARCH region and the hypothesis  $p = 1/2$  is always strongly rejected, evidence against  $p = 1$  is not always overwhelmingly strong. The 80s exponent estimate is the highest with 1.94 (the standard error is 0.19) with the persistence parameter estimate  $\gamma$  somehow not exceeding one only for this subsample. However, with the 90s subsample, the exponent estimate is relatively small at 1.36 (the standard error is 0.19) albeit still higher than one, and the BIC actually selects the GARCH (but not the SQGARCH) over the unconstrained CEVGARCH. The LR test also strongly rejects both the SQGARCH and the GARCH restrictions on the exponent for the 70s and 80s subsamples, but its  $p$ -value of 1.74% for the test of the GARCH restriction with the 90s subsample is not as small.

### 5.1.6 Prediction of the conditional distribution of the next-period volatility

Lastly, we show in Figure 2 the confidence intervals based on the 5th and 95th percentiles of the conditional distributions of the next-period conditional standard deviation for various levels of current conditional standard deviation<sup>15</sup>. The conditional standard deviations are annualized. The three specifications of the variance of variance lead to different estimates of the conditional distribution of the next-period volatility, particularly when the given current volatility is high (and estimates of unobserved current volatility tend to differ across the three specifications in such a way to magnify, rather than reduce, these discrepancies), which illustrates the importance of accurate modeling of the variance of variance. Notice, also, how the intervals widen as the current level of volatility increases in each case. The SQGARCH's intervals become vanishingly small relative to the deterministic component of time-varying volatility whereas the GARCH's intervals maintain a certain relative size and the intervals of the CEVGARCH with its super-GARCH exponent widens fast, reducing the relative importance of the deterministic component, until the shock damper becomes effective. A visual inspection of Figure 2 helps us understand the phenomenon of variance-of-variance-induced strict stationarity.

## 5.2 Deutsche mark/US dollar exchange rate

Next, we briefly discuss the results of estimation using the DM/US dollar exchange rate data, summarized in Table 3. As in the case of the S&P 500 data, the estimate 1.88 of the volatility exponent  $p$  in the unconstrained CEVGARCH model far exceeds one, while the estimate 1.007 of the persistence parameter  $\gamma$  is slightly above one (but not significantly so at usual levels). The LR test statistics reject both the SQGARCH and the GARCH restrictions at the 1% and 5% levels respectively. The asymmetry parameter  $\phi$  does not have a clear interpretation for exchange rate returns, and as expected, the estimates of  $\phi$  are not significantly different from zero. So the high value of the volatility exponent does not appear to be limited to the stock market data. The Ljung-Box statistics for the hypothesis of no serial correlations up to order 10 in the squared residuals are again dramatically reduced after standardization by their respective conditional variance estimates. In terms of the BIC the GARCH is the best and the CEVGARCH is behind the SQGARCH so that the increase in the maximized log likelihood value due to the flexible exponent is perhaps not enough to justify the added model complexity for this dataset.

---

<sup>15</sup>In each of three case, the estimated process is assumed to be the true DGP. The cumulative distribution function of  $\eta_t = z_t^2 + \phi z_t - 1$  where  $z_t$  is assumed to be a standardized  $t$ -distributed random variable is easily obtained analytically, and is inverted for constructing the intervals.

## 6 Concluding remarks

In the paper, we have developed a new empirical framework for investigating how the sensitivity of the financial market volatility to shocks varies with the volatility level. For this purpose, we first introduced the SQGARCH model, which is an ARCH analogue of the continuous-time SQSV diffusion model. In this model, the variance of variance is a linear function of the conditional variance. We have demonstrated some of the different distributional implications of this variance-of-variance specification, relative to the benchmark case of the more familiar GARCH(1,1) model, and then introduced possibilities for generalization in several directions, including affine, CEV, exponential-CEV, asymmetric and  $(P, Q)$  order. Focusing on the CEVGARCH(1,1) model, which is an ARCH analogue of the continuous-time CEVSV diffusion model, we have established conditions for positivity, stationarity, the existence of finite moments,  $\beta$ -mixing, and other probabilistic properties of the conditional variance process, and noted that strict stationarity of the integrated or mildly explosive CEVGARCH model may be considered a discrete-time analogue of the volatility-induced stationarity phenomenon studied by Conley et al. [29] in the context of continuous-time interest rate models. Our theoretical exercise serves as an example that demonstrates the wide applicability of tools developed in the general state-space Markov chains literature to time series processes defined with highly nonlinear stochastic difference equations. We have also established a diffusion limit for the CEVGARCH(1,1) model, giving a more rigorous foundation for the intuitive link between the continuous-time CEVSV diffusion model and the discrete-time CEVGARCH model.

Paralleling the findings reported for the SQSV diffusion model in the continuous-time estimation literature, our empirical results using data of the daily log returns on the S&P 500 index and the DM/US dollar exchange rate time series find little support for the SQGARCH(1,1) model and suggest that the GARCH(1,1) model provides a better description of time-varying volatility than the SQGARCH(1,1) model does. There may be gains in forecast accuracy by going to the other side of the GARCH volatility exponent since the exponent in the CEV specification of the conditional variance equation appears to exceed one when it is allowed to be flexible and estimated from the data. According to our preliminary empirical results for various other international stock market indices and currencies not examined in this paper, high volatility exponent estimates in the super-GARCH region appear to be universal. If the fast variance-of-variance growth is indeed a true common feature of various financial markets, then a challenge would be to find an economic explanation of the mechanism that gives rise to it.

The lack of empirical support for the SQGARCH model seems to call for a re-examination of the popular practice of using the continuous-time SQSV diffusion in modeling the stock and currency market volatility dynamics. Most of the continuous-time finance researchers have opted to remedy the limitations of the SQSV diffusion model in matching data by adding a price jump component while keeping the square-root specification for the volatility diffusion function intact.

However, the two types of extensions, the SQSV diffusion with a price jump component and the high volatility-exponent CEV without a price jump component, have different static and dynamic implications for the distributions of returns and their volatility, and consequently for the pricing and hedging of related financial derivatives. Our empirical evidence for high elasticity of variance of variance and excess kurtosis that persists after the ARCH-induced effect is accounted for seems to suggest that, if we are to adopt a continuous-time model, a CEVSV jump-diffusion model that incorporates both types of extensions may contribute towards more accurate modeling of the stock index and exchange rate time series.

In this paper, we have not evaluated the performance of the CEVGARCH model with the flexibly estimated exponent value in predicting volatility either in-sample or out-of-sample. Of particular interest is the performance during periods surrounding extremely large stock price or exchange rate movements such as that of October 19, 1987 when the differences in the exponent value are expected to lead to large differences in the volatility prediction. The overprediction of the volatility levels by some models for the periods following the October 1987 crash (see Schwert [80]) may disappear under the estimated high-exponent CEVGARCH model.

## Appendix

As defined in Section 4,  $X := \{X_t : t \in \mathbb{Z}_+\}$  denotes a Markov chain on a general state space  $(S, \mathcal{S})$ , i.e., a measurable space where  $S$  is an arbitrary set and  $\mathcal{S}$  is a countably generated  $\sigma$ -algebra of subsets of  $S$ , with time-homogenous transition densities. First, we show that under some conditions the CEVGARCH(1,1) process is irreducible and aperiodic.

**Definition A.1 (Irreducibility):**  $X$  is called  $\psi$ -irreducible if there exists a  $\sigma$ -finite measure  $\psi$  on  $(S, \mathcal{S})$  with  $\psi(S) > 0$  such that  $P^t(x, A) > 0$  for some  $t$  for all  $x \in S$  and all  $A \in \mathcal{S}$  with  $\psi(A) > 0$ .  $X$  is called irreducible if it is  $\psi$ -irreducible for some  $\psi$  (call it an irreducibility measure). If all irreducibility measures are absolutely continuous w.r.t.  $\psi$ , then  $\psi$  is called a maximal irreducibility measure for  $X$ .

We will reserve the notation  $\psi$  for a maximal irreducibility measure. A maximal irreducibility measure always exists for an irreducible Markov chain (see Nummelin [76, p.13, Proposition 13]), and is unique up to the equivalence of measures.

**Definition A.2 (Aperiodicity):** An irreducible Markov chain  $X$  is called aperiodic if the largest integer  $d$  such that (i)-(iii) hold is one.

- (i)  $S_i \in \mathcal{S}$  for  $i = 1, \dots, d$  and  $S_i \cap S_j = \emptyset$  for  $i \neq j$ .
- (ii) for  $x \in S_i$ ,  $P(x, S_{i+1}) = 1$ ,  $i = 0, 1, \dots, d-1 \pmod{d}$ ,
- (iii)  $\psi\left(\left(\cup_{i=1}^d S_i\right)^c\right) = 0$ .

**Lemma A.1:** The CEVGARCH(1,1) process of Definitions 4.1-4.3 with  $\{\eta_t\}$  satisfying Assumption 4.1 is irreducible and aperiodic.

Proof:

Define a continuous function  $m : G \rightarrow G$ ,  $m(h) := \omega + \gamma h + \alpha H(h) \underline{\eta}$  and denote a solution of the equation  $m(h) - h = 0$  on  $G$  by  $h^*$  (Note that  $H(h) = h^p$  for any  $h \in G$  when  $p \in (0, 1]$ ). For the GARCH case ( $p = 1$ ), clearly  $h^* = \omega \{1 - (\gamma + \alpha \underline{\eta})\}^{-1}$  is the unique solution. For the strictly sub-GARCH case ( $p \in (0, 1)$ ), there is a unique solution  $h^*$  since  $\gamma \leq 1$  and  $\underline{\eta} < 0$  by assumption so that  $m(h_L) - h_L \geq 0$ ,  $m(h) - h < 0$  for all  $h$  large enough, the slope  $m'(h) = \gamma + \alpha \underline{\eta} p h^{p-1}$  of  $m(h)$  is monotone increasing as well as being strictly less than unity for all  $h \in G$ . Also for the strictly super-GARCH case, there is a unique solution  $h^*$ . Regardless of the given value of  $p$ , it clearly holds that  $m(h) - h < 0$  for all  $h > h^*$  and  $m(h) - h > 0$  for all  $h \in (h_L, h^*)$ .

Next, let  $\vartheta$  be the Lebesgue measure restricted on  $W := [h^*, h^* + b]$  for some finite constant  $b > 0$ .  $\vartheta$  is a nontrivial (i.e.,  $\vartheta(G) > 0$ ), finite measure on  $(G, \mathcal{G})$ . For the GARCH case, by the absolute continuity of the distribution of  $\eta_t$  over  $(\underline{\eta}, \bar{\eta})$ ,  $\forall t$ , the distribution of  $h_1$  conditional on  $h_0 = h$ ,  $\forall h \in G$ , is absolutely continuous with a positive density over  $(\omega + (\gamma + \alpha \underline{\eta})h, \omega + (\gamma + \alpha \bar{\eta})h)$  and the distribution of  $h_t$ ,  $t \in \mathbb{N}$ , conditional on  $h_0 = h$  is absolutely continuous with a positive density over  $(\omega \sum_{s=0}^{t-1} (\gamma + \alpha \underline{\eta})^s + (\gamma + \alpha \underline{\eta})^t h, \omega \sum_{s=0}^{t-1} (\gamma + \alpha \bar{\eta})^s + (\gamma + \alpha \bar{\eta})^t h)$ . Since  $\omega \sum_{s=0}^{t-1} (\gamma + \alpha \underline{\eta})^s + (\gamma + \alpha \underline{\eta})^t h \rightarrow h^*$  as  $t \rightarrow \infty$ ,  $\forall h \in G$ , it follows that, if we select a small enough  $b$ , for any  $A \in \mathcal{G}$ ,  $A \subset W$  with  $\vartheta(A) > 0$ , there exists a finite number  $t \in \mathbb{N}$  such that  $P^t(h, A) > 0$ . So  $\{h_t\}$  is

irreducible. We also have  $P^{t+1}(h, A) > 0$  whenever  $P^t(h, A) > 0$  holds so that  $\{h_t\}$  is aperiodic by the sufficiency part of Proposition A1.2 in Chan [24, p.456], whose proof is valid without any modification for a general state space. Proceeding similarly for the strictly sub-GARCH and strictly super-GARCH cases, we obtain the conclusion.  $\square$

The measure  $\vartheta$  defined in each of the three cases in the proof of Lemma A.1 above is an irreducibility measure, and will be used in the proof of Proposition 4.1 given below. The following fact is the main tool that we use for delivering geometric ergodicity of the CEVGARCH(1,1) process. The term ‘‘small set,’’ a general state-space analogue of a single point in a countable state-space, refers to a specific notion, which plays a central role in the general state-space Markov chains theory. However, we refer the reader to Nummelin [76, p.15, Definition 2.3] for its definition since we will rely on the fact (see Feigin and Tweedie [47, p.3]) that if  $S$  for  $X$  is a locally compact complete separable metric space,  $\mathcal{S} = \mathcal{B}(S)$ , and  $X$  is weakly continuous, i.e.,  $E[u(X_{t+1}) | X_t = x]$  is a continuous function of  $x$  for any bounded continuous function  $u : S \rightarrow \mathbb{R}$ , then any relatively compact subset  $A \in \mathcal{S}$  with  $\psi(A) > 0$  is small. Our state space  $[h_L, \infty)$  equipped with the Euclidean metric is a locally compact complete separable metric space and every bounded subset of it is relatively compact.

**Theorem A.1** (Chan [23, Theorem 1]): *Suppose that  $X$  is irreducible and aperiodic and that there exist a measurable function  $g : S \rightarrow [0, \infty]$ , a constant  $r > 1$ , and a small set  $C \in \mathcal{S}$  such that*

$$\sup_{x \in C^c} \{E[rg(X_{t+1}) | X_t = x] - g(x)\} < 0, \quad (\text{A1})$$

$$\sup_{x \in C} \int_{C^c} g(w) P(x, dw) < \infty, \quad (\text{A2})$$

and  $g(x)$  is bounded away from zero and  $+\infty$  on  $C$ . Then  $X$  is geometrically ergodic and there exist constants  $c_1 < 1$ ,  $c_2$ , and  $c_3$  such that  $\|P^t(x, \cdot) - \pi(\cdot)\| \leq c_1^t (c_2 + c_3 g(x))$  for all  $x \in S$  and all  $t \in \mathbb{Z}_+$ .

A nonnegative measurable function  $g$  that satisfies the inequalities in Theorem A.1 or Theorem A.2 below (or some variants of them) is given various names. We will call such a function a *Lyapunov function*.

Proof of Proposition 4.1:

By Lemma A.1,  $\{h_t\}$  is irreducible and aperiodic. Define a function  $f(h, e) := \omega + \beta h + \alpha H(h)e$ . Then,  $f(h, e)$  is continuous in  $h$  and  $e$  so that  $\{h_t\}$  is weakly continuous by the bounded convergence theorem. We also have  $\vartheta(V) > 0$  (and  $\psi(V) > 0$  by the absolute continuity of  $\vartheta$  w.r.t.  $\psi$ ) for  $V := [h_L, h_L + c]$  where  $c > 0$  is a finite constant large enough. Hence,  $V$  is a small set. If we find a Lyapunov function that satisfies (A1) and (A2) in Theorem A.1, then we may conclude that  $\{h_t\}$  is geometrically ergodic. Here, it turns out that  $g_1(h) := h$  works as a Lyapunov function since  $h_t$  is positive and bounded away from zero and  $+\infty$  on  $V$ , and if we take a large enough  $c$ , there exists  $r > 1$  such that

$$\begin{aligned}
& \sup_{h \in V^c} \{rE [g_1 (h_{t+1}) | h_t = h] - h\} \\
&= \sup_{h \in V^c} \{rE [(\omega + \gamma h + \alpha H (h) \eta_t)] - h\} \\
&= \sup_{h \in V^c} h (r (\omega h^{-1} + \gamma) - 1) < 0.
\end{aligned}$$

and

$$\begin{aligned}
\sup_{h \in V} \int_{V^c} g_1 (w) P (h, dw) &\leq \sup_{h \in V} E [\omega + \gamma h + \alpha H (h) \eta_t] \\
&= \sup_{h \in V} (\omega + \gamma h) < \infty
\end{aligned}$$

where  $V^c := G \setminus V$ , since  $\gamma < 1$  by assumption. Next, define  $g_2 (h) := h^p$ , then  $g_2$  works as a Lyapunov function if we show, by taking a large enough  $c$  again, that

$$\begin{aligned}
& \sup_{h \in V^c} \{rE [g_2 (h_{t+1}) | h_t = h] - h^q\} \tag{A3} \\
&= \sup_{h \in V^c} \{rE [(\omega + \gamma h + \alpha H (h) \eta_t)^q] - h^q\} \\
&= \sup_{h \in V^c} h^q (rE [(\omega h^{-1} + \gamma + \alpha H (h) h^{-1} \eta_t)^q] - 1) < 0,
\end{aligned}$$

and

$$\sup_{h \in V} \int_{V^c} g_2 (w) P (h, dw) \leq \sup_{h \in V} E [(\omega + \gamma h + \alpha h^p \eta_t)^q] < \infty. \tag{A4}$$

hold. To verify (A3), (A4), consider the three cases separately.

(i) The strictly sub-GARCH(1,1) case ( $p \in (0, 1)$ ):

Since  $p \in (0, 1)$ , clearly we have  $\lim_{h \rightarrow \infty} (\omega h^{-1} + \gamma + \alpha h^{p-1} \eta_t)^q = \gamma^q$  a.s. Define a variable  $U := (\omega h_L^{-1} + \gamma + \alpha h_L^{p-1} |\eta_t|)^q$ . If  $q \geq 0$ , then  $0 < (\omega h^{-1} + \gamma + \alpha h^{p-1} \eta_t)^q \leq (\omega h^{-1} + \gamma + \alpha h^{p-1} |\eta_t|)^q \leq U$  a.s. for all  $h \in G$ . Consider the two cases  $q \geq 1$  and  $0 \leq q < 1$  separately. First, if  $q \geq 1$ ,  $E[U] \leq (\omega h_L^{-1} + \gamma + \alpha h_L^{p-1} (E[|\eta_t|^q])^{1/q})^q < \infty$  by Minkowski's inequality and the assumption  $E[|\eta_t|^q] < \infty$ . On the other hand, if  $0 \leq q < 1$ ,  $E[U] \leq (\omega h_L^{-1} + \gamma + \alpha h_L^{p-1} E[|\eta_t|])^q < \infty$  as well by Jensen's inequality and the assumption  $E[|\eta_t|^q] < \infty$ . So, we have (A4), and, by the dominated convergence theorem (see, e.g., Durrett [40], p.16) and the assumption  $\gamma < 1$ , we also have  $\lim_{h \rightarrow \infty} E[(\omega h^{-1} + \gamma + \alpha h^{p-1} \eta_t)^q] = \gamma^q < 1$  so that, for all  $h \in V^c$  large enough, there exists  $r > 1$  such that  $rE[(\omega + \gamma h + \alpha h \eta_t)^q] - h^q = h^q (rE[(\omega h^{-1} + \gamma + \alpha h^{p-1} \eta_t)^q] - 1) < 0$ . As for the case of  $q < 0$ , first we note that  $\omega h^{-1} + \gamma + \alpha h^{p-1} \eta_t \geq b_l$  a.s. for all  $h \in G$  for some  $b_l > 0$  since  $\omega h^{-1} + \gamma + \alpha h^{p-1} \eta_L > 0$ ,  $\omega h_L^{-1} + \gamma + \alpha h_L^{p-1} \eta_L > 0$ , and  $\lim_{h \rightarrow \infty} (\omega h^{-1} + \gamma + \alpha h^{p-1} \eta_L) = \gamma > 0$ . Hence,  $0 < (\omega h^{-1} + \gamma + \alpha h^{p-1} \eta_t)^q \leq b_u$  for all  $h \in G$  for some  $b_u > 0$  so that, we again obtain (A3) and (A4) by the bounded convergence theorem (see, e.g., Durrett [40], p.17).

(ii) The GARCH(1,1) case ( $p = 1$ ):

Clearly, we have  $\lim_{h \rightarrow \infty} (\omega h^{-1} + \gamma + \alpha \eta_t)^q = (\gamma + \alpha \eta_t)^q$  a.s. Since Definition 4.2 imposes the condition  $\eta_L \geq -\gamma \alpha^{-1}$ , we have  $\gamma + \alpha \eta_t > 0$  a.s. so that  $0 < (\omega h^{-1} + \gamma + \alpha \eta_t)^q \leq (\omega h_L^{-1} + \gamma + \alpha \eta_t)^q$  a.s. for all  $h \in G$  if  $q \geq 0$  and  $0 < (\omega h^{-1} + \gamma + \alpha \eta_t)^q < (\gamma + \alpha \eta_t)^q$  a.s. if  $q < 0$ . Also, since  $E[(\gamma + \alpha \eta_t)^q] < 1$  by assumption, we have  $E[(\omega h_L^{-1} + \gamma + \alpha \eta_t)^q] \leq (\omega h_L^{-1} + (E[(\gamma + \alpha \eta_t)^q])^{1/q})^q < \infty$  if  $q \geq 1$  by Minkowski's inequality,  $E[(\omega h_L^{-1} + \gamma + \alpha \eta_t)^q] \leq (\omega h_L^{-1} + \gamma)^q < \infty$  if  $0 \leq q < 1$  by Jensen's inequality. So by the dominated convergence theorem,  $\lim_{h \rightarrow \infty} E[(\omega h^{-1} + \gamma + \alpha \eta_t)^q] = E[(\gamma + \alpha \eta_t)^q] < 1$ , and (A3) and (A4) holds.

(iii) The strictly super-GARCH(1,1) case ( $p > 1$ ):

Clearly,  $\lim_{h \rightarrow \infty} (\omega h^{-1} + \gamma + \alpha H(h) h^{-1} \eta_t)^q = \lim_{h \rightarrow \infty} (\omega h^{-1} + \gamma + (h_L h^{-1} - \omega h^{-1} - \gamma) \eta_L^{-1} \eta_t)^q = \gamma^q (1 - \eta_L^{-1} \eta_t)^q$  a.s. If  $q \geq 0$ ,  $(\omega h^{-1} + \gamma + (h_L h^{-1} - \omega h^{-1} - \gamma) \eta_L^{-1} \eta_t)^q = ((\omega h^{-1} + \gamma - h_L h^{-1}) \cdot (1 - \eta_L^{-1} \eta_t) + h_L h^{-1})^q \leq (1 + (\omega h_L^{-1} + \gamma) (1 - \eta_L^{-1} \eta_t))^q$ . And since  $E[(1 - \eta_L^{-1} \eta_t)^q] < \infty$  by assumption,  $E[(1 + (\omega h_L^{-1} + \gamma) (1 - \eta_L^{-1} \eta_t))^q] \leq \left(1 + (\omega h_L^{-1} + \gamma) (E[(1 - \eta_L^{-1} \eta_t)^q])^{1/q}\right)^q < \infty$  by Minkowski's inequality if  $q \geq 1$ ,  $E[(1 + (\omega h_L^{-1} + \gamma) (1 - \eta_L^{-1} \eta_t))^q] \leq (1 + \omega h_L^{-1} + \gamma)^q < \infty$  by Jensen's inequality if  $0 \leq q < 1$ . If  $q < 0$ , then  $(\omega h^{-1} + \gamma + (h_L h^{-1} - \omega h^{-1} - \gamma) \eta_L^{-1} \eta_t)^q = (((\omega - h_L) h^{-1} + \gamma) (1 - \eta_L^{-1} \eta_t) + h_L h^{-1})^q \leq \gamma^q (1 - \eta_L^{-1} \eta_t)^q$ . So it holds, by the dominated convergence theorem, that  $\lim_{h \rightarrow \infty} E[(\omega h^{-1} + \gamma + (h_L h^{-1} - \omega h^{-1} - \gamma) \eta_L^{-1} \eta_t)^q] = \gamma^q E[(1 - \eta_L^{-1} \eta_t)^q] < 1$ .  $\square$

For proving Proposition 4.2, we use the following fact, which is essentially Theorem 4 (i) in Tweedie [86].

**Theorem A.2:** *Suppose that  $X$  is irreducible and aperiodic and that there exist a measurable function  $g : S \rightarrow [0, \infty]$  and a small set  $C \in \mathcal{S}$  such that*

$$\sup_{x \in C^c} \{E[g(X_{t+1}) \mid X_t = x] - g(x)\} < 0 \quad (\text{A5})$$

and

$$\sup_{x \in C} \int_{C^c} g(y) P(x, dy) < \infty. \quad (\text{A6})$$

Then  $X$  is Harris ergodic.

When the state space  $S$  is a subset of  $\mathbb{R}$ ,  $\ln(1 + |\cdot|)$  serves as another candidate for a Lyapunov function; See Feigin and Tweedie [47] for an application. The following corollary specializes Theorem A.2 for this function.

**Corollary A.1:** *Let  $X$  be an irreducible and aperiodic Markov chain on  $(S, \mathcal{S})$ ,  $S \subset \mathbb{R}$ ,  $\mathcal{S} := \mathcal{B}(S)$ , with time homogeneous transition probabilities. Suppose that there exists a small set  $C \in \mathcal{S}$  such that*

$$\sup_{x \in C^c} \{E[\ln(1 + |X_{t+1}|) \mid X_t = x] - \ln(1 + |x|)\} < 0, \quad (\text{A7})$$

$$\sup_{x \in C} \int_{C^c} \ln(1 + |y|) P(x, dy) < \infty \quad (\text{A8})$$

where  $C^c := S \setminus C$ . Then  $X$  is Harris ergodic.

Proof of Proposition 4.2:

Define a set  $C := [x_L, x_L + c]$  for some constant  $c \in (0, \infty)$  large enough so that  $\psi(C) > 0$  holds.  $C$  is a small set since  $X$  is weakly continuous by the continuity of  $\mu_X(x) + \sigma_X(x, e)$  in  $x$  for each fixed  $e$  and the bounded convergence theorem. (A8) holds since  $\int_{C^c} \ln(1 + |y|) P(x, dy) \leq E[\ln(1 + X_{t+1}) | X_t = x] < \ln E[1 + X_{t+1} | X_t = x] = \ln(1 + \mu_X(x)) \leq b < \infty$  for some constant  $b$  for all  $x \in C$  by Jensen's inequality,  $E[\sigma_X(x, e_{t+1})] = 0$ , and the continuity of  $\mu_X(x)$ . Since  $E\left[\ln\left(\frac{\mu_X(x)}{x} + \frac{\sigma_X(x, z_t)}{x}\right)\right] < 0$  holds for all  $x$  large enough by assumption, (A7) holds by making  $c$  large enough and the conclusion follows from Corollary A.1.  $\square$

## References

- [1] Aït-Sahalia, Y., 1996, Nonparametric pricing of interest derivative securities, *Econometrica* **64**, 527-560.
- [2] Aït-Sahalia, Y., 1996, Testing continuous-time models of the spot interest rate, *Review of Financial Studies* **9**, 385-426.
- [3] Aït-Sahalia, Y., 1999, Transition densities for interest rate and other nonlinear diffusions, *Journal of Finance* **54**, 1361-1395.
- [4] Andersen, T.G., L. Benzoni, and J. Lund, 2002, An empirical investigation of continuous-time equity return models, *Journal of Finance* **57**, 1239-1284.
- [5] Andreou, E., and E. Ghysels, 2002, Detecting multiple breaks in financial market dynamics, *Journal of Applied Econometrics* **17**, 579-600.
- [6] Bakshi, G., C. Cao, and Z. Chen, 1997, Empirical performance of alternative option pricing models, *Journal of Finance* **52**, 2003-2049.
- [7] Bates, D., 1996, Jumps and stochastic volatility: Exchange rate process implicit in Deutsche Mark options, *Review of Financial Studies* **9**, 69-107.
- [8] Bates, D., 2000, Post-'87 crash fears in S&P 500 futures options, *Journal of Econometrics* **94**, 181-238.
- [9] Bera, A., and M. Higgins, 1995, On ARCH models: Properties, estimation and testing, in L. Oxley, D. George, C. Roberts, and S. Sayer, eds., *Surveys in Econometrics*, Oxford: Basil Blackwell.
- [10] Black, F., 1976, Studies in stock price volatility changes, *Proceedings of the 1976 Business Meeting of the Business and Economics Statistics Section*, American Statistical Association, 177-181.
- [11] Black, F., and Scholes, M., 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* **81**, 637-654.
- [12] Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31**, 307-327.
- [13] Bollerslev, T., 1987, A conditional heteroskedastic time series model for speculative prices and rates of return, *Review of Economics and Statistics* **69**, 542-547.
- [14] Bollerslev, T., R. Chou, and K. Kroner, 1992, ARCH modeling in finance: A review of the theory and empirical evidence, *Journal of Econometrics* **52**, 5-59.

- [15] Bollerslev, T., R.F. Engle, and D. Nelson, 1994, ARCH models, in R. Engle and D. McFadden, eds., *Handbook of Econometrics, Volume IV*, Amsterdam: Elsevier Science.
- [16] Brenner, R., R. Harjes, and K. Kroner, 1996, Another look at models of the short-term interest rate, *Journal of Financial and Quantitative Analysis* **31**, 85-107.
- [17] Brown, R., and S. Schaefer, 1993, Interest rate volatility and the term structure of interest rates, *Philosophical Transactions of the Royal Society, Physical Sciences and Engineering* **347**, 563-576.
- [18] Borovkov, A.A., 1998, *Ergodicity and stability of stochastic processes*, New York: John Wiley & Sons.
- [19] Campbell, J., and L. Hentschel, 1992, No news is good news: An asymmetric model of changing volatility in stock returns, *Journal of Financial Economics* **31**, 281-318.
- [20] Carrasco, M., and X. Chen, 2002, Mixing and moment properties of various GARCH and stochastic volatility models, *Econometric Theory* **18**, 17-39.
- [21] Chan, K.C., G.A. Karolyi, F.A. Longstaff, and A.B. Sanders, 1992, An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance* **48**, 1209-1227.
- [22] Chan, K.S., 1986, *Topics in Nonlinear Time Series Analysis*, Ph.D. Dissertation, Princeton University.
- [23] Chan, K.S., 1989, A note on the geometric ergodicity of a Markov chain, *Advances in Probability* **21**, 702-704.
- [24] Chan, K.S., 1990, Deterministic stability, stochastic stability, and ergodicity, in H. Tong, *Nonlinear Time Series: A Dynamical System Approach*, Oxford: Oxford University Press.
- [25] Chan, K.S., and H. Tong, 1994, A note on noisy chaos, *Journal of the Royal Statistical Society B* **56**, 301-311.
- [26] Chan, K.S., and H. Tong, 2001, *Chaos: A Statistical Perspective*, New York: Springer-Verlag.
- [27] Christie, A., 1982, The stochastic behavior of common stock variance: Value, leverage and interest rate effects, *Journal of Financial Economics* **10**, 407-437.
- [28] Chernov, M., and E. Ghysels, 2000, A study towards a unified approach to the joint estimation of objective and risk-neutral measures for the purpose of options valuation, *Journal of Financial Economics* **56**, 407-458.
- [29] Conley, T.G, L.P. Hansen, E.G.J. Luttmer, and J.A. Scheinkman, 1997, Short-term interest rates as subordinated diffusions, *Review of Financial Studies* **10**, 525-577.

- [30] Cox, J., 1975, Notes on option pricing I: Constant elasticity of variance diffusion, Working paper, Stanford University.
- [31] Cox, J., J. Ingersoll, and S. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* **53**, 385-408.
- [32] Dai, Q., and K.J. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* **55**, 1943-1973.
- [33] Davydov, Y.A., 1973, Mixing conditions for Markov chains, *Theory of Probability and Its Applications* **18**, 312-328.
- [34] Ding, Z., C.W.J. Granger, and R.F. Engle, 1993, A long memory property of stock market returns and a new model, *Journal of Empirical Finance* **1**, 83-106.
- [35] Drost, F.C., and T.E. Nijman, 1993, Temporal aggregation of GARCH processes, *Econometrica* **61**, 909-927.
- [36] Drost, F.C., and B.J.M. Werker, 1996, Closing the GARCH gap: Continuous time GARCH modeling, *Journal of Econometrics* **74**, 31-57.
- [37] Duan, J., 1995, The GARCH option pricing model, *Mathematical Finance* **5**, 13-32.
- [38] Duan, J., 1996, A unified theory of option pricing under stochastic volatility: From GARCH to diffusion, Working paper.
- [39] Duan, J., 1997, Augmented GARCH( $p, q$ ) process and its diffusion limit, *Journal of Econometrics* **79**, 97-127.
- [40] Durrett, R., 1996, *Probability: Theory and examples, Second Edition*, Belmont: Wadsworth Publishing Co.
- [41] Duffie, D., and R. Kan, 1996, A yield-factor model of interest rates, *Mathematical Finance* **6**, 379-406.
- [42] Duffie, D., J. Pan, and K. Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* **68**, 1343-1376.
- [43] Engle, R., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation, *Econometrica* **50**, 987-1008.
- [44] Engle, R.F., and G. Lee, 1996, Estimating diffusion models of stochastic volatility, in P. Rossi, ed., *Modelling Stock Market Volatility*, San Diego: Academic Press.

- [45] Engle, R.F., D. Lilien, and R. Robins, 1987, Estimating time varying risk premia in the term structure: The ARCH-M model, *Econometrica* **55**, 391-407.
- [46] Engle, R.F., and V. Ng, 1993, Measuring and testing the impact of news on volatility, *Journal of Finance* **48**, 1749-1778.
- [47] Feigin, P.D., and R.L. Tweedie, 1985, Random coefficient autoregressive processes: A Markov chain analysis of stationary and finiteness of moments, *Journal of Time Series Analysis* **6**, 1-14.
- [48] Feller, W., 1951, Two singular diffusion problems, *Annals of Mathematics* **54**, 173-182.
- [49] Fonseca, G., and R.L. Tweedie, 2001, Stationary measures for non-irreducible non-continuous Markov chains with time series applications, Working paper.
- [50] Friedman, B., and D. Laibson, 1989, Economic implications of extraordinary movements in stock prices, *Brookings Papers on Economic Activity* **1989**, 137-189.
- [51] Gallant, A., and Tauchen, G., 1996, Which moments to match, *Econometric Theory* **12**, 657-681.
- [52] Glosten, L.R., R. Jagannathan, and D.E. Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* **48**, 1779-1801.
- [53] Gouriéroux, C., 1997, *ARCH Models and Financial Applications*, New York: Springer-Verlag.
- [54] Gouriéroux, C., A. Monfort, and E. Renault, 1993, Indirect inference, *Journal of Applied Econometrics* **8**, S85-S118.
- [55] Granger, C.W.J., and N. Hyung, 1999, Occasional structural breaks and long memory, Discussion Paper 99-14, Department of Economics, University of California, San Diego.
- [56] Hamilton, J.D., 1994, *Time Series Analysis*, Princeton University Press, Princeton, NJ.
- [57] Hamilton, J.D., and R. Susmel, 1994, Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics* **64**, 307-333.
- [58] Hansen, P.R., and A. Lunde, 2001, A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?, Working Paper No. 01-04, Department of Economics, Brown University.
- [59] Hansen, L.P., and J.A. Scheinkman, 1995, Back to the future: Generating moment implications for continuous-time Markov processes, *Econometrica* **63**, 767-804.

- [60] Hentschel, L., 1995, All in the family: Nesting symmetric and asymmetric GARCH models, *Journal of Financial Economics* **39**, 77-104.
- [61] Heston, S.L, 1993, A closed form solution for options with stochastic volatility with applications bond and currency options, *Review of Financial Studies* **6**, 327-343.
- [62] Heston, S., and S. Nandi, 2000, A closed-form GARCH option pricing model, *Review of Financial Studies* **13**, 585-625.
- [63] Hull, J., and A. White, 1987, The pricing of options on assets with stochastic volatilities, *Journal of Finance* **42**, 281-300.
- [64] Jiang, J.J., and J.L. Knight, 2000, Estimation of continuous time processes via the empirical characteristic function, Working paper.
- [65] Jones, C., 2000, The dynamics of stochastic volatility: Evidence from underlying and options markets, Working paper.
- [66] Lee, S., and B.E. Hansen, 1994, Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator, *Econometric Theory* **10**, 29-52.
- [67] Lamoureux, C.G., and W.D. Lastrapes, 1990, Persistence in variance, structural change, and the GARCH model, *Journal of Business and Economic Statistics* **8**, 225-234.
- [68] Li, W.K., S. Ling, and M. McAleer, 2002, Recent theoretical results for time series models with GARCH errors, *Journal of Economic Surveys* **16**, 245-269.
- [69] Meyn, S., and R.L. Tweedie, 1993, *Markov chains and stochastic stability*, London: Springer-Verlag.
- [70] Merton, R., 1973, Rational theory of option pricing, *Bell Journal of Economics and Management Science* **4**, 141-183.
- [71] Nandi, S, How important is the correlation between returns and volatility in a stochastic volatility model? Empirical evidence from pricing and hedging in the S&P 500 index options market, *Journal of Banking & Finance* **22**, 589-610.
- [72] Nelson, D.B., 1990, ARCH models as diffusion approximations, *Journal of Econometrics* **45**, 7-38.
- [73] Nelson, D.B., 1990, Stationarity and persistence in the GARCH (1,1) model, *Econometric Theory* **6**, 318-334.
- [74] Nelson, D.B., 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* **59**, 347-370.

- [75] Nelson, D.B., and D.P. Foster, 1994, Asymptotic filtering theory for univariate ARCH models, *Econometrica* **62**, 1-41.
- [76] Nummelin, 1984, *General irreducible Markov chains and non-negative operators*, Cambridge: Cambridge University Press.
- [77] Pan, J., 2002, The jump-risk premia implicit in options: Evidence from an integrated time series study, *Journal of Financial Economics* **63**, 3-50.
- [78] Palm, F.C., 1996, GARCH models of volatility, in G.S. Maddala and C.R. Rao, eds., *Statistical Methods in Finance*, Amsterdam: Elsevier Science.
- [79] Quinn, B.G., 1982, A note on the existence of strictly stationary solutions to bilinear equations, *Journal of Time Series Analysis* **3**, 249-252.
- [80] Schwert, G.W., 1990, Stock volatility and the crash of '87, *Review of Financial Studies* **3**, 77-102.
- [81] Singleton, K., 2000, Estimation of affine asset pricing models using the empirical characteristic function, *Journal of Econometrics* **102**, 111-141.
- [82] Stein, E., and J. Stein, 1991, Stock price distribution with stochastic volatility: An analytic approach, *Review of Financial Studies* **4**, 727-753.
- [83] Taylor, S., 1986, *Modeling financial time series*, New York: Wiley.
- [84] Tong, H., 1981, A note on a Markov bilinear stochastic process in discrete time, *Journal of Time Series Analysis* **2**, 279-284.
- [85] Tong, H., 1990, *Non-linear Time Series: A Dynamical System Approach*, Oxford: Oxford University Press.
- [86] Tweedie, R.L., 1983, Criteria for rates of convergence of Markov chains, with applications to queueing and storage theory, in J.F.C. King and G.E.H. Reuter, eds., *Probability, Statistics and Analysis*, Cambridge: Cambridge University Press.
- [87] Tweedie, R.L., 2001, Drift conditions and invariant measures for Markov chains, *Stochastic Processes and their Applications* **92**, 345-354.
- [88] Wang, Y., 2002, Asymptotic nonequivalence of GARCH models and diffusions, *Annals of Statistics* **30**, 754-783.
- [89] Wiggins, J., 1987, Option values under stochastic volatility: Theory and empirical estimates, *Journal of Financial Economics* **19**, 351-372.

- [90] Zakoïan, J.M., 1994, Threshold heteroskedastic models, *Journal of Economic Dynamics and Control* **18**, 931-955.

**Table 1: CEVGARCH(1,1) Model Estimation Results**  
**Daily Returns on the S&P 500 Index: Jan. 1970 - Dec. 1999**

Parameter	SQGARCH(1,1) $p=0.5$		GARCH(1,1) $p=1.0$		CEVGARCH(1,1)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
$\mu$	0.03865	(0.00857)	0.03919	(0.00852)	0.03843	(0.00851)
$\omega$	0.01471	(0.00238)	0.00909	(0.00183)	0.00025	(0.00118)
$\gamma$	0.98086	(0.00292)	0.98979	(0.00282)	1.00654	(0.00303)
$\alpha$	0.03983	(0.00423)	0.05177	(0.00529)	0.06501	(0.00643)
$\phi$	-1.09523	(0.15590)	-1.07439	(0.15408)	-1.06948	(0.12443)
$p$					1.71208	(0.09841)
$\nu$	7.43973	(0.60114)	7.52439	(0.58031)	7.73637	(0.58165)
$\kappa_z$	4.74432		4.70242		4.60584	
$\kappa_z^*$	9.59383	5.72987	7.57255	5.76146	5.90811	5.58593
Log likelihood	-9955.43001		-9924.24664		-9889.18299	
AIC	2.46265		2.45853		2.45416	
BIC	2.46815		2.46403		2.46057	
LR	132.49404	(0.00000)	70.12729	(0.00000)		
$Q_{10}(\varepsilon_t^2)$	503.68002	(0.00000)	503.65942	0.00000	503.6882368	(0.00000)
$Q_{10}(\varepsilon_t^2/h_t)$	23.23164	(0.00992)	10.99620	0.35781	5.67909	(0.84147)

$\kappa_z$ : Kurtosis of the standardized shock  $z_t$  implied by  $\nu$ .

$\kappa_z^*$ : The sample kurtosis of the standardized residuals (The number on the second column for each model is obtained from the sample without the October 19, 1987).

LR: The likelihood ratio test statistics for the hypotheses of  $p = 1/2$  and  $p = 1$  respectively for the SQGARCH and the GARCH models. In brackets are the  $p$ -values.

$Q_{10}(\varepsilon_t^2)$ : The Ljung-Box statistics for the hypothesis of no serial correlations of orders up to the 10th for the squared residuals.

$Q_{10}(\varepsilon_t^2/h_t)$ : The Ljung-Box statistics for the hypothesis of no serial correlations of orders up to the 10th for the squared standardized residuals. In brackets are the  $p$ -values.

**Table 2-1: CEVGARCH(1,1) Model Estimation Results**  
**Daily Returns on the S&P 500 Stock Index: Jan. 1970 - Dec. 1979**

Parameter	SQGARCH(1,1) $p=0.5$		GARCH(1,1) $p=1.0$		CEVGARCH(1,1)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
$\mu$	0.01274	(0.01464)	0.01363	(0.01451)	0.01308	(0.01448)
$\omega$	0.01155	(0.00311)	0.00914	(0.00261)	0.00273	(0.00272)
$\gamma$	0.98367	(0.00439)	0.98703	(0.00460)	1.00070	(0.00683)
$\alpha$	0.03983	(0.00685)	0.05335	(0.00928)	0.07657	(0.01286)
$\phi$	-1.10280	(0.24623)	-1.16595	(0.25034)	-1.05802	(0.21768)
$p$					1.66389	(0.16661)
$\nu$	14.07305	(3.61830)	15.09703	(4.32921)	15.09929	(4.36358)
$\kappa_z$	3.59565		3.54069		3.54058	
$\kappa_z^*$	3.72738		3.72869		3.78227	
Log likelihood	-3093.142		-3076.867		-3068.702	
AIC	2.38184		2.37535		2.37290	
BIC	2.39577		2.38929		2.38915	
LR	48.887977	(0.00000)	16.32898	(0.00005)		
$Q_{10}(\varepsilon^2)$	748.31876	(0.00000)	749.04149	(0.00000)	748.59441	(0.00000)
$Q_{10}(\varepsilon_t^2/h_t)$	19.29262	(0.03670)	21.29377	(0.01914)	23.05371	(0.01055)

$\kappa_z$ : Kurtosis of the standardized shock  $z_t$  implied by  $\nu$ .

$\kappa_z^*$ : The sample kurtosis of the standardized residuals.

LR: The likelihood ratio test statistics for the hypotheses of  $p = 1/2$  and  $p = 1$  respectively for the SQGARCH and the GARCH models. In brackets are the  $p$ -values.

$Q_{10}(\varepsilon_t^2)$ : The Ljung-Box statistics for the hypothesis of no serial correlations of orders up to the 10th for the squared residuals.

$Q_{10}(\varepsilon_t^2/h_t)$ : The Ljung-Box statistics for the hypothesis of no serial correlations of orders up to the 10th for the squared standardized residuals. In brackets are the  $p$ -values.

**Table 2-2: CEVGARCH(1,1) Model Estimation Results  
Daily Returns on the S&P 500 Index: Jan.1980-Dec.1989**

Parameter	SQGARCH(1,1) $p=0.5$		GARCH(1,1) $p=1.0$		CEVGARCH(1,1)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
$\mu$	0.05077	(0.01664)	0.05083	(0.01659)	0.04856	(0.01655)
$\omega$	0.02929	(0.00770)	0.02807	(0.00811)	0.00437	(0.00619)
$\gamma$	0.96671	(0.00809)	0.96828	(0.00916)	0.99959	(0.00888)
$\alpha$	0.03324	(0.00751)	0.04006	(0.00882)	0.04639	(0.00910)
$\phi$	-0.79038	(0.38345)	-0.88636	(0.36743)	-1.01516	(0.24469)
$p$					1.93918	(0.18519)
$\nu$	5.88606	(0.64433)	5.95406	(0.70179)	6.33283	(0.72514)
$\kappa_z$	6.18124		6.07053		5.57198	
$\kappa_z^*$	20.87390	6.43000	15.00387	6.39000	8.26641	6.68000
Log likelihood	-3779.91766		-3775.10431		-3765.39863	
AIC	2.64588		2.64398		2.64093	
BIC	2.65974		2.65784		2.65710	
LR	29.03806	(0.00000)	19.41136	(0.00001)		
$Q_{10}(\varepsilon^2)$	154.16949	(0.00000)	154.16860	(0.00000)	154.20485	(0.00000)
$Q_{10}(\varepsilon_t^2/h_t)$	14.79788	(0.13961)	11.58822	(0.31356)	5.76552	(0.83456)

$\kappa_z$ : Kurtosis of the standardized shock  $z_t$  implied by  $\nu$ .

$\kappa_z^*$ : The sample kurtosis of the standardized residuals (The number on the second column for each model is obtained from the sample without the October 19, 1987.

LR: The likelihood ratio test statistics for the hypotheses of  $p = 1/2$  and  $p = 1$  respectively for the SQGARCH and the GARCH models. In brackets are the  $p$ -values.

$Q_{10}(\varepsilon_t^2)$ : The Ljung-Box statistics for the hypothesis of no serial correlations of orders up to the 10th for the squared residuals.

$Q_{10}(\varepsilon_t^2/h_t)$ : The Ljung-Box statistics for the hypothesis of no serial correlations of orders up to the 10th for the squared standardized residuals. In brackets are the  $p$ -values.

**Table 2-3: CEVGARCH(1,1) Model Estimation Results**  
**Daily Returns on the S&P 500 Stock Index: Jan. 1990 - Dec.1999**

Parameter	SQGARCH(1,1) p=0.5		GARCH(1,1) p=1.0		CEVGARCH(1,1)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
$\mu$	0.05272	(0.01362)	0.05212	(0.01352)	0.05268	(0.01355)
$\omega$	0.00944	(0.00380)	0.00610	(0.00244)	0.00231	(0.00244)
$\gamma$	0.98756	(0.00507)	0.99466	(0.00432)	1.00296	(0.00615)
$\alpha$	0.03729	(0.00766)	0.05396	(0.00996)	0.06446	(0.01230)
$\phi$	-1.33626	(0.29331)	-1.36598	(0.26649)	-1.30191	(0.25415)
$p$					1.36732	(0.19217)
$\nu$	6.52411	(0.84250)	6.65381	(0.87951)	6.71930	(0.88306)
$\kappa_z$	5.37708		5.26090		5.20645	
$\kappa_z^*$	5.41987		5.22204		5.07643	
Log likelihood	-3024.42876		-3013.24885		-3010.41868	
AIC	2.34632		2.35773		2.34157	
BIC	2.36018		2.35575		2.35773	
LR	28.02015	(0.00000)	5.66033	(0.01735)		
$Q_{10}(\varepsilon^2)$	408.03963	(0.00000)	408.06164	(0.00000)	408.04105	(0.00000)
$Q_{10}(\varepsilon_t^2/h_t)$	3.96156	(0.94906)	3.30788	(0.97322)	3.77907	(0.95676)

$\kappa_z$ : Kurtosis of the standardized shock  $z_t$  implied by  $\nu$ .

$\kappa_z^*$ : The sample kurtosis of the standardized residuals.

LR: The likelihood ratio test statistics for the hypotheses of  $p = 1/2$  and  $p = 1$  respectively for the SQGARCH and the GARCH models. In brackets are the  $p$ -values.

$Q_{10}(\varepsilon_t^2)$ : The Ljung-Box statistics for the hypothesis of no serial correlations of orders up to the 10th for the squared residuals.

$Q_{10}(\varepsilon_t^2/h_t)$ : The Ljung-Box statistics for the hypothesis of no serial correlations of orders up to the 10th for the squared standardized residuals. In brackets are the  $p$ -values.

**Table 3: CEVGARCH(1,1) Model Estimation Results**  
**Daily Returns on the DM/US\$ Exchange Rate: Jan.1990-Dec.1999**

Parameter	SQGARCH(1,1) $p=0.5$		GARCH(1,1) $p=1.0$		CEVGARCH(1,1)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
$\mu$	0.01298	(0.01109)	0.01329	(0.01105)	0.01377	(0.01106)
$\omega$	0.00614	(0.00242)	0.00487	(0.00216)	0.00000	(0.00218)
$\gamma$	0.98609	(0.00569)	0.98998	(0.00582)	1.00661	(0.00851)
$\alpha$	0.02678	(0.00521)	0.04807	(0.00918)	0.11874	(0.04064)
$\phi$	-0.21999	(0.26949)	-0.19668	(0.25927)	-0.18824	(0.23968)
$p$					1.88061	(0.32917)
$\nu$	6.25294	(0.74427)	6.32090	(0.78584)	6.29630	(0.80574)
$\kappa_z$	5.66319		5.58521		5.61290	
$\kappa_z^*$	4.47409		4.43081		4.39664	
Log likelihood	-1832.37605		-1828.02072		-1824.87913	
AIC	1.86407		1.86237		1.86192	
BIC	1.87775		1.87605		1.87789	
LR	14.9938	(0.00011)	6.28318	(0.01218)		
$Q_{10}(\varepsilon^2)$	272.30429	(0.00000)	272.24408	(0.00000)	272.14928	(0.00000)
$Q_{10}(\varepsilon_t^2/h_t)$	7.05396	(0.72034)	7.41637	(0.68564)	8.20897	(0.60843)

$\kappa_z$ : Kurtosis of the standardized shock  $z_t$  implied by  $\nu$ .

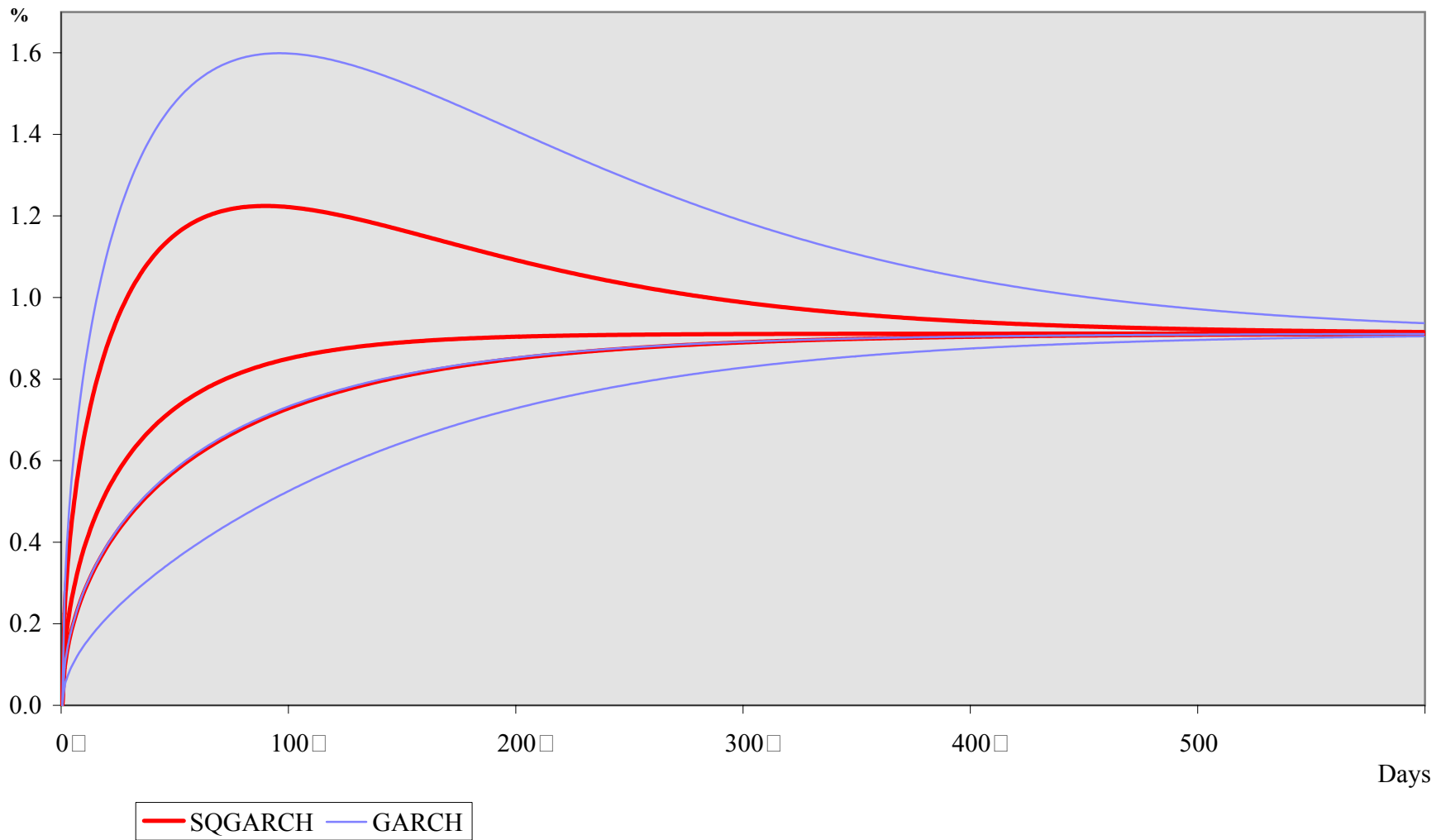
$\kappa_z^*$ : The sample kurtosis of the standardized residuals.

LR: The likelihood ratio test statistics for the hypotheses of  $p = 1/2$  and  $p = 1$  respectively for the SQGARCH and the GARCH models. In brackets are the  $p$ -values.

$Q_{10}(\varepsilon_t^2)$ : The Ljung-Box statistics for the hypothesis of no serial correlations of orders up to the 10th for the squared residuals.

$Q_{10}(\varepsilon_t^2/h_t)$ : The Ljung-Box statistics for the hypothesis of no serial correlations of orders up to the 10th for the squared standardized residuals. In brackets are the  $p$ -values.

**Figure 1:**  
**Term Structure of Conditional Standard Deviation of Conditional Variance:**  
**GARCH versus SQGARCH**



**Figure 2: Confidence Intervals for the Next-Period Conditional Standard Deviation (Annualized %)**

