Journal of Econometrics 45 (1990) 7–38. North-Holland

ARCH MODELS AS DIFFUSION APPROXIMATIONS*

Daniel B. Nelson

University of Chicago, Chicago, IL 60637, USA

This paper investigates the convergence of stochastic difference equations (e.g., ARCH) to stochastic differential equations as the length of the discrete time intervals between observations goes to zero. These results are applied to the GARCH(1, 1) model of Bollerslev (1986) and to the AR(1) Exponential ARCH model of Nelson (1989). In their continuous time limits, the conditional variance processes in these models have stationary distributions that are inverted gamma and lognormal, respectively. In addition, a class of diffusion approximations based on the Exponential ARCH model is developed.

1. Introduction

Many econometric studies [e.g., Officer (1973), Black (1976), Engle and Bollerslev (1986), and French, Schwert, and Stambaugh (1987)] have documented that financial time series tend to be highly heteroskedastic. This has important implications for many areas of macroeconomics and finance, including the term structure of interest rates [e.g., Barsky (1989), Abel (1988)], irreversible investment [e.g., Bernanke (1983), McDonald and Siegel (1986)], options pricing [e.g., Wiggins (1987)], and dynamic capital asset pricing theory [e.g., Merton (1973), Cox, Ingersoll, and Ross (1985)]. Many of these theoretical models have made extensive use of the Ito stochastic calculus [see, e.g., Liptser and Shiryayev (1977)], which provides an elegant and relatively straightforward means to analyze the properties of many diffusion processes.[1]

Econometricians have also been very active in developing models of conditional heteroskedasticity. The most widely used models of dynamic

*This paper is a revision of parts of chapter III of Nelson (1988a). The Department of Education and the University of Chicago Graduate School of Business provided research support. I am indebted to John Cox, Richard Dudley, Ricardo Garcia-Badaracco, Mark Nelson, Richard Parkus, James Poterba, Peter Rossi, G. William Schwert, Daniel Stroock, S.R.S. Varadhan, Jeffrey Wooldridge, Arnold Zellner, an editor (Angelo Melino), and a referee for helpful comments. Seminar participants at the 1987 Econometric Society Winter Meetings, at the NBER Conference on Econometric Methods and Financial Time Series, and at workshops at M.I.T., Chicago, Princeton, Wharton, Stanford, Berkeley, Northwestern, Yale, and Rochester also made useful suggestions. Any remaining errors are mine.

[1]For formal definitions of diffusions and Ito processes, see, e.g., Liptser and Shiryayev (1977) and Arnold (1974).

conditional variance have been the ARCH models first introduced by Engle (1982). In its most general form [see Engle (1982, eqs. 1–5)], a univariate ARCH model makes conditional variance at time t a function of exogenous and lagged endogenous variables, time, parameters, and past residuals. Formally, let e_t be a sequence of (orthogonal) prediction errors, b a vector of parameters, x_t a vector of exogenous and lagged endogenous variables, and σ_t^2 the variance of e_t given information at time t:

$$e_t = \sigma_t Z_t, \tag{1.1}$$

$$Z_t \sim \text{i.i.d.} \quad \text{with} \quad \mathrm{E}(Z_t) = 0, \quad \operatorname{var}(Z_t) = 1, \tag{1.2}$$

$$\begin{aligned}\sigma_t^2 &= \sigma^2(e_{t-1}, e_{t-2}, \dots, x_t, t, b) \\ &= \sigma^2(\sigma_{t-1}Z_{t-1}, \sigma_{t-2}Z_{t-2}, \dots, x_t, t, b).\end{aligned} \tag{1.3}$$

Many different parameterizations for the function $\sigma^2(\cdot, \dots, \cdot)$ have been used in the literature, including the original ARCH(p) specification of Engle (1982), the GARCH and GARCH-M models of Bollerslev (1986) and Engle and Bollerslev (1986), respectively, the Log-GARCH models of Pantula (1986) and Geweke (1986), and the Exponential ARCH model of Nelson (1989).

In contrast to the stochastic differential equation models so frequently found in the theoretical finance literature, ARCH models are discrete time stochastic difference equation systems. It is clear why empiricists have favored the discrete time approach of ARCH: virtually all economic time series data are recorded only at discrete intervals, and a discrete time ARCH likelihood function is usually easy to compute and maximize. By contrast, the likelihood of a nonlinear stochastic differential equation system observed at discrete intervals can be very difficult to derive, especially when there are unobservable state variables (e.g., conditional variance) in the system.[2]

To date, relatively little work has been done on the relation between the continuous time nonlinear stochastic differential systems, used in so much of the theoretical literature, and the ARCH stochastic difference equation systems, favored by empiricists. Indeed, the two literatures have developed quite independently, with little attempt to reconcile the discrete and continuous models. In this paper we partially bridge the gap by developing conditions under which ARCH stochastic difference equations systems converge in distribution to Ito processes as the length of the discrete time intervals goes to zero.

[2]But see Lo (1988) and Duffie and Singleton (1989) for some progress in estimating such models.

What do we hope to gain from such an enterprise? First, as we've indicated above, it may be much easier to do estimation and forecasting with a discrete time ARCH model than with a diffusion model observed at discrete intervals. We may therefore, as the title of the paper suggests, want to use ARCH models as diffusion approximations.

Second, in some cases we may find that distributional results are available for the diffusion limit of a sequence of ARCH processes that are not available for the discrete time ARCH processes themselves. In such cases, we may be able to use diffusion processes as ARCH approximations. For example, consider the much used GARCH(1, 1) model of Bollerslev (1986): though Nelson (1988b) and Sampson (1988) have recently developed necessary and sufficient conditions for the strict stationarity of the conditional variance process, relatively little is known about the stationary distribution. In continuous time, however, we will see later in the paper that the stationary distribution for the GARCH(1, 1) conditional variance process is an inverted gamma. This implies that in a conditionally normal GARCH(1, 1) process observed at short time intervals, the innovations process is approximately distributed as a Student t. In the Exponential ARCH model of Nelson (1989), on the other hand, the conditional variance in continuous time is lognormal, so that in the discrete time model when time intervals are short, the stationary distribution of the innovations is approximately a normal–lognormal mixture, as in the model of Clark (1973).

In section 2, we first present conditions developed by Stroock and Varadhan (1979) for a sequence of stochastic difference equations to converge weakly to an Ito process, and then present an alternate, somewhat simpler, set of conditions. We apply these results to find and analyze the diffusion limit of the GARCH(1, 1) process. In section 3, we introduce a class of ARCH diffusion approximations, and show that they can approximate quite a wide variety of Ito processes. Two examples are provided, the first based on the AR(1) Exponential ARCH model of Nelson (1989) and the second a model with the same diffusion limit as GARCH(1, 1). Section 4 is a brief conclusion. Appendix A summarizes some regularity conditions needed for the convergence results in section 2, and all proofs are in appendix B.

2. Convergence of stochastic difference equations to stochastic differential equations

In this section we present general conditions for a sequence of finite-dimensional discrete time Markov processes $\{_hX_t\}_{h\downarrow 0}$ to converge weakly to an Ito process. As we noted in the Introduction, these are drawn largely from Stroock and Varadhan (1979). Kushner (1984) and Ethier and Kurtz (1986) have extended these convergence theorems to jump-diffusion processes, but it would be beyond the scope of this paper to deal with this more general case.

2.1. The main convergence result

Our formal setup is as follows: Let $D([0,\infty), R^n)$ be the space of mappings from $[0,\infty)$ into R^n that are continuous from the right with finite left limits, and let $B(R^n)$ denote the Borel sets on R^n. D is a metric space when endowed with the Skorohod metric [Billingsley (1968)]. For each $h > 0$, let M_{kh} be the σ-algebra generated by kh, ${}_hX_0, {}_hX_h, {}_hX_{2h}, \ldots, {}_hX_{kh}$, and let ν_h be a probability measure on $(R^n, B(R^n))$. For each $h > 0$ and each $k = 0, 1, 2, \ldots$, let $\Pi_{h,kh}(x, \cdot)$ be a transition function on R^n, i.e.,

a) $\Pi_{h,kh}(x, \cdot)$ is a probability measure on $(R^n, B(R^n))$ for all $x \in R^n$,
b) $\Pi_{h,kh}(\cdot, \Gamma)$ is $B(R^n)$ measurable for all $\Gamma \in B(R^n)$.

For each $h > 0$, let P_h be the probability measure on $D([0,\infty), R^n)$ such that

$$P_h[{}_hX_0 \in \Gamma] = \nu_h(\Gamma) \quad \text{for any} \quad \Gamma \in B(R^n), \tag{2.1}$$

$$P_h\left[{}_hX_t = {}_hX_{kh}, kh \leq t < (k+1)h\right] = 1, \tag{2.2}$$

$$P_h\left[{}_hX_{(k+1)h} \in \Gamma | M_{kh}\right] = \Pi_{h,kh}({}_hX_{kh}, \Gamma) \quad \text{almost surely under } P_h$$

$$\text{for all} \quad k \geq 0 \quad \text{and} \quad \Gamma \in B(R^n). \tag{2.3}$$

For each $h > 0$, (2.1) specifies the distribution of the random starting point and (2.3) the transition probabilities of the n-dimensional discrete time Markov process ${}_hX_{kh}$. We form the continuous time process ${}_hX_t$ from the discrete time process ${}_hX_{kh}$ by (2.2), making ${}_hX_t$ a step function with jumps at times h, $2h$, $3h$, and so on. In this somewhat complicated setup, our notation must keep track of three distinct kinds of processes:

(a) the sequence of discrete time processes $\{{}_hX_{kh}\}$ that depend both on h and on the (discrete) time index kh, $k = 0, 1, 2, \ldots$,
(b) the sequence of continuous time processes $\{{}_hX_t\}$ formed as step functions from the discrete time process in (a) using (2.2) (this process also depends on h and on a (continuous) time index t, $t \geq 0$),
(c) a limiting diffusion process X_t to which, under conditions given below, the sequence of processes $\{{}_hX_t\}_{h \downarrow 0}$ weakly converges.

To accommodate these different processes, we indicate dependence on h to the lower left of X and dependence on the time index to the lower right.

Next, let A' denote the transpose of the matrix A and define the vector/matrix norm $\|A\|$ as

$$\|A\| = \begin{cases} [A'A]^{1/2} & \text{when } A \text{ is a column vector,} \\ [\operatorname{trace}(A'A)]^{1/2} & \text{when } A \text{ is a matrix.} \end{cases} \tag{2.4}$$

For each $h > 0$ and each $\varepsilon > 0$, define

$$a_h(x,t) \equiv h^{-1} \int_{\|y-x\| \le 1} (y-x)(y-x)' \Pi_{h,h[t/h]}(x, \mathrm{d}y), \tag{2.5}$$

$$b_h(x,t) \equiv h^{-1} \int_{\|y-x\| \le 1} (y-x) \Pi_{h,h[t/h]}(x, \mathrm{d}y), \tag{2.6}$$

$$\Delta_{h,\varepsilon}(x,t) \equiv h^{-1} \int_{\|y-x\| \ge \varepsilon} \Pi_{h,h[t/h]}(x, \mathrm{d}y), \tag{2.7}$$

where $[t/h]$ is the integer part of t/h, i.e., the largest integer $k \le t/h$.

The integration in (2.5)–(2.6) is taken over $\|y-x\| \le 1$ rather than over R^n because the usual conditional moments may not be finite.[3] $a_h(x,t)$ and $b_h(x,t)$ are measures of the (truncated) second moment and drift per unit of time, respectively. The convergence results we present below will require that $a_h(x,t)$ and $b_h(x,t)$ converge to finite limits, and that $\Delta_{h,\varepsilon}(x,t)$ goes to zero for all $\varepsilon > 0$. $\Delta_{h,\varepsilon}(x,t)$ is a measure of the probability per unit of time of a jump of size ε or greater. Since diffusion processes have sample paths that are continuous with probability one, it should not be surprising that the probability of discrete jumps of any fixed size ε ($\varepsilon > 0$) or greater must go to zero.

The first convergence result of this section will require the following assumptions:

Assumption 1. There exists a continuous, measurable mapping $a(x,t)$ from $R^n \times [0,\infty)$ into the space of $n \times n$ nonnegative definite symmetric matrices and a continuous, measurable mapping $b(x,t)$ from $R^n \times [0,\infty)$ into R^n such that for

[3] E.g., $X_t = \exp[\exp(W_t)]$, with W_t a Brownian motion, is a diffusion but has no moments of any order over any time interval of positive length.

all $R > 0$, $T > 0$, and $\varepsilon > 0$

$$\lim_{h \downarrow 0} \sup_{\|x\| \le R,\, 0 \le t \le T} \| a_h(x,t) - a(x,t) \| = 0, \tag{2.8}$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \le R,\, 0 \le t \le T} \| b_h(x,t) - b(x,t) \| = 0, \tag{2.9}$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \le R,\, 0 \le t \le T} \Delta_{h,\varepsilon}(x,t) = 0. \tag{2.10}$$

Eqs. (2.8) and (2.9) require that the second moment and drift per unit of time converge uniformly on compact sets to well-behaved functions of time and the state variables x. (2.10) requires that the probability per unit of time of a jump of size ε or greater vanishes uniformly on compacts for all $\varepsilon > 0$, so the sample paths of the limit process are continuous with probability one.

Assumption 2. There exists a continuous, measurable mapping $\sigma(x,t)$ from $R^n \times [0,\infty)$ into the space of $n \times n$ matrices, such that for all $x \in R^n$ and all $t \ge 0$,

$$a(x,t) = \sigma(x,t)\sigma(x,t)'. \tag{2.11}$$

Assumption 2 requires that the function $a(x,t)$, the second moment per unit of time of the limit process, has a well-behaved matrix square root $\sigma(x,t)$. The next assumption requires the probability measures ν_h of the random starting points ${}_hX_0$ to converge to a limit measure ν_0 as $h \downarrow 0$:

Assumption 3. As $h \downarrow 0$, ${}_hX_0$ converges in distribution to a random variable X_0 with probability measure ν_0 on $(R^n, B(R^n))$.

We have specified a probability measure ν_0 for the initial value of the limit process X_t, an instantaneous drift function $b(x,t)$, an instantaneous covariance[4] matrix $a(x,t)$, and have guaranteed that the limit process, if it exists, will have sample paths that are continuous with probability one. At this point, there are two things that can go wrong: first, a limit process may not exist, because when taken together ν_0, $a(x,t)$, and $b(x,t)$ may imply that the process explodes (with positive probability) to infinity in finite time. Second, ν_0, $a(x,t)$, and $b(x,t)$ may not *uniquely* define a limit process. For example, we may need additional information in the form of boundary conditions.[5]

[4]As $h \downarrow 0$, the difference between the conditional covariance matrix and the conditional second moment matrix vanishes, so that in the continuous time limit the two are identical.

[5]For examples of explosion and nonuniqueness, see Stroock and Varadhan (1979, sect. 10.0 and problem 6.7.7).

There is an extensive literature on conditions under which ν_0, $a(x,t)$, and $b(x,t)$ uniquely define a limiting diffusion [e.g., see Stroock and Varadhan (1979)]. Appendix A summarizes a few of these conditions.

Assumption 4. *ν_0, $a(x,t)$, and $b(x,t)$ uniquely specify the distribution of a diffusion process X_t, with initial distribution ν_0, diffusion matrix $a(x,t)$, and drift vector $b(x,t)$.*

Theorem 2.1. *Under Assumptions 1 through 4, the sequence of ${}_hX_t$ processes defined by (2.1)–(2.3) converges weakly (i.e., in distribution) as $h \downarrow 0$ to the X_t process defined by the stochastic integral equation*

$$X_t = X_0 + \int_0^t b(X_s, s)\,\mathrm{d}s + \int_0^t \sigma(X_s, s)\,\mathrm{d}W_{n,s}, \tag{2.12}$$

where $W_{n,t}$ is an n-dimensional standard Brownian motion,[6] independent of X_0, and where for any $\Gamma \in B(R^n)$, $\mathrm{P}(X_0 \in \Gamma) = \nu_0(\Gamma)$. Such an X_t process exists and is distributionally unique. This distribution does not depend on the choice of $\sigma(\cdot,\cdot)$ made in Assumption 2. Finally, X_t remains finite in finite time intervals almost surely, i.e., for all $T > 0$,

$$\mathrm{P}\left[\sup_{0 \le t \le T} \|X_t\| < \infty\right] = 1. \tag{2.13}$$

The convergence in distribution in Theorem 2.1 is not merely a convergence in distribution of $\{{}_hX_t\}$ for a fixed value of t: rather for any T, $0 \le T < \infty$, the probability laws generating the *entire sample paths* $\{{}_hX_t\}$, $0 \le t \le T$, converge to the probability law generating the sample path of X_t, $0 \le t \le T$.[7] From this point on, we will denote this type of convergence in distribution by the symbol $\Rightarrow$, while $\xrightarrow{\mathrm{d}}$ will denote convergence in distribution of random variables in R^1 or R^n.

2.2. An alternate set of conditions

Assumption 1 is cumbersome to verify, since the integrals are taken over $\|y - x\| \le 1$ and $\|y - x\| \ge \varepsilon$. This somewhat limits the usefulness of Theorem

[6] I.e., $W_{n,t}$ is an n-dimensional Brownian motion with, for all t, $\mathrm{E}[W_{n,t}] = W_{n,0} = 0_n$ and $\mathrm{d}W_{n,t}\,\mathrm{d}W'_{n,t} = I_{n,n}\,\mathrm{d}t$, where $I_{n,n}$ is an $n \times n$ identity matrix.

[7] Weak convergence implies, for example, that given any times $t_1, t_2, \dots, t_n > 0$, the joint distributions of $\{{}_hX_{t_1}, {}_hX_{t_2}, \dots, {}_hX_{t_n}\}$ converge to the distribution of $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ as $h \downarrow 0$. More generally, weak convergence implies that if $f(\cdot)$ is a continuous functional of the sample path of ${}_hX_t$, then $f({}_hX_t)$ converges in distribution to $f(X_t)$ as $h \downarrow 0$. See Billingsley (1968).

2.1. We therefore introduce a simpler, though somewhat less general, alternative to Assumption 1. First, for each i, $i = 1, \ldots, n$, each $\delta > 0$, and each $h > 0$, define

$$c_{h,i,\delta}(x,t) \equiv h^{-1} \int_{R^n} |(y-x)_i|^{2+\delta} \Pi_{h,h[t/h]}(x, \mathrm{d}y), \tag{2.14}$$

where $(y-x)_i$ is the ith element of the vector $(y-x)$. If for some $\delta > 0$, and all i, $i = 1, \ldots, n$, $c_{h,i,\delta}(x,t)$ is finite, then the following integrals will be well-defined and finite:

$$a_h^*(x,t) \equiv h^{-1} \int_{R^n} (y-x)(y-x)' \Pi_{h,h[t/h]}(x, \mathrm{d}y), \tag{2.15}$$

$$b_h^*(x,t) \equiv h^{-1} \int_{R^n} (y-x) \Pi_{h,h[t/h]}(x, \mathrm{d}y). \tag{2.16}$$

Like $a(\cdot, \cdot)$ and $b(\cdot, \cdot)$, $a^*(\cdot, \cdot)$ and $b^*(\cdot, \cdot)$ are measures of the second moment and drift per unit of time, with the difference that the integrals are taken over R^n instead of $\|y-x\| \leq 1$.

Assumption 5. There exists a $\delta > 0$ *such that for each* $R > 0$, *each* $T > 0$, *and each* $i = 1, \ldots, n$,

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R,\, 0 \leq t \leq T} c_{h,i,\delta}(x,t) = 0. \tag{2.17}$$

Further, there exists a continuous mapping $a(x,t)$ *from* $R^n \times [0, \infty)$ *into the space of* $n \times n$ *nonnegative definite symmetric matrices and a continuous mapping* $b(x,t)$ *from* $R^n \times [0, \infty)$ *into* R^n *such that*

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R,\, 0 \leq t \leq T} \| a_h^*(x,t) - a(x,t) \| = 0, \tag{2.18}$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R,\, 0 \leq t \leq T} \| b_h^*(x,t) - b(x,t) \| = 0. \tag{2.19}$$

Theorem 2.2. Under Assumptions 2 through 5, the conclusions of Theorem 2.1 hold.[8]

Theorems 2.1 and 2.2 provide a relatively simple way to prove convergence to an Ito process limit. In essence, Theorem 2.2 says that if the drift and

[8]This Theorem was inspired by a standard result for diffusions (as opposed to a sequence of discrete time Markov processes). See Arnold (1974, p. 40).

second moment of ${}_hX_{kh}$ per unit of time converge to well-behaved limits, and if the first differences of ${}_hX_{kh}$ have an absolute moment higher than two that collapses to zero at an approrpiate rate as $h \downarrow 0$, then the ${}_hX_t$ processes converge in distribution to the solution of the stochastic integral equation system (2.12).

2.3. An example: GARCH(1, 1)-M[9]

In discrete time, the GARCH(1,1)-M process of Engle and Bollerslev (1986) for the log of cumulative excess returns Y_t on a portfolio is

$$Y_t = Y_{t-1} + c\sigma_t^2 + \sigma_t Z_t, \tag{2.20}$$

$$\sigma_{t+1}^2 = \omega + \sigma_t^2[\beta + \alpha Z_t^2], \tag{2.21}$$

where $\{Z_t\} \sim$ i.i.d. N(0, 1).

Now we consider the properties of the stochastic difference equation system (2.20)–(2.21) as we partition time more and more finely. We allow the parameters of the system α, β, and ω to depend on h, and make both the drift term in (2.20) and the variance of Z_t proportional to h:

$${}_hY_{kh} = {}_hY_{(k-1)h} + h \cdot c \cdot {}_h\sigma_{kh}^2 + {}_h\sigma_{kh} \cdot {}_hZ_{kh}, \tag{2.22}$$

$${}_h\sigma_{(k+1)h}^2 = \omega_h + {}_h\sigma_{kh}^2[\beta_h + h^{-1}\alpha_h \cdot {}_hZ_{kh}^2], \tag{2.23}$$

and

$$\mathrm{P}[({}_hY_0, {}_h\sigma_0^2) \in \Gamma] = \nu_h(\Gamma) \quad \text{for any} \quad \Gamma \in B(R^2), \tag{2.24}$$

where $\{{}_hZ_{kh}\} \sim$ i.i.d. N(0, h). We also assume that the sequence of measures $\{\nu_h\}_{h \downarrow 0}$ satisfies Assumption 3, and that for each $h \geq 0$, $\nu_h((Y_0, \sigma_0^2)$: $\sigma_0^2 > 0)$ = 1. Finally, we create the continuous time processes ${}_hY_t$ and ${}_h\sigma_t^2$ by

$${}_hY_t \equiv {}_hY_{kh} \quad \text{and} \quad {}_h\sigma_t^2 \equiv {}_h\sigma_{kh}^2 \quad \text{for} \quad kh \leq t < (k+1)h. \tag{2.25}$$

We allow ω, α, and β to depend on h because our object is to discover which sequences $\{\omega_h, \alpha_h, \beta_h\}$ make the $\{{}_h\sigma_t^2, {}_hY_t\}$ process (in which ${}_h\sigma_t^2$ now represents variance per unit of time and $c \cdot {}_h\sigma_t^2$ the risk premium per unit of time) converge in distribution to an Ito process limit as h goes to zero. For all h, however, we require that ω_h, α_h, and β_h be nonnegative, which ensures that the σ_t^2 process remains positive with probability one [Bollerslev (1986)].

[9]Nelson (1988a, ch. 3) noted that no conditionally heteroskedastic diffusion limit had previously been found for GARCH processes, and boldly conjectured that none was possible. Unfortunately, this bold conjecture turned out to be wrong, and this paper corrects the error. As Sandburg (1969) said: 'I can eat crow, but I don't hanker after it.'

The discrete time process (2.22)–(2.24) is clearly Markovian, and the drift per unit of time [conditioned on information at time $(k-1)h$] is given by

$$E\left[h^{-1}\left({}_hY_{kh} - {}_hY_{(k-1)h}\right)|M_{kh}\right] = c \cdot {}_h\sigma_{kh}^2, \tag{2.26}$$

$$E\left[h^{-1}\left({}_h\sigma_{(k+1)h}^2 - {}_h\sigma_{kh}^2\right)|M_{kh}\right] = h^{-1}\omega_h + h^{-1}(\beta_h + \alpha_h - 1)_h\sigma_{kh}^2, \tag{2.27}$$

where M_{kh} is the σ-algebra generated by kh, ${}_hY_0, \ldots, {}_hY_{(k-1)h}$, and ${}_h\sigma_0^2, \ldots, {}_h\sigma_{kh}^2$. For the drift per unit of time to converge as required by Assumption 5, the limits

$$\lim_{h\downarrow 0} h^{-1}\omega_h = \omega \geq 0, \tag{2.28}$$

$$\lim_{h\downarrow 0} h^{-1}(1 - \beta_h - \alpha_h) = \theta, \tag{2.29}$$

must exist and be finite. Note that ω is required to be nonnegative, whereas θ can be of either sign.

The second moment per unit of time is given by

$$E\left[\left({}_h\sigma_{(k+1)h}^2 - {}_h\sigma_{kh}^2\right)^2/h\,|M_{kh}\right]$$

$$= h^{-1}\omega_h^2 + 2h^{-1}\omega_h(\beta_h + \alpha_h - 1)_h\sigma_{kh}^2$$

$$+ h^{-1}(\beta_h + \alpha_h - 1)^2{}_h\sigma_{kh}^4 + 2h^{-1}\alpha_h^2{}_h\sigma_{kh}^4, \tag{2.30}$$

$$E\left[h^{-1}\left({}_hY_{kh} - {}_hY_{(k-1)h}\right)^2|M_{kh}\right] = hc^2{}_h\sigma_{kh}^4 + {}_h\sigma_{kh}^2, \tag{2.31}$$

$$E\left[h^{-1}\left({}_hY_{kh} - {}_hY_{(k-1)h}\right)\left({}_h\sigma_{(k+1)h}^2 - {}_h\sigma_{kh}^2\right)|M_{kh}\right]$$

$$= c \cdot {}_h\sigma_{kh}^2 \cdot \omega_h + c \cdot {}_h\sigma_{kh}^4(\beta_h + \alpha_h - 1). \tag{2.32}$$

Substituting from (2.28)–(2.29) into (2.30)–(2.32) and assuming that

$$\lim_{h\downarrow 0} 2h^{-1}\alpha_h^2 = \alpha^2 > 0, \tag{2.33}$$

exists and is finite, we have

$$\mathrm{E}\left[h^{-1}\left({}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh}\right)^2 | M_{kh}\right] = \alpha^2{}_h\sigma^4_{kh} + o(1), \tag{2.34}$$

$$\mathrm{E}\left[h^{-1}\left({}_hY_{kh} - {}_hY_{(k-1)h}\right)^2 | M_{kh}\right] = {}_h\sigma^2_{kh} + o(1), \tag{2.35}$$

$$\mathrm{E}\left[h^{-1}\left({}_hY_{kh} - {}_hY_{(k-1)h}\right)\left({}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh}\right) | M_{kh}\right] = o(1), \tag{2.36}$$

where the o(1) terms vanish uniformly on compact sets.

It is not difficult to find $\{\omega_h, \alpha_h, \beta_h\}$ sequences satisfying (2.28), (2.29), and (2.33). For example, set $\omega_h = \omega h$, $\beta_h = 1 - \alpha(h/2)^{1/2} - \theta h$, and $\alpha_h = \alpha(h/2)^{1/2}$. It is also straightforward though tedious to verify that the limits of $\mathrm{E}[h^{-1}({}_hY_{kh} - {}_hY_{(k-1)h})^4 | M_{kh}]$ and $\mathrm{E}[h^{-1}({}_h\sigma^2_{(k+1)h} - {}_h\sigma^2_{kh})^4 | M_{kh}]$ exist and converge to zero, so that with $\delta = 2$,

$$b(Y, \sigma^2) \equiv \begin{bmatrix} c \cdot \sigma^2 \\ \omega - \theta\sigma^2 \end{bmatrix}, \tag{2.37}$$

$$a(Y, \sigma^2) \equiv \begin{bmatrix} \sigma^2 & 0 \\ 0 & \alpha^2\sigma^4 \end{bmatrix}, \tag{2.38}$$

and with α_h, β_h, and ω_h satisfying (2.28), (2.29), and (2.33), Assumption 5 holds. Setting $\sigma(\cdot, \cdot)$ in Assumption 2 equal to the element-by-element square root of $a(Y, \sigma^2)$, Assumption 2 holds as well. (2.37)–(2.38) suggest a limit diffusion of the form

$$\mathrm{d}Y_t = c \cdot \sigma_t^2\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_{1,t}, \tag{2.39}$$

$$\mathrm{d}\sigma_t^2 = \left(\omega - \theta\sigma_t^2\right)\mathrm{d}t + \alpha\sigma_t^2\,\mathrm{d}W_{2,t}, \tag{2.40}$$

$$\mathrm{P}\left[\left(Y_0, \sigma_0^2\right) \in \Gamma\right] = \nu_0(\Gamma) \quad \text{for any} \quad \Gamma \in B(R^2), \tag{2.41}$$

where $W_{1,t}$ and $W_{2,t}$ are independent standard Brownian motions, independent of the initial values (Y_0, σ_0^2).

Unfortunately, none of the conditions A–D in appendix A directly apply to this diffusion. To verify distributional uniqueness, it helps to define $V_t \equiv \ln(\sigma_t^2)$ and apply Ito's Lemma to (2.39)–(2.41):

$$\mathrm{d}Y_t = c \cdot \exp(V_t)\,\mathrm{d}t + \exp(V_t/2)\,\mathrm{d}W_{1,t}, \tag{2.39'}$$

$$\mathrm{d}V_t = \left[\omega \cdot \exp(-V_t) - \theta - \alpha^2/2\right]\mathrm{d}t + \alpha\,\mathrm{d}W_{2,t}, \tag{2.40'}$$

$$\mathrm{P}\left[\left(Y_0, \exp(V_0)\right) \in \Gamma\right] = \nu_0(\Gamma) \quad \text{for any} \quad \Gamma \in B(R^2). \tag{2.41'}$$

It is easy to check that condition B and the nonexplosion condition in appendix A hold for (2.39′)–(2.41′).[10] Since distributional uniqueness holds for (2.39′)–(2.41′), it must by the Continuous Mapping Theorem [Billingsley (1968)] hold for (2.39)–(2.41) as well.[11] We now apply Theorem 2.2 to conclude that $\{{}_hY_t, {}_h\sigma_t^2\} \Rightarrow (Y_t, \sigma_t^2)$ as $h \downarrow 0$.

While no closed form exists for the stationary distribution of GARCH(1, 1) in discrete time, an application of the results of Wong (1964) allows us to solve for the stationary distribution of σ_t^2 implied by (2.40):

Theorem 2.3. *For each $t \geq 0$, define the conditional precision process $\lambda_t \equiv \sigma_t^{-2}$, where σ_t^2 is generated by the stochastic differential equation system (2.39)–(2.41). Then if $2\theta/\alpha^2 > -1$ and $\omega > 0$,*

$$\lambda_t \xrightarrow{\text{d}} \Gamma(1 + 2\theta/\alpha^2, 2\omega/\alpha^2) \quad as \quad t \to \infty, \tag{2.42}$$

where $\Gamma(\cdot,\cdot)$ is the Gamma distribution.[12,13] *If $\lambda_0 \sim \Gamma(1 + 2\theta/\alpha^2, 2\omega/\alpha^2)$, then the σ_t^2 and λ_t processes generated by (2.39)–(2.41) are strictly stationary, and for all $t \geq 0$,*

$$\lambda_t \sim \Gamma(1 + 2\theta/\alpha^2, 2\omega/\alpha^2). \tag{2.43}$$

If (a) the distribution of ${}_h\lambda_0 \equiv {}_h\sigma_0^{-2}$ converges to a $\Gamma(1 + 2\theta/\alpha^2, 2\omega/\alpha^2)$ as $h \downarrow 0$, (b) the sequence $\{\omega_h, \alpha_h, \beta_h\}_{h \downarrow 0}$ satisfies (2.28), (2.29), and (2.31), and

[10] For the Liapunov function $\varphi(Y,V)$ in the nonexplosion condition, use $\varphi(Y,V) \equiv K + f(Y) \cdot |Y| + f(V) \cdot \exp(|V|)$, where $f(x) \equiv \exp(-1/|x|)$ if $x \neq 0$ and $\equiv 0$ otherwise. $\varphi(\cdot,\cdot)$ is arbitrarily differentiable, despite the presence of the $|V|$ and $|Y|$ terms. To verify the inequality (A.7), use the fact that for large Y and V, $\partial\varphi(Y,V)/\partial Y \approx \operatorname{sign}(Y)$, $\partial^2\varphi(Y,V)/\partial Y^2 = 0$, $\partial\varphi(Y,V)/\partial V \approx \operatorname{sign}(V) \cdot \exp(|V|)$, and $\partial^2\varphi(Y,V)/\partial V^2 \approx \exp(|V|)$.

[11] That is, as long as $\nu_h((Y_0, \sigma_0^2)\colon \sigma_0^2 > 0) = 1$, which we have already assumed. Using a simple Taylor series argument, it is also easy to check that Assumption 1 holds for the transformed variables ${}_hV_{kh} \equiv \ln({}_h\sigma_{kh}^2)$ and ${}_hY_{kh}$: i.e., the jump sizes go to zero and the local drifts and second moments converge to the limits implied by (2.39′)–(2.41′).

[12] $x \sim \Gamma(r,s)$ means that the probability density function for x is given by $f(x) = s^r \cdot x^{(r-1)} \cdot \exp(-sx)/\Gamma(r)$ for $x > 0$, where $\Gamma(\cdot)$ is the gamma function.

[13] When $\omega = 0$, we have $\mathrm{d}[\ln(\sigma_t^2)] = -(\alpha^2/2 + \theta)\mathrm{d}t + \alpha\,\mathrm{d}W_t$, so that $\sigma_t^2 \to 0$, $\sigma_t^2 \to \infty$, or $\ln(\sigma_t^2)$ is a driftless Brownian motion as $(2\theta/\alpha^2 + 1) > 0$, < 0, or $= 0$, respectively. This is the continuous time analog to the results of Nelson (1988b) for GARCH(1, 1) in discrete time. The stationarity condition $2\theta/\alpha^2 > -1$ is also the continuous time analog of the discrete time stationarity condition in Nelson (1988b).

(c) $2\theta/\alpha^2 > -1$, *and* $\omega > 0$, *then in the discrete time system (2.22)–(2.24),*

$$ {}_h\sigma_{kh}^{-2} \xrightarrow{\mathrm{d}} \Gamma(1 + 2\theta/\alpha^2, 2\omega/\alpha^2), \tag{2.44}$$

$$ h^{-1/2}{}_hZ_{kh} \cdot {}_h\sigma_{kh} \cdot \left[(2\theta + \alpha^2)/2\omega\right]^{1/2} \xrightarrow{\mathrm{d}} t(2 + 4\theta/\alpha^2), \tag{2.45}$$

for any constant value of kh as $h \downarrow 0$, *where* $t(2 + 4\theta/\alpha^2)$ *is the Student t distribution with* $2 + 4\theta/\alpha^2$ *degrees of freedom.*

Finally, if (b) and (c) are satisfied and in addition there exists a $d > 0$ *such that* $\limsup \mathrm{E}[{}_h\sigma_0^{2d}] < \infty$, *then*

$$ {}_h\sigma_{kh}^{-2} \xrightarrow{\mathrm{d}} \Gamma(1 + 2\theta/\alpha^2, 2\omega/\alpha^2), \tag{2.46}$$

$$ h^{-1/2}{}_hZ_{kh} \cdot {}_h\sigma_{kh} \cdot \left[(2\theta + \alpha^2)/2\omega\right]^{1/2} \xrightarrow{\mathrm{d}} t(2 + 4\theta/\alpha^2), \tag{2.47}$$

as $h \downarrow 0$ *and* $kh \to \infty$.[14]

Theorem 2.3 tells us that although the innovations process ${}_hZ_{kh} \cdot {}_h\sigma_{kh}$ is conditionally normal, its unconditional distribution is approximately Student t when h is small and kh is large. The degrees of freedom term $(2 + 4\theta/\alpha^2)$ is also instructive: the Student t distribution has a finite variance when it has more than two degrees of freedom, which it does in the GARCH(1, 1) limit case only when $\theta > 0$. The case $\theta = 0$ corresponds to the diffusion limit of the IGARCH(1, 1) model of Engle and Bollerslev (1986), which separates the covariance stationary GARCH(1, 1) case ($\theta > 0$) from the strictly stationary but not covariance stationary GARCH(1, 1) model of Nelson (1988b) and Sampson (1988) ($\theta < 0$). In discrete time, the condition for the strict stationarity of the conditionally normal GARCH(1, 1) is that $\mathrm{E}[\ln(\beta_h + \alpha_h Z^2)] < 0$, $Z \sim \mathrm{N}(0, 1)$ [Nelson (1988b)], whereas in continuous time this requirement is that $(2 + 4\theta/\alpha^2) > 0$ (i.e., the t distribution has strictly positive, though not necessarily integer, degrees of freedom).

3. A class of univariate Ito process approximations

In this section we develop a class of ARCH processes based on the Exponential ARCH model of Nelson (1989), and use Theorem 2.2 to show

[14] Unfortunately, Theorems 2.1 and 2.2 do not allow us to simultaneously take $t \to \infty$ and $h \downarrow 0$, so that (2.46) and (2.47) do not follow directly from Theorem 2.1. However, Kushner (1984, ch. 6) gives conditions for the steady state (i.e., $t = \infty$) distributions of the ${}_hX_t$ processes to converge to the steady state distribution of X_t. In appendix B, Kushner's results are employed to prove (2.46) and (2.47).

how these processes can approximate a wide variety of stochastic differential equations.

3.1. Exponential ARCH

In a conditionally normal Exponential ARCH model, we have an innovations process ε_t and its conditional variance σ_t^2 given by

$$\varepsilon_t = \sigma_t Z_t, \tag{3.1}$$

$$Z_t \sim \text{i.i.d. } \mathrm{N}(0,1), \tag{3.2}$$

$$\ln(\sigma_t^2) = \alpha_t + \sum_{k=1}^{\infty} \beta_k g(Z_{t-k}), \qquad \beta_1 \equiv 1, \tag{3.3}$$

where $\{\alpha_t\}_{t=-\infty,\infty}$ and $\{\beta_k\}_{k=1,\infty}$ are real, nonstochastic scalar sequences, and

$$g(Z_t) \equiv \theta Z_t + \gamma\left[|Z_t| - (2/\pi)^{1/2}\right], \tag{3.4}$$

i.e., the log of σ_t^2 is linear in lagged Z_t's, $|Z_t|$'s, and some function of time. $g(Z_t)$ is a zero mean, i.i.d. innovations term, with two zero mean, serially and contemporaneously uncorrelated components, θZ_t and $\gamma[|Z_t| - (2/\pi)^{1/2}]$. The two components allow any desired degree of conditional correlation between Z_t and the change in $\ln(\sigma_t^2)$. In GARCH models, this conditional correlation is always fixed at zero. Black (1976) noted that changes in the level of the stock market are negatively correlated with changes in its volatility. The two terms in $g(\cdot)$ allow Exponential ARCH to capture this correlation in a way that GARCH cannot, since in GARCH models the change in σ_t^2 is driven by Z_t^2, which is uncorrelated with Z_t.

3.2. ARCH diffusion approximations

The class of diffusion approximations that we now propose is a variant of the Exponential ARCH model (3.1)–(3.4). Like Exponential ARCH, this class of models uses Z_t and $|Z_t|$ as the driving noise processes. Unlike Exponential ARCH however, the log-linearity assumption for σ_t^2 is relaxed, while a finite-order Markov representation for the system is imposed, ruling out, for example, a fractionally differenced [Hosking (1981)] representation for $\ln(\sigma_t^2)$.

First, define the stochastic differential equation system

$$\mathrm{d}S_t = f(S_t, Y_t, t)\,\mathrm{d}t + g(S_t, Y_t, t)\,\mathrm{d}W_{1,t}, \tag{3.5}$$

$$\mathrm{d}Y_t = F(S_t, Y_t, t)\,\mathrm{d}t + G(S_t, Y_t, t)\,\mathrm{d}W_{2,t}, \tag{3.6}$$

$$\begin{bmatrix} \mathrm{d}W_{1,t} \\ \mathrm{d}W_{2,t} \end{bmatrix} [\mathrm{d}W_{1,t} \quad \mathrm{d}W_{2,t}'] = \begin{bmatrix} 1 & \Omega_{1,2} \\ \Omega_{2,1} & \Omega_{2,2} \end{bmatrix} \mathrm{d}t \equiv \Omega\,\mathrm{d}t, \tag{3.7}$$

where Ω is an $(n+1) \times (n+1)$ positive semi-definite matrix of rank two or less, Y is an n-dimensional vector of (unobservable) state variables, S is an (observable) scalar process, W_1 is a one-dimensional standard Brownian motion, and W_2 is an n-dimensional Brownian motion. $f(\cdot,\cdot,\cdot)$ and $g(\cdot,\cdot,\cdot)$ are real-valued, continuous scalar functions, and $F(\cdot,\cdot,\cdot)$ and $G(\cdot,\cdot,\cdot)$ are real, continuous $n \times 1$ and $n \times n$ valued functions, respectively. We also assume that (S_0, Y_0) are random variables with joint probability measure ν_0, and are independent of the W_1 and W_2 processes. Define the vector and matrix functions b and a by

$$b(s,y,t) \equiv [f(s,y,t) \quad F(s,y,t)']', \tag{3.8}$$

$$a(s,y,t) = \begin{bmatrix} g^2 & g\Omega_{1,2}G' \\ G\Omega_{2,1}g & G\Omega_{2,2}G' \end{bmatrix}. \tag{3.9}$$

$b(\cdot,\cdot,\cdot)$ is an $(n+1) \times 1$ vector and $a(\cdot,\cdot,\cdot)$ is an $(n+1) \times (n+1)$ matrix.

Next we propose a sequence of approximating processes that converge to (3.5)–(3.7) in distribution as $h \downarrow 0$. We model ${}_hS_t$ and ${}_hY_t$ as step functions with jumps at times h, $2h$, $3h$, and so on. The jumps are given by

$${}_hS_{(k+1)h} = {}_hS_{kh} + f({}_hS_{kh}, {}_hY_{kh}, kh)h + g({}_hS_{kh}, {}_hY_{kh}, kh)\,{}_hZ_{kh}, \tag{3.10}$$

$${}_hY_{(k+1)h} = {}_hY_{kh} + F({}_hS_{kh}, {}_hY_{kh}, kh)h + G({}_hS_{kh}, {}_hY_{kh}, kh)\,{}_hZ^*_{kh}, \tag{3.11}$$

where

$${}_hZ_{kh} \sim \text{i.i.d. } \mathrm{N}(0,h), \tag{3.12}$$

$${}_hZ^*_{kh} \equiv \begin{bmatrix} \theta_1 \cdot {}_hZ_{kh} + \gamma_1\left[|{}_hZ_{kh}| - (2h/\pi)^{1/2}\right] \\ \cdots \\ \theta_n \cdot {}_hZ_{kh} + \gamma_n\left[|{}_hZ_{kh}| - (2h/\pi)^{1/2}\right] \end{bmatrix}, \tag{3.13}$$

and $\{\theta_1, \gamma_1, \ldots, \theta_n, \gamma_n\}$ are selected so that

$$\mathrm{E}\begin{bmatrix} {}_hZ_{kh} \\ {}_hZ^*_{kh} \end{bmatrix}[{}_hZ_{kh} \quad {}_hZ^{*\prime}_{kh}] = \Omega h. \tag{3.14}$$

We convert the discrete time process $[{}_hS_{kh}, {}_hY'_{kh}]$ into a continuous time process by defining

$${}_hS_t \equiv {}_hS_{kh}, \quad {}_hY_t \equiv {}_hY_{kh} \quad \text{for} \quad kh \leq t < (k+1)h. \tag{3.15}$$

Theorem 3.1. If $b(s, y, t)$ and $a(s, y, t)$ satisfy Assumption 4, with $x \equiv [s, y']'$, and if the joint probability measures ν_h of the starting values $({}_hS_0, {}_hY'_0)$ converge to the measure ν_0 as $h \downarrow 0$, then $({}_hS_t, {}_hY'_t) \Rightarrow (S_t, Y'_t)$ as $h \downarrow 0$.

The proof of Theorem 3.1 is a straightforward application of Theorem 2.2. The details are in appendix B. Intuitively, $({}_hS_{(k+1)h} - {}_hS_{kh})$, $({}_hY_{(k+1)h} - {}_hY_{kh})$, and h are the discrete time counterparts of $\mathrm{d}S$, $\mathrm{d}Y$, and $\mathrm{d}t$, respectively. It is a little bit harder to see why ${}_hZ_{kh}$ and ${}_hZ^*_{kh}$ are the discrete time counterparts of $\mathrm{d}W_{1,t}$ and $\mathrm{d}W_{2,t}$. To see that they are, the following result may be helpful:

Theorem 3.2. Let W_t be a standard Brownian motion on $[0,1]$. *For each $k = 1, 2, \ldots$ and each $t \in [0,1]$ define $Q_{k,t}$ by*

$$Q_{k,t} \equiv (1 - 2/\pi)^{-1/2} \sum_{j=1}^{[kt]} \left\{ |W_{(j+1)/k} - W_{j/k}| - (2/\pi k)^{1/2} \right\}, \tag{3.16}$$

where $[kt]$ is the integer part of kt. Then as $k \to \infty$,

$$[Q_{k,t}, W_t] \Rightarrow W_t^{**}, \tag{3.17}$$

where W_t^{**} is a two-dimensional standard Brownian motion on $[0,1]$.[15]

Although $Q_{k,t}$ is a function of the path of W_t, it converges in distribution to a Brownian motion independent of W_t as $k \to \infty$. To manufacture a

[15]Richard Dudley pointed out that $Q_{k,t}$ does not converge in probability to any Brownian motion, since as $k \to \infty$, $[Q_{k,t}, Q_{k^2,t}, W_t] \Rightarrow W_t^{***}$, a three-dimensional standard Brownian motion. In other words, the Q function corresponding to a fine partition of time is asymptotically independent of the Q function corresponding to a partition that is much finer still.

sequence of processes that converges to a Brownian motion that is imperfectly correlated with W_t, we take a linear combination of W_t and $Q_{k,t}$. Heuristically, this is why ${}_hZ_{kh}$ and ${}_hZ_{kh}^*$ can serve as the discrete time counterparts of $\mathrm{d}W_{1,t}$ and $\mathrm{d}W_{2,t}$, respectively. Though ${}_hZ_{kh}$ and $|{}_hZ_{kh}| - \mathrm{E}|{}_hZ_{kh}|$ are uncorrelated but not independent for each $h > 0$, their partial sums are independent in the limit as $h \downarrow 0$. Thus, although there is only one observable random process $({}_hS_{kh})$, we are able to approximate continuous time systems driven by more than one Brownian motion.

One note of caution is in order on the use of the diffusion approximation just developed: Suppose that the true data-generating process in some model is a diffusion of the form (3.5)–(3.7) over a time span $[0, T]$, and that we observe the diffusion only at discrete intervals of length h, and wish to estimate the diffusion's coefficients. Theorem 3.1 takes θ, γ, f, g, F, and G as nonrandom, and does *not* tell us that we can fit a stochastic difference equation model of the form (3.10)–(3.14) to the observed data and consistently estimate the parameters of the diffusion as $h \downarrow 0$ and $T \to \infty$. There may be general conditions under which this procedure yields consistent parameter estimates, but these conditions are as yet unknown.

For the sake of simplicity, we have restricted the rank of Ω to be less than or equal to two. This allowed us to have only two driving noise terms, ${}_hZ_{kh}$ and $|{}_hZ_{kh}|$. We could easily allow Ω to have a higher rank by considering arbitrary piecewise linear functions of ${}_hZ_{kh}$.

Another easy generalization of Theorem 3.1 is to relax the requirement that the ${}_hZ_{kh}$ are $\mathrm{N}(0, h)$. We could instead allow $h^{-1/2}{}_hZ_{kh}$ to be i.i.d. from any symmetric distribution (the same for each h) with mean zero, variance one, and a finite absolute moment of order higher than two. We then substitute $\mathrm{E}|h^{-1/2}{}_hZ_{kh}|$ for $(2/\pi)^{1/2}$ in (3.13).

The approximation scheme (3.10)–(3.15) is closely related to the Euler approximation for stochastic differential equations discussed, for example, in Duffie and Singleton (1989) and Pardoux and Talay (1985). There are two differences between the Euler approximation and the approximation (3.10)–(3.15): First, the Euler approximation does not include the $|{}_hZ_{kh}|$ terms in (3.10)–(3.15).[16] Second, the Euler approximation assumes global Lipschitz continuity of f, F, g, and G, while Theorem 3.1 does not. In the Euler approximation, however, the convergence result analogous to Theorem 3.1 is stronger in that the approximating sequence converges in probability rather than in distribution. The assumption of Lipschitz continuity, however, rules out several interesting cases, including the next example:

[16]The referee pointed out that in a simulation context, inclusion of the $|{}_hZ_{kh}|$ terms can be interpreted as a trick to reduce the number of pseudo-random numbers needed to carry out the simulation.

3.3. AR(1) Exponential ARCH

Next, consider a simple model of changing volatility on a portfolio, in which the log of the conditional variance follows a continuous time AR(1) [i.e., an Ornstein–Uhlenbeck process (Arnold (1974))]. Variations on the basic model are found in Wiggins (1987), Nelson (1989), and Gennotte and Marsh (1987):

$$d[\ln(S_t)] = \theta\sigma_t^2\, dt + \sigma_t\, dW_{1,t}, \tag{3.18}$$

$$d[\ln(\sigma_t^2)] = -\beta[\ln(\sigma_t^2) - \alpha]\, dt + dW_{2,t}, \tag{3.19}$$

$$P[(\ln(S_0), \ln(\sigma_0^2)) \in \Gamma] = \nu_0(\Gamma) \quad \text{for any} \quad \Gamma \in B(R^2), \tag{3.20}$$

where S_t is the value of the portfolio at time t, $W_{1,t}$, and $W_{2,t}$ are Brownian motions with

$$\begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} [dW_{1,t} \quad dW_{2,t}] = \begin{bmatrix} 1 & C_{12} \\ C_{12} & C_{22} \end{bmatrix} dt \equiv C\, dt, \tag{3.21}$$

and $C_{22} \geq C_{12}^2$. σ_t^2 is the instantaneous variance per unit of time and $\theta\sigma_t^2$ is the instantaneous risk premium.[17]

The diffusion (3.18)–(3.21) does not satisfy global Lipschitz conditions and cannot be transformed to satisfy them. Hence, the standard theorems on convergence of the Euler approximation do not apply. Theorem 3.1 can be applied, however, to find a sequence of ARCH models that converge weakly (rather than path by path) to (3.18)–(3.21). Since $\ln(\sigma_t^2)$ follows a continuous time AR(1) process in (3.19), our discrete time models will also make $\{\ln({}_h\sigma_{kh}^2)\}_{k=0,\infty}$ an AR(1) process. Specifically, we assume that ${}_h\sigma_{kh}^2$ follows an AR(1) Exponential ARCH process [Nelson (1989)]. For each h we have

$$\ln[{}_hS_{kh}] = \ln[{}_hS_{(k-1)h}] + h\theta_h\,{}_h\sigma_{kh}^2 + {}_h\sigma_{kh} \cdot {}_hZ_{kh}, \tag{3.22}$$

$$\begin{aligned} \ln[{}_h\sigma_{(k+1)h}^2] = \ln({}_h\sigma_{kh}^2) - \beta[\ln({}_h\sigma_{kh}^2) - \alpha]h + C_{12} \cdot {}_hZ_{kh} \\ + \gamma[|{}_hZ_{kh}| - (2h/\pi)^{1/2}], \end{aligned} \tag{3.23}$$

$$P[(\ln({}_hS_0), \ln({}_h\sigma_0^2)) \in \Gamma] = \nu_h(\Gamma) \quad \text{for any} \quad \Gamma \in B(R^2), \tag{3.24}$$

[17] The conditions under which this is the equilibrium risk premium are very restrictive, however – see Gennotte and Marsh (1987).

where $\gamma \equiv [(C_{22} - C_{12}^2)/(1 - 2/\pi)]^{1/2}$ and ${}_hZ_{kh} \sim$ i.i.d. $\mathrm{N}(0, h)$. It is easy to check that

$$\mathrm{E}\begin{bmatrix} {}_hZ_{kh} \\ C_{12}\cdot {}_hZ_{kh} + \gamma\left[|{}_hZ_{kh}| - (2h/\pi)^{1/2}\right] \end{bmatrix}$$

$$\times\left[{}_hZ_{kh} \quad C_{12}\cdot {}_hZ_{kh} + \gamma\left[|{}_hZ_{kh}| - (2h/\pi)^{1/2}\right]\right]$$

$$= h\begin{bmatrix} 1 & C_{12} \\ C_{12} & C_{22} \end{bmatrix}, \tag{3.25}$$

which is the discrete time analog of (3.21). Once again, we create the continuous time step function processes ${}_hS_t$ and ${}_h\sigma_t^2$. As before, $\ln({}_hS_{kh}) - \ln({}_hS_{(k-1)h})$, $\ln[{}_h\sigma^2_{(k+1)h}] - \ln({}_h\sigma^2_{kh})$, ${}_hZ_{kh}$, $(C_{12}\cdot {}_hZ_{kh} + \gamma[|{}_hZ_{kh}| - (2h/\pi)^{1/2}])$, and h are the discrete time counterparts of $\mathrm{d}(\ln(S))$, $\mathrm{d}(\ln(\sigma^2))$, $\mathrm{d}W_{1,t}$, $\mathrm{d}W_{2,t}$, and $\mathrm{d}t$, respectively.

Theorem 3.3. If $\nu_h \to \nu_0$, then $\{{}_hS_t, {}_h\sigma_t^2\} \Rightarrow \{S_t, \sigma_t^2\}$.

Recall from section 2 that the continuous time limit of the GARCH(1,1) process has an inverted Gamma stationary distribution, so that although the normalized discrete time innovations process $h^{-1/2}{}_hZ_{kh}\cdot {}_h\sigma_{kh}$ is normally distributed given ${}_h\sigma_{kh}$, its unconditional distribution is approximately Student t when h is small and kh is large. In the Exponential ARCH case, if $\ln(\sigma_0^2)$ is normally distributed, then the $\ln(\sigma_t^2)$ process is Gaussian, and even if $\ln(\sigma_0^2)$ is not normally distributed, then a Gaussian stationary limit distribution for $\ln(\sigma_t^2)$ exists as long as $\beta > 0$ [Arnold (1974, sect. 8.3)].

Theorem 3.4. For a fixed initial condition ${}_h\sigma_0^2 \equiv \sigma_0^2$ for all h, and for any finite β,

$$h^{-1/2}{}_hZ_{kh}\cdot {}_h\sigma_{kh} \xrightarrow{\mathrm{d}} Z\cdot\sigma_{kh}, \tag{3.26}$$

for fixed kh and $h \downarrow 0$, where $Z \sim \mathrm{N}(0,1)$ independent of σ_{kh}, and

$$\ln(\sigma_{kh}) \sim \mathrm{N}\left(\mathrm{e}^{-\beta kh}(\ln(\sigma_0) - \alpha/2) + \alpha/2, C_{22}(1 - \mathrm{e}^{-2\beta kh})/8\beta\right). \tag{3.27}$$

If $\beta > 0$, then in the discrete time model (3.22)–(3.25),

$$h^{-1/2}{}_hZ_{kh}\cdot{}_h\sigma_{kh} \xrightarrow{\mathrm{d}} Z\cdot\sigma, \tag{3.28}$$

as $h\downarrow 0$ and $kh\to\infty$, where Z and σ are independent random variables with

$$Z\sim \mathrm{N}(0,1), \qquad \ln(\sigma)\sim \mathrm{N}(\alpha/2, C_{22}/8\beta). \tag{3.29}$$

In contrast to the GARCH(1, 1) model, in which the innovations terms are approximately Student when the time interval between observations is short, the Exponential ARCH model produces innovations that are approximately a Normal–Lognormal mixture. The Normal–Lognormal mixture has been used in modelling stock returns by a number of researchers, e.g., Clark (1973). In Clark's model, however, the conditional variance process was assumed to be i.i.d., whereas it is serially correlated in the Exponential ARCH case.

It is a simple exercise to apply the results of this section to show that Exponential ARCH processes, in which $\ln(\sigma^2)$ is either a finite-order AR or else is the sum of a finite number of finite-order AR processes, have well-defined continuous time limits. For a discussion of continuous time AR processes and their relation to discrete time AR processes, see Priestley (1981, ch. 3, sects. 3.7.4, 3.7.5). Passage to continuous time of a sequence of Exponential ARCH models in which $\ln(\sigma_t^2)$ follows an AR(k) is directly analogous to the case discussed in Priestley.

3.4. GARCH(1, 1) revisited

As a final example, consider again the Ito process limit of the GARCH(1, 1) model discussed in section 2. As before, define $V_t \equiv \ln(\sigma_t^2)$. Applying Ito's Lemma, the limiting diffusion becomes (2.39′)–(2.41′). Our discrete time system is now

$${}_hY_{kh} = {}_hY_{(k-1)h} + h\cdot c\cdot \exp({}_hV_{kh}) + \exp({}_hV_{kh}/2)\cdot{}_hZ_{kh}, \tag{3.30}$$

$$\begin{aligned}{}_hV_{(k+1)h} = {}_hV_{kh} &+ h\left[\omega\cdot\exp(-{}_hV_{kh}) - \theta - \alpha^2/2\right]\\ &+ \alpha\left[|{}_hZ_{kh}| - (2h/\pi)^{1/2}\right]\left[1-2/\pi\right]^{-1/2},\end{aligned} \tag{3.31}$$

$$\mathrm{P}\left[({}_hY_0, \exp({}_hV_0))\in\Gamma\right] = \nu_h(\Gamma) \quad \text{for any} \quad \Gamma\in B(R^2), \tag{3.32}$$

where $\{_h Z_{kh}\} \sim$ i.i.d. $N(0, h)$. As in earlier examples, we create the continuous time step function processes $_hY_t$ and $_hV_t$ from the discrete time processes $_hY_{kh}$ and $_hV_{kh}$.

Theorem 3.5. *If* $\nu_h \to \nu_0$ *as* $h \downarrow 0$, *then* $[_hY_t, {}_hV_t] \Rightarrow [Y_t, V_t]$.

4. Conclusion

In this paper, we have presented some basic tools for investigating the relationship between stochastic difference equations and Ito processes, and have made two concrete applications: First, we have derived the approximate distributions of GARCH(1, 1) and AR(1) Exponential ARCH for small sampling intervals. Second, we have introduced a class of ARCH models (in which Exponential ARCH is a special case) which can approximate a wide variety of stochastic differential equations.

In recent years, researchers in many disciplines have developed and applied methods of approximating Ito processes with stochastic difference equations (and visa versa). For example, see Kushner (1984) for applications to signal processing and Ethier and Kurtz (1986, ch. 10) for applications to genetics. Apart from the options pricing literature [e.g., Cox, Ross, and Rubinstein (1979)], however, economists have made relatively little use of these tools, though they can be very useful, especially in understanding nonlinear time series models such as ARCH.

While the results in this paper contribute to our understanding of the relation between stochastic difference and differential equations, several limitations that have yet to be overcome. For example, it would be useful to extend the results of the paper to jump-diffusion processes. Results on jump-diffusion approximation have been developed, e.g., by Kushner (1984) and Ethier and Kurtz (1986). It would be interesting to see how these could be integrated into the ARCH framework. A second generalization would be to extend the results of section 3 to multivariate systems. Perhaps most useful would be to find circumstances under which we can use the diffusion approximations of section 3 to estimate consistently the parameters of discretely sampled diffusions. These tasks await future research.

Appendix A: Distributional uniqueness

In this appendix, we state conditions under which ν_0, $a(x,t)$, and $b(x,t)$ uniquely specify the distribution of a stochastic integral equation of the form given in Theorem 2.1. The conditions are of two types: first, those that ensure distributional uniqueness of the process on compact sets, and second, a condition that ensures that the limit process doesn't fly off to infinity in finite time. The assumptions made in Theorem A.1 below to ensure distributional uniqueness are considerably less restrictive than the Lipschitz and Growth conditions that are often assumed [e.g., Liptser and Shiryayev (1977, theorem 4.6, corollary)] to ensure strong (pathwise) existence and uniqueness of solutions to stochastic differential equations. Our conditions are less restrictive largely because we are interested only in distributional existence and uniqueness. The source of these results is Stroock and Varadhan (1979) (henceforth S&V). For a much fuller treatment of these and related conditions, see S&V and Ethier and Kurtz (1986) (henceforth E&K).

First, we present several conditions, any one of which ensures weak existence and uniqueness of the process X_t on compact sets:

Condition A. *Let $a(x,t)$ and $b(x,t)$ be continuous in both x and t with two continuous partial derivatives with respect to x.*

Condition B. *Let $a(x,t)$ and $b(x,t)$ be locally bounded [i.e., bounded on bounded (x,t) sets] and measurable, and let $a(x,t)$ be continuous in x. Let I_n be the $n \times n$ identity matrix. For every $R > 0$, and every $T > 0$, let there be a number $\Lambda_{T,R} > 0$ such that for all (x,t) satisfying $0 \le t \le T$ and $\|x\| \le R$, $a(x,t) - \Lambda_{T,R} \cdot I_n$ is positive definite.*

Condition C. *Let $a(x,t)$ and $b(x,t)$ be locally bounded and measurable, and let $\sigma(x,t)$ be a locally bounded, measurable function satisfying $a(x,t) = \sigma(x,t)'\sigma(x,t)$. For every $R > 0$ and every $T > 0$, let there be a number $\Lambda_{T,R} > 0$ such that*

$$\sup_{0 \le t \le T,\, \|x\| \le R,\, \|y\| \le R} \|\sigma(y,t) - \sigma(x,t)\|$$

$$+ \|b(y,t) - b(x,t)\| - \Lambda_{T,R}\|y - x\| \le 0. \tag{A.1}$$

Condition D. *Let x be a scalar (i.e., $n = 1$), let $a(x,t)$ and $b(x,t)$ be locally bounded and measurable, and let $\sigma(x,t) \equiv a(x,t)^{1/2}$. Let there be an increasing, nonnegative function $\rho(u)$ from $[0,\infty)$ into $[0,\infty)$ such that*

$$\rho(u) > 0 \quad \textit{for} \quad u > 0, \tag{A.2}$$

$$\lim_{\varepsilon \downarrow 0} \int_{\varepsilon}^{1} [\rho(u)]^{-2}\, \mathrm{d}u = \infty. \tag{A.3}$$

For every $R>0$ and ever $T>0$, let there be a number $\Lambda_{T,R}>0$ such that

$$\sup_{|x|\leq R,\,|y|\leq R,\,0\leq t\leq T} |\sigma(x,t)-\sigma(y,t)|-\Lambda_{T,R}\cdot\rho(|x-y|)\leq 0, \qquad (A.4)$$

$$\sup_{|x|<R,\,|y|<R,\,0\leq t\leq T} |b(x,t)-b(y,t)|-\Lambda_{T,R}\cdot|x-y|\leq 0. \qquad (A.5)$$

Next, we state a condition that ensures that the limit process does not run off to infinity in finite time. S&V Theorem 10.2.3 provides a second (somewhat more complicated) such condition.

Nonexplosion condition. There exists a nonnegative function $\varphi(x,t)$ that is differentiable with respect to t and twice differentiable with respect to x such that for each $T>0$,

$$\lim_{\|x\|\to\infty}\ \inf_{0\leq t\leq T} \varphi(x,t)=\infty, \qquad (A.6)$$

and there exists a positive, locally bounded function $\lambda(T)$ such that for each $T>0$, all $x\in R^n$ and all t, $0\leq t\leq T$,

$$\sum_{i=1}^{n} b_i(x,t)\frac{\partial\varphi(x,t)}{\partial x_i}+\tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(x,t)\frac{\partial^2\varphi(x,t)}{\partial x_i\,\partial x_j}$$

$$+\frac{\varphi(x,t)}{\partial t}\leq\lambda(T)\varphi(x,t). \qquad (A.7)$$

The left-hand side of (A.7) is the instantaneous drift of $\varphi(X_t,t)$ when $X_t=x$. The right-hand side of (A.7) bounds this drift to grow at most linearly with $\varphi(X_t,t)$, which can be used to ensure that $\varphi(X_t,t)$ does not explode in finite time (see the proof of S&V Theorem 10.2.1). (A.6) guarantees that if $\varphi(X_t,t)$ does not explode, neither will X_t. Since the inequality (A.7) holds uniformly for $x\in R^n$, S&V Theorem 10.2.1 easily extends to a random initial condition.

Theorem A.1. Let the nonexplosion condition be satisfied, and let one or more of conditions A, B, C, and D be satisfied. Then a solution to the stochastic integral equation

$$X_t=X_0+\int_0^t b(X_s,s)\,\mathrm{d}s+\int_0^t\sigma(X_s,s)\,\mathrm{d}W_{n,s}, \qquad (A.8)$$

where $\{W_{n,t}\}$ is an n-dimensional standard Brownian motion independent of X_0, and where for any $\Gamma \in B(R^n)$, $P(X_0 \in \Gamma) = \nu_0(\Gamma)$, has a unique weak-sense solution [Liptser and Shiryayev (1977), Ethier and Kurtz (1986)], in that all solutions to (A.8) have the same probability law. This distribution does not depend on the choice of $\sigma(\cdot,\cdot)$. Finally, X_t remains finite in finite time intervals almost surely.

Appendix B: Proofs

Proof of Theorem 2.1

For the case when condition (2.2) is replaced by

$$P_h\Big[{}_hX_t = \big\{h^{-1}((k+1)h - t)\,{}_hX_{kh} + h^{-1}(t - kh)\,{}_hX_{(k+1)h}\big\},$$

$$kh \le t < (k+1)h\Big] = 1. \tag{2.2'}$$

S&V Theorem 11.2.3 proved that when 1) $a(\cdot,\cdot)$, $b(\cdot,\cdot)$, and a fixed starting point X_0 uniquely specify the distribution of a diffusion, 2) ${}_hX_0 = X_0$ for all h, and 3) Assumption 1 holds, then ${}_hX_t$ converges weakly to X_t, the diffusion generated by $a(\cdot,\cdot)$, $b(\cdot,\cdot)$, and X_0. That this still holds when (2.2') is replaced by (2.2), and when the assumption of a fixed initial state is weakened to Assumption 3, is proved in E&K, chapter 7, Corollary 4.2. [*Note:* In E&K and S&V, time appeared in the transition functions as an element of x, not as a separate argument. It is a simple exercise to show that the change of notation is made without loss of generality. See S&V, p. 166, Problem 6.7.2, and p. 266.]

Under Assumption 2, X_t has the stochastic integral representation (2.12) by E&K, chapter 5, Theorem 3.3 and Corollary 3.4. Finally, since the distribution of X_t is characterized only by ν_0, $a(\cdot,\cdot)$, and $b(\cdot,\cdot)$, $\sigma(\cdot,\cdot)$ only enters through the $a(\cdot,\cdot)$ function. Therefore the distribution of X_t does not depend on our choice of $\sigma(\cdot,\cdot)$ as long as $\sigma(\cdot,\cdot)\sigma(\cdot,\cdot)' = a(\cdot,\cdot)$.

Proof of Theorem 2.2

To prove Theorem 2.2 it is sufficient to show that Assumption 5 implies Assumption 1. The Theorem then follows immediately by Theorem 2.1. To show that Assumption 5 implies Assumption 1, it is sufficient to show that

(2.17)–(2.19) imply that for every i, $i=1,\dots,n$, every R, $0<R<\infty$, every T, $0<T<\infty$, and every $\varepsilon>0$,

$$\lim_{h\downarrow 0}\ \sup_{\|x\|\le R,\,0\le t\le T} h^{-1}\int_{\|y-x\|\ge 1}(y-x)_i^2\Pi_{h,h[t/h]}(x,\mathrm{d}y)=0, \tag{B.1}$$

$$\lim_{h\downarrow 0}\ \sup_{\|x\|\le R,\,0\le t\le T} h^{-1}\int_{\|y-x\|\ge 1}|y-x|_i\Pi_{h,h[t/h]}(x,\mathrm{d}y)=0, \tag{B.2}$$

$$\lim_{h\downarrow 0}\ \sup_{\|x\|\le R,\,0\le t\le T} \Delta_{h,\varepsilon}(x,t)=0. \tag{B.3}$$

By Markov's Inequality,

$$\Delta_{h,\varepsilon}(x,t)\le h^{-1}\varepsilon^{-(2+\delta)}\int_{R^n}\|y-x\|^{2+\delta}\Pi_{h,h[t/h]}(x,\mathrm{d}y). \tag{B.4}$$

By Minkowski's Integral Inequality,

$$\le\varepsilon^{-(2+\delta)}\left[\sum_{i=1}^{n}\left[c_{h,i,\delta}(x,t)\right]^{2/(2+\delta)}\right]^{(2+\delta)/2}. \tag{B.5}$$

By (2.17), there is some $\delta>0$ such that for all $R,T>0$, this expression vanishes for every $\varepsilon>0$ as $h\downarrow 0$ uniformly on $\|x\|\le R$, $0\le t\le T$, proving (B.3).

Next, we prove (B.2). By Hölder's Integral Inequality,

$$h^{-1}\int_{\|y-x\|\ge 1}|y-x|_i\Pi_{h,h[t/h]}(x,\mathrm{d}y)$$

$$\le\left[c_{h,i,\delta}(x,t)\right]^{1/(2+\delta)}\left[\Delta_{h,1}(x,t)\right]^{(1+\delta)/(2+\delta)}, \tag{B.6}$$

which vanishes in the same manner (i.e., uniformly on compacts) as (B.5). Finally, we must prove (B.1). Again by Hölder's Integral Inequality,

$$h^{-1}\int_{\|y-x\|\ge 1}(y-x)_i^2\Pi_{h,h[t/h]}(x,\mathrm{d}y)$$

$$\le\left[c_{h,i,\delta}(x,t)\right]^{2/(2+\delta)}\cdot\left[\Delta_{h,1}(x,t)\right]^{\delta/(2+\delta)}, \tag{B.7}$$

which vanishes in the same manner as (B.5) and (B.6).

Proof of Theorem 2.3

Let $A(x)$ and $B(x)$ be a real-valued scalar functions on R^1, with $B(x) \geq 0$ for all $x \in R^1$. Let x_t be the stochastic process given by the stochastic differential equation

$$\mathrm{d}x_t = A(x_t)\,\mathrm{d}t + [2 \cdot B(x_t)]^{1/2}\,\mathrm{d}W_t, \tag{B.8}$$

where W_t is a standard Brownian motion. The corresponding Fokker–Planck equation [Arnold (1974)] is

$$\partial^2[B(x)p]/\partial x^2 - \partial[A(x)p]/\partial x = \partial p/\partial t, \tag{B.9}$$

where $p = p(x|x_0, t)$, $0 \leq t < \infty$, is the probability density of x_t given x_0. Wong (1964) showed that if there is a stationary density for x_t, i.e., a density $p^*(x)$ such that

$$p^*(x) = \lim_{t \to \infty} p(x|x_0, t), \tag{B.10}$$

then it satisfies the differential equation

$$\mathrm{d}[p^*(x) \cdot B(x)]/\mathrm{d}x = A(x) \cdot p^*(x). \tag{B.11}$$

Further, if a density $p^*(x)$ satisfying (B.11) is a member of the Pearson family of distributions [Johnson and Kotz (1970)], then it is the stationary density of the x_t process. For the GARCH(1, 1) diffusion limit (2.40), we have

$$A(\sigma^2) = \omega - \theta\sigma^2, \tag{B.12}$$

$$B(\sigma^2) = \sigma^4\alpha^2/2. \tag{B.13}$$

The results of Wong (1964, sect. 2.F) suggest that we try a density $p^*(\sigma^2)$ of the form $p^*(\sigma^2) \propto \sigma^{2a} \cdot \exp(b/\sigma^2)$ for some a and b. This guess turns out to be correct, and the result is an inverted Gamma density. A change of variables $\lambda \equiv \sigma^{-2}$ then yields the stationary density $f(\lambda)$:

$$f(\lambda) = \Gamma(1 + 2\theta/\alpha^2, 2\omega/\alpha^2), \tag{B.14}$$

where $\Gamma(\cdot,\cdot)$ is the Gamma density. (2.42)–(2.43) follow immediately. Because ${}_h\sigma^2_{kh} \xrightarrow{\mathrm{d}} \sigma^2_t$ for $kh = t$ as $h \downarrow 0$, (2.44) immediately follows. Since $h^{-1/2}{}_hZ_{kh} \sim \mathrm{N}(0,1)$ and is independent of λ, we integrate λ out to obtain (2.45).

By Kushner (1984, ch. 6, theorem 6), (2.46) and (2.47) will follow if we can show that the sequence of random functions $\{{}_h\sigma_t^2\}_{h\downarrow 0}$ is tight in $D[0,\infty)$. By Kushner (1984, ch. 6, theorem 3), (2.46)–(2.47) will hold if we can find a twice-continuously differentiable Liapunov function $\varphi(\cdot)$ defined on $[0,\infty)$ and strictly positive numbers h^*, Δ, and η such that

$$\min_{0\le\sigma^2}\varphi(\sigma^2) = 0, \tag{B.15}$$

$$\lim_{\sigma^2\to\infty}\varphi(\sigma^2) = \infty, \tag{B.16}$$

$$\sup_{0<h\le h^*}\mathrm{E}\big[\varphi\big({}_h\sigma_0^2\big)\big] < \infty, \tag{B.17}$$

and for all ${}_h\sigma_{kh}^2 \ge 0$ and all h, $0<h<h^*$,

$$h^{-1}\mathrm{E}\big[\varphi\big({}_h\sigma_{(k+1)h}^2\big) - \varphi\big({}_h\sigma_{kh}^2\big)|M_{kh}\big] < \Delta - \eta\cdot\varphi\big({}_h\sigma_{kh}^2\big). \tag{B.18}$$

Define, for any d, $0<d<1$,

$$\varphi_d(\sigma^2) \equiv \begin{cases} 0 & \text{if} \quad \sigma^2 = 0, \\ \sigma^{2d}\cdot\exp(-1/\sigma^2) & \text{otherwise}. \end{cases} \tag{B.19}$$

For each $d>0$, $\varphi_d(\sigma^2)$ and its derivatives of arbitrary order are continuous on $[0,\infty)$. $\varphi_d(\sigma^2)$ also satisfies (B.15)–(B.16). (B.18) was already assumed in the statement of Theorem 2.3. All that remains therefore is to verify that (B.18) holds for some $d>0$.

To simplify notation, we write $\varphi(\sigma^2)$, σ_+^2, and σ^2 instead of $\varphi_d(\sigma^2)$, ${}_h\sigma_{(k+1)h}^2$, and ${}_h\sigma_{kh}^2$. By S&V Lemma 11.2.1,

$$h^{-1}\mathrm{E}\big[\varphi\big(\sigma_+^2\big) - \varphi(\sigma^2)|M_{kh}\big] - \big(\alpha^2\sigma^4/2\big)\cdot\varphi''(\sigma^2)$$
$$-(\omega-\theta\sigma^2)\cdot\varphi'(\sigma^2)\to 0, \tag{B.20}$$

uniformly on compacts for fixed d. Therefore, if we confine σ^2 to any compact set $\Theta\subset[0,\infty)$, (B.20) reduces to

$$(\alpha^2\sigma^4/2)\cdot\varphi'' + (\omega-\theta\sigma^2)\cdot\varphi' + o(1) < \Delta - \eta\cdot\varphi. \tag{B.18$'$}$$

Since $\varphi(\cdot)$ and its derivatives are locally bounded, it is clear that for any compact Θ, a finite Δ can be found such that (B.20) holds uniformly on

$\sigma^2 \in \Theta$. Now consider (B.18) for large values of σ^2: expanding $h^{-1}[\varphi(\sigma_+^2) - \varphi(\sigma^2)]$ in a four-term Taylor series and taking expectations, (B.18) is satisfied if the inequality

$$\begin{aligned} & h^{-1}\mathrm{E}\left[\varphi(\sigma_+^2) - \varphi(\sigma^2)|M_{kh}\right] \\ & = -d\theta\sigma^{2d} + \mathrm{O}(\sigma^{-2}) + d(d-1)(\alpha^2/2)\sigma^{2d} \\ & \quad + \sigma^{2d}\left[\mathrm{O}(h^{1/2}) + \mathrm{O}(\sigma^{-2}) \cdot \mathrm{O}(h)\right] \\ & \quad + h^{-1}d(d-1)(d-2)(d-3) \\ & \quad \cdot \mathrm{E}\left[\left(\varphi(\sigma_+^2) - \varphi(\sigma^2)\right)^4 \cdot \sup_{0 \le \delta \le 1} \left(\delta\sigma^2 + (1-\delta)\sigma_+^2\right)^d | M_{kh}\right] \\ & < \Delta - \eta \cdot \sigma^{2d}, \end{aligned} \tag{B.21}$$

is satisfied. The last term on the left-hand side of (B.21) is nonpositive as long as $0 < d < 1$. Clearly also, a finite Δ can be chosen that dominates the $\mathrm{O}(\cdot)$ and $\mathrm{o}(\cdot)$ terms. (B.18) will therefore hold if we can find a d, $0 < d < 1$, and a $\eta > 0$ such that

$$(\eta - \theta) \cdot \sigma^{2d} + (d-1)(\alpha^2/2) \cdot \sigma^{2d} < 0, \tag{B.22}$$

which we can always do if $2\theta/\alpha^2 > -1$.

Proof of Theorem 3.1

We need only verify Assumptions 2 and 5. (3.5)–(3.7) factor $a(s,y,t)$ into $\sigma(s,y,t)\sigma(s,y,t)'$, satisfying Assumption 2. To verify Assumption 5, we need to check that $b_h^*(\cdot,\cdot,\cdot)$ and $a_h^*(\cdot,\cdot,\cdot)$ converge, and that $c_{h,i,\delta}(\cdot,\cdot,\cdot) \to 0$ uniformly on compacts,

$$b_h^*(s,y,t) = \begin{bmatrix} f(s,y,t) \\ F(s,y,t) \end{bmatrix}. \tag{B.23}$$

Since $b_h^*(\cdot,\cdot,\cdot) = b(\cdot,\cdot,\cdot)$, (2.19) is satisfied. Similarly, we have (dropping the function arguments for the sake of notational clarity):

$$a_h^*(s,y,t) = \begin{bmatrix} hf^2 + g^2 & hfF' + G\Omega_{1,2}g \\ hFf + g\Omega_{2,1}G' & G\Omega_{2,2}G' \end{bmatrix}. \tag{B.24}$$

Since f, F, g, and G are locally bounded, it is clear that

$$\lim_{h\downarrow 0} \sup_{\|x\|\le R,\, 0\le t\le T} \|a_h^*(s,y,t)-a(s,y,t)\|=0. \tag{B.25}$$

Finally, choose $\delta=1$. Stacking the elements of $c_{h,i,1}(\cdot,\cdot,\cdot)$ in a vector, we have

$$c_{h,1}({}_hS_{kh},{}_hY_{kh},t)=h^{-1}\mathrm{E}\begin{bmatrix}|hf+g_hZ_{kh}|^3\\ |hF+G_hZ_{kh}^*|^3\end{bmatrix}=\mathrm{O}(h^{1/2}), \tag{B.26}$$

uniformly on compacts, so that (2.17) is satisfied.

Proof of Theorem 3.2

Let $\{Z_j\}_{j=1,\infty}$ be an i.i.d. N(0, 1) sequence, and note that

$$Q_{k,t}\sim(1-2/\pi)^{-1/2}[kt]^{-1/2}\sum_{j=1}^{[kt]}\left[|Z_j|-\mathrm{E}|Z_j|\right], \tag{B.27}$$

where $\sim$ denotes equality in distribution. Since $\mathrm{E}|Z_j|=(2/\pi)^{1/2}$, the Theorem follows immediately by Donsker's Theorem [Billingsley (1968)].

Proof of Theorem 3.3

Define $V_t\equiv\ln(\sigma_t^2)$ and $S^*\equiv\ln(S_t)$ and substitute into (3.19). Our equation system becomes

$$\mathrm{d}S^*=\theta e^V\mathrm{d}t+e^{V/2}\mathrm{d}W_1, \tag{B.28}$$

$$\mathrm{d}V=-\beta(V-\alpha)\,\mathrm{d}t+\mathrm{d}W_2, \tag{B.29}$$

with

$$b=\left[\theta\mathrm{e}^V,-\beta(V-\alpha)\right]', \tag{B.30}$$

$$a=\begin{bmatrix}\mathrm{e}^V & \mathrm{e}^{V/2}C_{12}\\ \mathrm{e}^{V/2}C_{12} & C_{22}\end{bmatrix}. \tag{B.31}$$

It is readily verified that condition B of appendix A holds. To verify the nonexplosion condition, define for $K>0$,

$$\varphi(V,S^*)\equiv K+f(S^*)\cdot|S^*|+f(V)\cdot\exp(|V|), \tag{B.32}$$

where $f(x) \equiv \exp(-1/|x|)$ if $x \neq 0$ and $\equiv 0$ otherwise. $\varphi(\cdot,\cdot)$ is nonnegative, arbitrarily differentiable, and satisfies (A.6). Its derivatives are locally bounded, so that positive K and λ can be chosen to satisfy (A.7) on any compact set. For large values of S^* and V, $\varphi_V \approx \operatorname{sign}(V)\cdot\exp(|V|)$, $\varphi_{VV} \approx \exp(|V|)$, $\varphi_{S^*} \approx \operatorname{sign}(S^*)$, $\varphi_{S^*S^*} \approx 0$, and $\varphi_{VS^*} = 0$, so that with $\lambda > 1 + \beta\alpha + C_{22}/2 + |\theta|$, a finite K can be found to satisfy (A.7). The result then follows by Theorem 2.2.

Proof of Theorem 3.4

Since $h^{-1/2}{}_hZ_{kh} \sim \mathrm{N}(0,1)$ and since ${}_h\sigma_{kh}$ and ${}_hZ_{kh}$ are independent, (3.26)–(3.27) follow immediately from Theorem 3.3 and from Arnold (1974, sect. 8.3). The proof of (3.28)–(3.29) proceeds along the same lines as the proof of (2.46)–(2.47) in Theorem 2.3: Label $V_+ \equiv \ln({}_h\sigma^2_{(k+1)h})$ and $V \equiv \ln({}_h\sigma^2_{kh})$, and define the Liapunov function $\varphi(V) \equiv [V-\alpha]^2$. We then have

$$h^{-1}\mathrm{E}[\varphi(V_+) - \varphi(V)|M_{kh}] = -2\cdot\beta\cdot\varphi(V) \pm h\beta^2\cdot\varphi^{1/2}(V) + C_{22}. \tag{B.33}$$

(B.18) then requires that there exist positive numbers Δ, η, and h^* such that

$$\sup_{0<h<h^*,\, V\in R^1} (\eta - 2\cdot\beta)\cdot\varphi(V) \pm h\beta^2\cdot\varphi^{1/2}(V) + C_{22} - \Delta < 0, \tag{B.34}$$

which is clearly possible as long as $\beta > 0$.

Proof of Theorem 3.5

We proved nonexplosion and condition B in section 2. The result then follows by a direct application of Theorem 3.1.

Proof of Theorem A.1

For a given ν_0, the conclusion of the Theorem is equivalent to the 'martingale problem' of S&V being well-posed (S&V, chapter 6, and E&K, chapters 4 and 5), or, equivalently, to a unique weak sense solution to the stochastic integral equation (2.12) [Ethier and Kurtz (1986, ch. 5, corollary 3.4)]. S&V assume that ν_0 puts all its mass at a given point, and E&K make the extension to a random initial condition. E&K, chapter 4, Problem 49 and the nonexplosion condition allow us to conclude that, if the martingale problem as defined in S&V is well-posed, then the martingale problem with random initial condition is well-posed for any ν_0 on R^n.

The Theorem using condition A follows by S&V, Corollary 6.3.3 and Theorem 10.2.1. Using condition B, it follows by S&V, Theorems 7.2.1 and 10.2.1. Using condition C, it follows by S&V, Theorems 6.3.4 and 10.2.1, and with condition D, it follows by S&V, Theorems 8.2.1 and 10.2.1.

References

Abel, A.B., 1988, Stock prices under time-varying dividend risk: An exact solution in an infinite-horizon general equilibrium model, Journal of Monetary Economics 22, 375–393.

Arnold, L., 1974, Stochastic differential equations: Theory and applications (Wiley, New York, NY).

Barsky, R.B., 1989, Why don't the prices of stocks and bonds move together?, American Economic Review 79, 1132–1146.

Bernanke, B.S., 1983, Irreversibility, uncertainty and cyclical investment, Quarterly Journal of Economics 98, 85–106.

Billingsley, P., 1968, Convergence of probability measures (Wiley, New York, NY).

Black, F., 1976, Studies of stock market volatility changes, Proceedings of the American Statistical Association, Business and Economic Statistics Section, 177–181.

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

Clark, P.K., 1973, A subordinated stochastic process model with finite variance for speculative prices, Econometrica 41, 135–155.

Cox, J., J. Ingersoll, and S. Ross, 1985, An intertemporal general equilibrium model of asset prices, Econometrica 53, 363–384.

Cox, J., S. Ross, and M. Rubinstein, 1979, Options pricing: A simplified approach, Journal of Financial Economics 7, 229–263.

Duffie, D. and K.J. Singleton, 1989, Simulated moments estimation of Markov models of asset prices, Mimeo. (Graduate School of Business, Stanford University, Stanford, CA).

Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, Econometrica 50, 987–1008.

Engle, R.F. and T. Bollerslev, 1986a, Modelling the persistence of conditional variances, Econometric Reviews 5, 1–50.

Ethier, S.N. and T.G. Kurtz, 1986, Markov processes: Characterization and convergence (Wiley, New York, NY).

French, K.R., G.W. Schwert, and R.F. Stambaugh, 1987, Expected stock returns and volatility, Journal of Financial Economics 19, 3–30.

Gennotte, G. and T.A. Marsh, Variations in economic uncertainty and risk premiums on capital assets, Mimeo. (University of California, Berkeley, CA).

Geweke, J., 1986, Modelling the persistence of conditional variances: A comment, Econometric Reviews 5, 57–61.

Hosking, J.R.M., 1981, Fractional differencing, Biometrika 68, 165–176.

Johnson, N.L. and S. Kotz, 1970, Distributions in statistics: Continuous univariate distributions – I (Wiley, New York, NY).

Kushner, H.J., 1984, Approximation and weak convergence methods for random processes, with applications to stochastic systems theory (M.I.T. Press, Cambridge, MA).

Liptser, R.S. and A.N. Shiryayev, 1977, Statistics of random processes, Vol. I (Springer Verlag, New York, NY).

Lo, A., 1988, Maximum likelihood estimation of generalized Ito processes with discretely sampled data, Econometric Theory 4, 231–247.

Merton, R.C., 1973, An intertemporal capital asset pricing model, Econometrica 41, 867–888.

Nelson, D.B., 1988a, The time series behavior of stock market volatility and returns, Unpublished doctoral dissertation (M.I.T., Cambridge, MA).

Nelson, D.B., 1988b, Stationarity and persistence in the GARCH(1, 1) model, Graduate School of Business working paper series in economics and econometrics no. 88-68 (University of Chicago, Chicago, IL), forthcoming in Econometric Theory.
Nelson, D.B., 1989, Conditional heteroskedasticity in asset returns: A new approach, Graduate School of Business working paper series in economics and econometrics no. 89-73 (University of Chicago, Chicago, IL), forthcoming in Econometrica.
Officer, R.R., 1973, The variability of the market factor of the New York Stock Exchange, Journal of Business 46, 434–453.
Pantula, S.G., 1986, Modelling the persistence of conditional variances: A comment, Econometric Reviews 5, 71–73.
Pardoux, E. and D. Talay, 1985, Discretization and simulation of stochastic differential equations, Acta Applicandae Mathematica 3, 23–47.
Priestley, M.B., 1981, Spectral analysis and time series (Academic Press, London).
Sampson, M., 1988, A stationarity condition for the GARCH(1, 1) model, Mimeo. (Department of Economics, Concordia University, Montreal).
Sandburg, C., 1969, The people, yes, The complete poems of Carl Sandburg (Harcourt, Brace, Jovanovich, San Diego, CA).
Stroock, D.W. and S.R.S. Varadhan, 1979, Multidimensional diffusion processes (Springer Verlag, Berlin).
Wong, E., 1964, The construction of a class of stationary Markov processes, in: R. Bellman, ed., Sixteenth symposia in applied mathematics – Stochastic processes in mathematical physics and engineering (American Mathematical Society, Providence, RI) 264–276.