

Filtering and forecasting with misspecified ARCH models I

Getting the right variance with the wrong model

Daniel B. Nelson*

NBER and University of Chicago, Chicago, IL 60637, USA

This paper investigates the properties of the conditional covariance estimates generated by a misspecified ARCH model. For example, suppose that we observe a diffusion process at discrete time intervals of length h . For each h , we use a GARCH(1,1) model to estimate the instantaneous conditional covariance matrix of the diffusion. Under mild regularity conditions, the difference between these conditional covariance estimates and the true conditional covariance converges to zero in probability as $h \downarrow 0$. Many other ARCH models (for example, Exponential ARCH) have similar consistency properties. This may well account for the success of ARCH models in short-term forecasting using high-frequency data, since even misspecified ARCH models can produce ‘good’ estimates of volatility.

1. Introduction

Most theories of asset pricing, for example the CAPM of Sharpe (1964) and Lintner (1965), the option pricing formula of Black and Scholes (1973), and the arbitrage pricing theory of Ross (1976), relate required returns on assets to their variances and covariances. An enormous literature in empirical

*I would like to thank Stuart Ethier, Dean Foster, Leo Kadanoff, G. Andrew Karolyi, Albert Kyle, Franz Palm, G. William Schwert, an editor (Robert Engle), two anonymous referees, participants at the 1989 NBER Summer Institute, the 1990 Conference on Statistical Models for Financial Volatility, the 1990 Paris ARCH conference, and at workshops at Northwestern, Queen’s University, the University of Chicago, and the University of Wisconsin at Madison for helpful comments. The usual disclaimer regarding errors applies. The Center for Research in Security Prices provided research support.

finance has explored the nature of this relation between risk and return. This literature has made it clear that the variability of returns and the degree of co-movement between assets change stochastically over time. Practical experience (as in the October stock market crashes of 1929, 1987, and 1989) points to the same conclusion. This realization has lead many researchers to recast asset pricing theory in terms of the *conditional* variances and covariances of returns. [See, for example, Engle, Ng, and Rothschild (1990), Harvey (1989), Hull and White (1987), Merton (1973), Shanken (1990), and Wiggins (1987).]

Researchers have examined the time series behavior of stock market volatility for other reasons as well: for example, Schwert (1989) investigates the connection between macroeconomic variables and market volatility. The nature of market volatility changes has also entered into public policy debates on financial market regulation, margin requirements, transactions taxes, etc. [See, e.g., Hardouvelis (1988), Hsieh and Miller (1990), Summers and Summers (1989), and the Summer 1988 issue of the *Journal of Economic Perspectives*.]

Since the seminal work of Engle (1982), ARCH models have become one of the most widely used means of modelling changing conditional variances.¹ As are all statistical models, ARCH models are at best a rough approximation to reality, and in applied work a misspecified model will inevitably be chosen. If we are concerned with accurately measuring conditional variances and covariances, it is important to ask how well a *misspecified* ARCH model can approximate the true structure of conditional heteroskedasticity in a time series.

This paper examines the ability of a misspecified ARCH model to estimate the conditional covariance matrix of a stochastic process. To anticipate the results below, ARCH models are remarkably robust to certain types of misspecification. In particular, *if the process is well approximated by a diffusion*, broad classes of ARCH models provide consistent (in a sense developed below) estimates of the conditional covariances. The ARCH models may seriously misspecify both the conditional mean of the process and the dynamic behavior of the conditional variances. In fact, the misspecification can be so severe that the ARCH models would make no sense as data-generating processes and may make terrible medium and long-term forecasts – without affecting the consistency of the one-step-ahead conditional covariance estimates. As we will see, however, this robustness does not extend to all ARCH models, and it breaks down altogether to the extent that the process is not well approximated by a diffusion.

¹The survey of applications of ARCH in finance by Bollerslev, Chou, and Kroner (1992) lists several hundred papers on ARCH in finance, most dated 1989 and after.

1.1. The initial setup

Define the $n \times 1$ diffusion process $\{X_t\}$ by the stochastic integral equation

$$X_t = X_0 + \int_0^t \mu(X_s) ds + \int_0^t \Omega^{1/2}(X_s) dW_s, \quad (1.1)$$

where $\{W_t\}$ is an $n \times 1$ standard Brownian motion, and $\mu(\cdot)$ and $\Omega^{1/2}(\cdot)$ are continuous functions from R^n into R^n and into the space of $n \times n$ matrices, respectively. Under mild moment conditions [see Arnold (1973)] $\Omega(X_t) \equiv \Omega^{1/2}(X_t) \cdot \Omega^{1/2}(X_t)'$ and $\mu(X_t)$ are, respectively, the instantaneous conditional covariance matrix and instantaneous conditional mean (per unit of time) of increments in the $\{X_t\}$ process. X_0 is assumed random with probability measure ν_0 independent of $\{W_t\}_{0 \leq t < \infty}$.

In applications, $\{X_t\}$ might include asset prices and other state variables describing an economy. Some elements of $\{X_t\}$ may never be directly observable. Others, we assume, are observable, but only at discrete time intervals of length h . For example, our model may define the stock price process in continuous time, while we observe only stock prices at discrete intervals of length h . Formally, we partition X_t as $[X'_{1:q,t} \ X'_{q+1:n,t}]'$, where $X_{1:q,t}$ consists of the first q elements of X_t (observable at intervals of length h) and $X_{q+1:n,t}$ consists of the last $n - q$ elements (which are never observable).² We partition μ and Ω accordingly.

For each t , our goal is to estimate $\Omega_{1:q,1:q}(X_t)$ (the instantaneous covariance of the increments in the observable variables $X_{1:q,t}$ given t and the history up to time t of the $\{X_t\}$ process) given the smaller information set $\{t, X_{1:q,0}, X_{1:q,h}, X_{1:q,2h}, \dots, X_{1:q,h[t/h]}\}$ (i.e., the time index and the past observed values of $X_{1:q,t}$). Under these assumptions, $\Omega_{1:q,1:q}(X_t)$ is unobservable since $X_{q+1:n,t}$ is unobservable, and unless t is an integer multiple of h , $X_{1:q,t}$ is as well – i.e., $\Omega_{1:q,1:q}(X_t)$ is a conditional covariance matrix, but is conditional on a larger information set than is possessed by the econometrician.

1.2. Consistent filters: A definition

In our examples, we generate conditional covariance estimates $\{\hat{\Omega}_{1:q,1:q,t}\}$ with a sequence of ARCH models, whose coefficients may depend on h . We say that a given sequence of ARCH conditional covariance estimates $\{\hat{\Omega}_{1:q,1:q,t}\}_{h \downarrow 0}$ provides a *consistent filter* for the $\{\Omega_{1:q,1:q,t}\}$ if for each

²Similarly, in our notation, $A_{i,j,t}$ is the $(i - j)$ th element of the matrix function A_t at time t , and $b_{i,t}$ is the i th element of the vector function b_t at time t . Where A_t can be written as $A(X_t)$, we use the notations interchangeably.

$t > 0$,

$${}_h\hat{\Omega}_{1:q,1:q,t} - \Omega_{1:q,1:q,t} \rightarrow 0_{q \times q} \quad (1.2)$$

(a $q \times q$ matrix of zeros) in probability for all $t > 0$ as $h \downarrow 0$. Thus, the ARCH model consistently estimates (pointwise in t) the true underlying conditional covariance as h , the time between observations, goes to zero. Note that our use of the term ‘estimate’ corresponds to its use in the filtering literature rather than the statistics literature; that is, an ARCH model with (given) fixed parameters produces ‘estimates’ of the true underlying conditional covariance matrix at each point in time in the same sense that a Kalman filter produces ‘estimates’ of unobserved state variables in a linear system.³

1.3. Some intuition on why misspecified ARCH models can provide consistent filters

In (1.1), we assumed that $\{X_t\}$ follows a diffusion process and that $\Omega(x)$ is continuous. The sample path of $\{\Omega_t\}$ is therefore continuous with probability one. Suppose we are trying to estimate the value of the $\Omega_{1,1,T}$, the time T instantaneous variance of increments in $\{X_{1,t}\}$, the first element of X_t , using information in the sample path of X_1 up to time T . Since the path of $\{\Omega_t\}$ is continuous almost surely, for every $\varepsilon > 0$ and every $T > 0$ there exists, with probability one, a random $\Delta(T) > 0$ such that

$$\sup_{T-\Delta(T) \leq s \leq T} |\Omega_{1,1,s} - \Omega_{1,1,T}| < \varepsilon. \quad (1.3)$$

That is, over a suitably small time interval, the change in $\Omega_{1,1}$ can be made arbitrarily small, as can the change in μ_1 , the drift in X_1 . Now take a small interval $[T - \Delta, T]$, chop it into M equal pieces, and estimate $\Omega_{1,1,T}$ by

$$\hat{\Omega}_{1,1,T}(\Delta, M) \equiv \Delta^{-1} \sum_{j=1}^M [X_{1,T-(j-1)\Delta/M} - X_{1,T-j\Delta/M}]^2. \quad (1.4)$$

For small Δ , μ_1 and $\Omega_{1,1}$ are effectively constant. Given $\mu_{1,T-\Delta}$ and $\Omega_{1,1,T-\Delta}$ the increments $X_{1,T-(j-1)\Delta/M} - X_{1,T-j\Delta/M}$ are approximately i.i.d. $N(M^{-1}\Delta\mu_{1,T-\Delta}, M^{-1}\Delta\Omega_{1,1,T-\Delta})$. Under suitable moment conditions, a law of large numbers yields $[\hat{\Omega}_{1,1,T}(\Delta, M) - \Omega_{1,1,T}] \rightarrow 0$ in probability as $\Delta \downarrow 0$ and $M \rightarrow \infty$. Note that failing to correct for the nonzero drift in X_1 does not interfere with consistency, since its effect on $\hat{\Omega}_{1,1,T}(\Delta, M)$ vanishes as $M \rightarrow \infty$ and $\Delta \downarrow 0$.

³See, e.g., the use of the term in Anderson and Moore (1979, ch. 2) or Arnold (1973, ch. 12).

Like $\hat{\Omega}_{1,1,T}(\Delta, M)$, the conditional covariance estimates produced by GARCH(1, 1) are long distributed lags of squared residuals. We can create a sequence of GARCH(1, 1) models that, like $\hat{\Omega}_{1,1,T}(\Delta, M)$, form estimates of $\Omega_{1,1,T}$ by averaging increasing numbers of squared residuals from the increasingly recent past. In this way, a sequence of GARCH(1, 1) models can consistently estimate $\Omega_{1,1,T}$, despite misspecifying the drift in X_1 , ignoring the other state variables $X_{2:n,T}$, and misspecifying the dynamics of $\Omega_{1,1,T}$, whose random behavior may be very different from that of a GARCH(1, 1).

To see that this consistency is not limited to GARCH(1, 1), consider the following estimation procedure inspired by the methods in Taylor (1986) and Schwert (1989): Instead of estimating $\Omega_{1,1,T}$ by a sum of lagged squared residuals, estimate $\Omega_{1,1,T}^{1/2}$ by a sum of lagged absolute residuals:

$$\hat{\Omega}_{1,1,T}^{1/2}(\Delta, M) \equiv (2\Delta M/\pi)^{-1/2} \sum_{j=1}^M |X_{1,T-(j-1)\Delta/M} - X_{1,T-j\Delta/M}|, \quad (1.5)$$

which again is consistent under sufficient moment conditions as $\Delta \downarrow 0$ and $M \rightarrow \infty$.

Thus far, we have presented heuristics: we have not specified regularity conditions on the diffusion (1.1) and on the sequence of ARCH models to guarantee consistency. In addition, there are other ARCH models, such as the Exponential ARCH (EGARCH) model of Nelson (1991), in which the conditional variance estimate is not a simple moving average of transformed residuals. To make our heuristics rigorous and broadly applicable requires more work. In section 2, we present basic results, developed by Ethier and Nagylaki (1988), on diffusion approximations. In section 3, we use these results to prove consistency theorems for three ARCH models: the multivariate GARCH(1, 1) process of Bollerslev, Engle, and Wooldridge (1988), the AR(1) Exponential ARCH model of Nelson (1991), and a variant of GARCH(1, 1) inspired by Taylor (1986) and Schwert (1989). Section 4 discusses some circumstances in which consistency breaks down, while section 5 concludes the paper. Proofs are gathered in the appendix.

2. Weak convergence of Markov processes to diffusions

In this section we present results, drawn largely from Stroock and Varadhan (1979) and Ethier and Nagylaki (1988), on the weak convergence of a sequence of Markov processes to a diffusion. We begin by defining our sequence of processes, which we divide into two pieces: $\{X_t\}$ and $\{Y_t\}$, $n \times 1$

and $m \times 1$ vector processes, respectively.⁴ $\{ {}_hX_t \}$ and $\{ {}_hY_t \}$ will be random step functions taking jumps at times $h, 2h, 3h$, and so on. In the convergence results below, we present sufficient conditions for $\{ {}_hX_t \} \Rightarrow \{ X_t \}$, where $\{ X_t \}$ is a solution of (1.1) and \Rightarrow denotes weak convergence. In the applications in section 3, $\{ {}_hX_t \}$ represents the underlying stochastic system generating the data, while $\{ {}_hY_t \}$ represents the difference between the true conditional covariance matrix of $\{ {}_hX_t \}$ and the estimate generated by an ARCH model. The conditions below will be used in section 3 to guarantee that for every $t > 0$, ${}_hY_t$ converges in probability to an $m \times 1$ vector of zeros as $h \downarrow 0$ – i.e., that for each $t > 0$, our conditional covariance estimator is consistent as $h \downarrow 0$.

Formally, let $D([0, \infty), R^n \times R^m)$ be the space of functions from $[0, \infty)$ into $R^n \times R^m$ that are right continuous with finite left limits (RCLL). D is a metric space when endowed with the Skorohod metric [see Ethier and Kurtz (1986, ch. 3) for formal definitions]. For each $h > 0$, let ${}_h\mathcal{L}_{kh}$ be the information set at time kh – i.e., the σ -algebra generated by ${}_hX_0, {}_hX_h, {}_hX_{2h}, \dots, {}_hX_{kh}$ and ${}_hY_0, {}_hY_h, {}_hY_{2h}, \dots, {}_hY_{kh}$. Let $B(R^n \times R^m)$ denote the Borel sets on $R^n \times R^m$, and let ν_h be a probability measure on $(R^n, B(R^n \times R^m))$. For each $h > 0$, let $\Pi_h(x, y, \cdot)$ be a transition function on $R^n \times R^m$ – i.e.:

- (a) $\Pi_h(x, y, \cdot)$ is a probability measure on $(R^n \times R^m, B(R^n \times R^m))$ for all $(x, y) \in R^n \times R^m$.
- (b) $\Pi_h(\cdot, \cdot, \Gamma)$ is $B(R^n \times R^m)$ measurable for all $\Gamma \in B(R^n \times R^m)$.

Let P_h be the probability measure on $D([0, \infty), R^n \times R^m)$ such that

$$P_h[({}_hX_0, {}_hY_0) \in \Gamma] = \nu_h(\Gamma) \quad \text{for any } \Gamma \in B(R^n \times R^m), \quad (2.1)$$

$$P_h[({}_hX_t, {}_hY_t) = ({}_hX_{kh}, {}_hY_{kh}), \quad kh \leq t < (k+1)h] = 1, \quad (2.2)$$

$$P_h[({}_hX_{(k+1)h}, {}_hY_{(k+1)h}) \in \Gamma | {}_h\mathcal{L}_{kh}] = \Pi_h({}_hX_{kh}, {}_hY_{kh}, \Gamma) \quad (2.3)$$

almost surely under P_h for all $k \geq 0$ and all $\Gamma \in B(R^n \times R^m)$.

⁴In our notation, $\{ {}_hX_t \}$ refers to ${}_hX_t$ as a random function of time – i.e., it refers to the ${}_hX_t$ process. On the other hand, ${}_hX_t$ (without the curly brackets) refers to the value that the $\{ {}_hX_t \}$ process takes at a particular time t .

It is possible to make the transition probability depend on time by making the time index kh an element of ${}_hX_{kh}$.

It is convenient to impose the following condition ruling out feedback from $\{{}_hY_{kh}\}$ to $\{{}_hX_{kh}\}$: for every Borel subset Γ_x of R^n and for all $h > 0$,

$$\Pi_h(x, y, \Gamma_x \times R^m) \text{ is independent of } y. \quad (2.4)$$

For each $h > 0$, the $(n + m)$ -dimensional discrete time Markov process $\{{}_hX_{kh}, {}_hY_{kh}\}$ has transition probabilities given by (2.3) and a random starting point $({}_hX_0, {}_hY_0)$ given by (2.1). We form the continuous time process $\{{}_hX_t, {}_hY_t\}$ from $\{{}_hX_{kh}, {}_hY_{kh}\}$ by (2.2), making $\{{}_hX_t, {}_hY_t\}$ a step function with jumps at times $h, 2h, 3h$, and so on.

Our notation must keep track of three distinct kinds of processes:

- (a) the sequence of discrete time processes $\{{}_hX_{kh}\}$ and $\{{}_hY_{kh}\}$ that depend both on h and on the (discrete) time index $kh, k = 0, 1, 2, \dots$,
- (b) the sequence of continuous time processes $\{{}_hX_t\}$ and $\{{}_hY_t\}$ formed as step functions from the discrete time process in (a) using (2.2), and
- (c) a limit diffusion process $\{X_t\}$, with $\{{}_hX_t\} \Rightarrow \{X_t\}$ as $h \downarrow 0$.

To understand the weak convergence results below, it is useful to think of $\{X_t\}$, a solution to (1.1), as a random variable taking a value in $D([0, \infty), R^n)$. Under regularity conditions explored in Stroock and Varadhan (1979) and Ethier and Kurtz (1986), four characteristics uniquely define the distribution of $\{X_t\}$:

- (a) starting point X_0 (or its distribution if X_0 is random),
- (b) drift function $\mu(x)$,
- (c) covariance function $\Omega(x)$, and
- (d) almost sure continuity of $\{X_t\}$ as a function of t .

We achieve $\{{}_hX_t\} \Rightarrow \{X_t\}$ by making $\{{}_hX_t\}$ match (a)–(d) in the limit as $h \downarrow 0$. To match (a), we require Assumption 1:

Assumption 1. $({}_hX_0, {}_hY_0) \Rightarrow (X_0, Y_0)$ as $h \downarrow 0$, where (X_0, Y_0) has probability measure ν_0 .

Next, we require that the discrete analogues of μ and Ω , μ_h and Ω_h , are well-defined for all $(x, y) \in R^{n+m}$:

$$\mu_h(x) \equiv h^{-1} E\left[{}_hX_{(k+1)h} - {}_hX_{kh} \mid {}_hX_{kh} = x, {}_hY_{kh} = y\right], \quad (2.5)$$

$$\begin{aligned} \Omega_h(x) \equiv h^{-1} E\left[({}_hX_{(k+1)h} - {}_hX_{kh}) \right. \\ \left. \times ({}_hX_{(k+1)h} - {}_hX_{kh}) \mid {}_hX_{kh} = x, {}_hY_{kh} = y\right], \end{aligned} \quad (2.6)$$

where the expectations in (2.5)–(2.6) are taken under P_h .⁵ μ_h and Ω_h are independent of y by (2.4). Next, define the norm of a $q \times r$ real matrix A as

$$\|A\| \equiv \left[\sum_{i=1, q} \sum_{j=1, r} A_{i,j}^2 \right]^{1/2}. \quad (2.7)$$

To match (b)–(c) in the limit as $h \downarrow 0$, we require:

Assumption 2. For every $\eta > 0$,

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq \eta} \|\mu_h(x) - \mu(x)\| = 0, \quad (2.8)$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq \eta} \|\Omega_h(x) - \Omega(x)\| = 0, \quad (2.9)$$

where $\mu(\cdot)$ and $\Omega(\cdot)$ are continuous.

Assumption 2 requires μ_h and Ω_h to converge uniformly on every bounded subset of R^n , which is much less restrictive than uniform convergence on R^n .

The following fourth moment condition guarantees sample path continuity of the limit process [see Arnold (1973, p. 40) and Nelson (1990a, theorem 2.2)], and so matches (d) in the limit as $h \downarrow 0$:

Assumption 3. For every $\eta > 0$ and all $i = 1, \dots, n$,

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq \eta} h^{-1} E\left[({}_hX_{i, (k+1)h} - {}_hX_{i, kh})^4 \mid {}_hX_{kh} = x\right] = 0. \quad (2.10)$$

⁵(2.6) defines $\Omega_h(x)$ as an uncentered conditional second moment. None of our results would be altered by using the centered moments instead [i.e., by replacing $({}_hX_{h(k+1)} - {}_hX_{hk})$ with $({}_hX_{h(k+1)} - {}_hX_{hk} - h\mu_h({}_hX_{hk}))$ in (2.6)], since the difference between the centered and uncentered second moments vanishes as $h \downarrow 0$.

Finally, (a)–(d) must completely characterize the distribution of $\{X_t\}$:

Assumption 4. There is a distributionally unique solution to (1.1).⁶

Theorem 2.1 (Stroock-Varadhan). Under Assumptions 1–4, $\{\mathbf{X}_t\} \Rightarrow \{X_t\}$ as $h \downarrow 0$.

We next turn to $\{Y_t\}$. First, for every $\delta > 0$, $h > 0$, and $(x, y) \in R^{n+m}$, let the following conditional expectations be well-defined:

$$c_{h,\delta}(x, y) \equiv h^{-\delta} E[hY_{(k+1)h} - {}_hY_{kh} | {}_hX_{kh} = x, {}_hY_{kh} = y], \quad (2.11)$$

$$\begin{aligned} d_{h,\delta}(x, y) \equiv & h^{-\delta} E[({}_hY_{(k+1)h} - {}_hY_{kh}) \\ & \times ({}_hY_{(k+1)h} - {}_hY_{kh})' | {}_hX_{kh} = x, {}_hY_{kh} = y]. \end{aligned} \quad (2.12)$$

Assumption 5. For some δ , $0 < \delta < 1$, and for every $\eta > 0$,

$$\lim_{h \downarrow 0} \sup_{\|(x', y')\| \leq \eta} \|c_{h,\delta}(x, y) - c(x, y)\| = 0, \quad (2.13)$$

where for all $x \in R^n$, $c(x, 0) = 0$, and

$$\lim_{h \downarrow 0} \sup_{\|(x', y')\| \leq \eta} \|d_{h,\delta}(x, y)\| = 0. \quad (2.14)$$

While the drift and conditional second moment of the increments in $\{{}_hX_{kh}\}$ are $O(h)$, Assumption 5 guarantees that the drift and (uncentered) second moment of the increments in $\{{}_hY_{kh}\}$ are $O(h^\delta)$ and $o(h^\delta)$, respectively. This has two important implications: first, $\{{}_hY_t\}$ operates on a faster time scale than $\{{}_hX_t\}$, since the drift (and possibly the variance) per unit of time of $\{{}_hY_t\}$ grow at a faster rate (as $h \downarrow 0$) than the drift and variance per unit of time of $\{{}_hX_t\}$. This implies that if $\{{}_hY_t\}$ mean-reverts to a vector of zeros, it does so with increasing speed as $h \downarrow 0$. Second, as $h \downarrow 0$ the drift of $\{{}_hY_t\}$ [which is $O(h^\delta)$] dominates the variance of $\{{}_hY_t\}$ [which is $o(h^\delta)$]. This allows us to approximate the behavior of $\{{}_hY_t\}$ by a *deterministic* differential equation.

⁶See Ethier and Kurtz (1986, pp. 290–291) for formal definitions. Several sets of sufficient conditions for Assumption 4 are summarized in appendix A of Nelson (1990a).

The next assumption assures that this differential equation is well-behaved, pulling $\{_h Y_t\}$ back to a vector of zeros:

Assumption 6. For each $x \in R^n$, $y \in R^m$, define the differential equation

$$dY(t, x, y)/dt = c(x, Y(t, x, y)), \quad (2.15)$$

with initial condition

$$Y(0, x, y) = y. \quad (2.16)$$

Then $0_{m \times 1}$ is a globally asymptotically stable solution of (2.15)–(2.16) for bounded values of x, y – i.e., for every $\eta > 0$,

$$\lim_{t \rightarrow \infty} \sup_{\|(x', y')\| \leq \eta} \|Y(t, x, y)\| = 0.^7 \quad (2.17)$$

Finally, we require a Lyapunov condition which guarantees that $\{_h Y_t\}$ does not diverge to infinity in finite time:⁸

Assumption 7. There exists a nonnegative function $\rho(x, y, h)$, twice differentiable in x and y , and a positive function $\lambda(\eta, h)$ such that

$$\lim_{\eta \rightarrow \infty} \liminf_{h \downarrow 0} \inf_{\|(x', y')\| \geq \eta} \rho(x, y, h) = \infty, \quad (2.18)$$

$$\limsup_{\eta \rightarrow \infty} \limsup_{h \downarrow 0} \lambda(\eta, h) < \infty, \quad (2.19)$$

$$\limsup_{h \downarrow 0} E[\rho(_h X_0, _h Y_0, h)] < \infty, \quad (2.20)$$

and for every $\eta > 0$,

$$\begin{aligned} & \limsup_{h \downarrow 0} \sup_{\|(x', y')\| \leq \eta} h^{-1} E[\rho(_h X_{(k+1)h}, _h Y_{(k+1)h}, h) \\ & - \rho(x, y, h)|_{_h X_{kh} = x, _h Y_{kh} = y}] - \lambda(\eta, h)\rho(x, y, h) \leq 0. \end{aligned} \quad (2.21)$$

⁷It is often possible to verify Assumption 6 without solving (2.15)–(2.16). For example, the following Lyapunov condition suffices: There is a nonnegative $V(x, y)$, bounded on compact (x, y) sets and differentiable in y , functions $\kappa(\eta)$ and $g(x)$, and a $\Delta > 0$ such that for each $N > 0$, $\inf_{0 \leq \eta \leq N} \kappa(\eta) > 0$, $\inf_{0 \leq \|x\| \leq \eta} g(x) > 0$, $\|y\|^\Delta g(x) \leq V(x, y)$, and for each $\eta > 0$ and all $y \in R^m$, $\sup_{\|x\| \leq \eta} \kappa(\eta)V(x, y) + (\partial V(x, y)/\partial y)' \cdot c(x, y) \leq 0$.

⁸This condition also guarantees that $\{X_t\}$ does not explode in finite time, but this was already implied by Assumption 4 [see Stroock and Varadhan (1979, ch. 10)].

Theorem 2.2 (Ethier–Nagylaki). *Let Assumptions 1–7 hold. Then*

$${}_h Y_t \rightarrow 0_{m \times 1} \quad \text{in probability as } h \downarrow 0 \quad \text{for every } t > 0. \quad (2.22)$$

3. Examples of consistent ARCH filters

In this section, we prove filter consistency for three ARCH models [multivariate GARCH(1,1), AR(1) EGARCH, and the Taylor–Schwert model] when (1.1) generates $\{X_t\}$ – i.e., we let $\{X_t\}$ be a solution to (1.1), and then define for all $t \geq 0$, ${}_h X_t \equiv X_{[t/h] \cdot h}$ where $[t/h]$ is the integer part of t/h . We then expand these results to allow $\{X_t\}$ to be generated by a stochastic difference equation.

3.1. Multivariate GARCH(1,1)

In the multivariate GARCH(1,1) model of Bollerslev, Engle, and Wooldridge (1988), conditional covariance estimates are formed by the recursion

$$\begin{aligned} \text{vech} \left[{}_h \hat{\Omega}_{1:q, 1:q, (k+1)h} \right] &= W_h + B_h \text{vech} \left[{}_h \hat{\Omega}_{1:q, 1:q, kh} \right] \\ &\quad + h^{-1} A_h \text{vech} \left[{}_h \hat{\xi}_{1:q, kh} {}_h \hat{\xi}_{1:q, kh}^T \right], \end{aligned} \quad (3.1)$$

where $\text{vech}(\cdot)$ stacks the columns of the lower portion of a symmetric matrix. W_h , B_h , and A_h are $\frac{1}{2}(q+1)q \times \frac{1}{2}(q+1)q$ matrices. The $\{{}_h \hat{\xi}_{1:q, kh}\}$ in (3.1) are fitted residuals obtained using the (possibly misspecified) drift $\hat{\mu}_h({}_h X_{kh}, {}_h \hat{\Omega}_{1:q, 1:q, kh})$:

$${}_h \hat{\xi}_{1:q, kh} \equiv {}_h X_{1:q, (k+1)h} - {}_h X_{1:q, kh} - h \cdot \hat{\mu}_h({}_h X_{kh}, {}_h \hat{\Omega}_{1:q, 1:q, kh}). \quad (3.2)$$

For some δ , $0 < \delta < 1$, let W_h , A_h , B_h , and $\hat{\mu}_h(\cdot, \cdot)$ satisfy

$$\hat{\mu}_h = o(h^{-1/2}),^9 \quad (3.3)$$

$$W_h = o(h^\delta), \quad (3.4)$$

$$I - B_h - A_h = o(h^\delta), \quad (3.5)$$

$$A_h = h^\delta A + o(h^\delta), \quad (3.6)$$

$^9 f(x, y, h) = O(h^\gamma)$ means that for every positive finite η , there is a $\lambda \geq 0$ such that

$$\limsup_{h \downarrow 0} \sup_{\|x', y'\| < \eta} h^{-\gamma} \|f(x, y, h)\| \leq \lambda.$$

That is, $h^{-\gamma} f$ is uniformly bounded in h for bounded (x, y) . If $\lambda = 0$, then we say that $f(x, y, h) = o(h^\gamma)$.

where I is a $\frac{1}{2}(q+1)q \times \frac{1}{2}(q+1)q$ identity matrix and A is a $\frac{1}{2}(q+1)q \times \frac{1}{2}(q+1)q$ matrix independent of h . Defining

$${}_h Y_t \equiv \text{vech} \left[{}_h \hat{\Omega}_{1:q, 1:q, t} - \Omega_{1:q, 1:q, t} \right], \quad (3.7)$$

we have under regularity conditions given below,

$$h^{-1} E \left[{}_h Y_{(k+1)h} - {}_h Y_{kh} | {}_h X_{kh} = x, {}_h Y_{kh} = y \right] = -h^{\delta-1} A \cdot y + O(1). \quad (3.8)$$

${}_h Y_t$ is the time t estimation error – i.e., the vech of the difference between the estimated and true conditional covariances. For each $h > 0$, $\{{}_h Y_t\}$ fluctuates randomly, but mean-reverts to a vector of zeros if A is well-behaved. As $h \downarrow 0$, the speed of this mean-reversion goes to infinity, implying under regularity conditions that for each $t > 0$, ${}_h Y_t$ converges in probability to a vector of zeros as $h \downarrow 0$. All of the filters presented in this section achieve consistency in this way – i.e., by increasingly rapid mean reversion in $\{{}_h Y_t\}$ as $h \downarrow 0$.

We next give regularity conditions on the diffusion (1.1), on $\hat{\mu}_h$, W_h , A_h , and B_h , and on the starting value ${}_h \hat{\Omega}_{1:q, 1:q, 0}$ guaranteeing that Assumptions 1–7 in section 2 hold. We are then able to apply Theorem 2.2 to prove that the measurement error ${}_h Y_t$ converges to zero in probability for each $t > 0$ as $h \downarrow 0$. We emphasize that the conditions given below, while *sufficient* for filter consistency, may or may not be *necessary*. We use Conditions 3.1–3.3 repeatedly in this section. Conditions 3.4–3.5 are specific to the GARCH(1, 1) case. While these conditions impose conditional moment and nonexplosion conditions on the diffusion (1.1), they do not impose stationarity or ergodicity.

Condition 3.1. For each $h > 0$, (1.1) generates $\{{}_h X_t\}$ and satisfies Assumptions 1, 3, and 4 of section 2.

Condition 3.2. For some $\varepsilon > 0$, $\limsup_{h \downarrow 0} E \| {}_h Y_0 \|^{2+\varepsilon} < \infty$.

Condition 3.3. There is a twice differentiable, nonnegative $\omega(x)$ and a $\theta > 0$ such that for every $\eta > 0$,

$$\liminf_{\|x\| \rightarrow \infty} \omega(x) = \infty, \quad (3.9)$$

$$\lim_{h \downarrow 0} E [\omega({}_h X_0)] < \infty, \quad (3.10)$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq \eta} h^{-1} E \left[\left(\left| \omega({}_h X_{(k+1)h}) - \omega(x) \right|^{1+\theta} | {}_h X_{kh} = x \right) \right] < \infty, \quad (3.11)$$

and there is a $\lambda > 0$ for all $x \in \mathbb{R}^n$,

$$\sum_{i=1}^n \mu_i(x) \frac{\partial \omega(x)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \Omega_{i,j}(x) \frac{\partial^2 \omega(x)}{\partial x_i \partial x_j} \leq \lambda \omega(x). \quad (3.12)$$

We use Conditions 3.2–3.3 to verify Assumption 7. Condition 3.2 is trivially satisfied if ${}_h\Omega_{1:q,1:q,0}$ and ${}_h\hat{\Omega}_{1:q,1:q,0}$ are nonrandom and independent of h . Though Condition 3.3 looks formidable, it is often easy to verify, since verifying Assumption 4 usually involves a condition close to Condition 3.3. [See the examples and discussion in Nelson (1990a, esp. app. A).]

Condition 3.4. For every $\eta > 0$, there is an $\varepsilon > 0$ such that for every i, j , $1 \leq i, j \leq q$,

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq \eta} h^{-1} \mathbb{E} \left[\left| \Omega_{i,j}({}_hX_{(k+1)h}) - \Omega_{i,j}(x) \right|^{2+\varepsilon} \middle| {}_hX_{kh} = x \right] = 0, \quad (3.13)$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq \eta} h^{-1} \mathbb{E} \left[\left| {}_hX_{i,(k+1)h} - x_i \right|^{4+\varepsilon} \middle| {}_hX_{kh} = x \right] = 0. \quad (3.14)$$

Condition 3.5. W_h , A_h , and B_h satisfy (3.3)–(3.5). All the eigenvalues of A have strictly positive real parts.

The differential equation of Assumption 6 is now

$$dY/dt = -A \cdot Y, \quad (3.15)$$

which has the unique solution $Y_t = \exp(-At)Y_0$ where $\exp(\cdot)$ is the matrix exponential. Condition 3.5 assures that Y satisfies Assumption 6 [Hochstadt (1975)].

Theorem 3.1. Let Conditions 3.1–3.5 hold. Then for each $t > 0$, $\|{}_hY_t\| \rightarrow 0$ in probability as $h \downarrow 0$.

The GARCH model is misspecified, since it is not the true data-generating process. Nevertheless, it achieves filter consistency as $h \downarrow 0$. The GARCH filter need not be heavily parameterized, regardless of how high q (the dimension of $X_{1:q,t}$) is. For example, let α be a positive number, set $\hat{\mu}_h \equiv 0_{q \times 1}$, $A_h = h^{1/2}\alpha I$, $B_h = (1 - h^{1/2}\alpha)I$, and $W_h = 0$, where ‘ I ’ and ‘ 0 ’ are $\frac{1}{2}(q+1)q \times \frac{1}{2}(q+1)q$ identity and zero matrices, respectively. Under Conditions 3.1–3.5, this defines a consistent filter for $\{\Omega_{1:q,1:q,t}\}$.

Under (3.3)–(3.5), $A_h + B_h$ converge to an identity matrix as $h \downarrow 0$. This also holds for GARCH models considered as data-generating processes

converging to a diffusion limit [Nelson (1990a)] and is consistent with findings in the empirical ARCH literature. For example, Baillie and Bollerslev (1989) estimated univariate GARCH(1, 1) models on exchange rate data for several currencies using various sampling intervals. The estimated values of $A_h + B_h$ using daily data were very close to one, less so using weekly data. Using monthly data, $A_h + B_h$ were much less than one.¹⁰ In interpreting results from ARCH models fit to high-frequency data, it is important to remember that values of $A_h + B_h$ close to one do not necessarily indicate nonstationarity – i.e., even for a sequence of covariance-stationary GARCH(1, 1) models converging to a covariance-stationary diffusion limit, $A_h + B_h$ converge to an identity matrix as $h \downarrow 0$.¹¹

Caution is required in applying our results when the parameters of the ARCH model (i.e., $\hat{\mu}_h$, W_h , A_h , and B_h) are random, as they are when they are estimated, for example by quasi-maximum likelihood. There are reasons to believe that our results apply when W_h , A_h , and B_h are estimated and the data are generated by a diffusion: first, because the class of consistent filters in Theorem 3.1 is very broad, and second, because better conditional covariance estimates should produce higher quasi-likelihoods. This heuristic is not a proof, however.

For a given GARCH model to make sense as a data-generating process, the conditional covariance matrix ${}_h\hat{\Omega}_{kh}$ must be nonnegative definite for all (h, k) with probability one. This need not be true of a consistent filter. It is easy to construct examples of sequences of GARCH(1, 1) models which are consistent filters, but in which ‘variance’ becomes *negative* with probability one *when the model is considered as a data-generating process*. For example, let a univariate IGARCH(1, 1) model [Bollerslev (1986)] generate $\{{}_h\sigma_{kh}^2\}$:

$$\alpha_h = h^{1/2}\alpha, \quad (3.16)$$

$$\beta_h = 1 - h^{1/2}\alpha, \quad 0 < \alpha < \infty, \quad 0 < h < 1/\alpha^2, \quad (3.17)$$

$$\omega_h = h \cdot \omega, \quad (3.18)$$

$${}_h\sigma_{(k+1)h}^2 = \omega_h + \beta_h \cdot {}_h\sigma_{kh}^2 + h^{-1}\alpha_h \cdot {}_h\sigma_{kh}^2 \cdot {}_h z_{kh}^2, \quad (3.19)$$

$${}_h z_{kh} \sim \text{i.i.d. } N(0, h). \quad (3.20)$$

¹⁰However, Baillie and Bollerslev (1991) find that when even finer time increments are used (i.e., hourly) MLE estimates of $A_h + B_h$ begin to move away from one. It may be that lumpy information and/or market microstructure effects destroy the validity of the diffusion approximation at extremely high frequencies. On the other hand, Chan, Chan, and Karolyi (1990) estimate higher-order GARCH models on intra-day data and find $\sum_j A_{j,h} + B_{j,h}$ close to one again.

¹¹The assumption of a diffusion limit here is critical – it may well be that convergence to, say, a jump-diffusion limit can occur without $A_h + B_h$ converging to an identity matrix.

When $\omega \geq 0$, ${}_h\sigma_{kh}^2$ remains nonnegative with probability one for all (h, k) , and the process has a well-defined diffusion limit [Nelson (1990a)]. For each h and k we can define ${}_h\xi_{kh} \equiv {}_h z_{kh} \cdot {}_h\sigma_{kh}$ and write

$$h^{-1} \text{var}[_h\xi_{kh} | {}_h\mathcal{L}_{kh}] = {}_h\sigma_{kh}^2. \quad (3.21)$$

When $\omega < 0$, however, (3.16)–(3.20) imply that for every h , $\{{}_h\sigma_{kh}^2\}$ eventually becomes negative (for large k) with probability one and remains negative forever [Nelson (1990b)]. Clearly such a model yields unacceptable long-term forecasts if we insist on interpreting $\{{}_h\sigma_{kh}^2\}$ as a conditional variance! Nevertheless, (3.16)–(3.19) define a consistent filter, even when $\omega < 0$.

The misspecification in the conditional mean and covariance processes permitted by Theorem 3.1 is even wider. For example, while the true drift per unit of time $\mu(x)$ is bounded for bounded values of x , (3.3) permits the drift per unit of time $\hat{\mu}_h$ assumed in the ARCH model to explode to infinity as $h \downarrow 0$. Next, we have ignored the state variables $X_{q+1:n,t}$. As a final example of the ways in which the ARCH model may misspecify the data-generating mechanism yet still provide consistent filtering, suppose we replace the vech operators in (3.1) with ordinary vec operators: we could then create ${}_h\hat{\Omega}_{1:q,1:q,t}$ matrices that are *asymmetric* with probability one for all (h, t) , and yet are consistent filters.¹²

3.2. AR(1) EGARCH

As a discrete time data-generating process, the AR(1) Exponential ARCH model is given by

$${}_hX_{1,kh} = {}_hX_{1,(k-1)h} + h \cdot \mu_h({}_hX_{1,(k-1)h}, {}_h\sigma_{kh}^2, kh) + {}_h z_{kh} \cdot {}_h\sigma_{kh}, \quad (3.22)$$

$$\begin{aligned} \ln({}_h\sigma_{(k+1)h}^2) &= \ln({}_h\sigma_{kh}^2) - \beta_h \cdot h \cdot [\ln({}_h\sigma_{kh}^2) - \alpha_h] \\ &\quad + \theta_h \cdot {}_h z_{kh} + \gamma_h [{}_h z_{kh} - m \cdot h^{1/2}], \end{aligned} \quad (3.23)$$

$${}_h z_{kh} \sim \text{i.i.d. } N(0, h), \quad (3.24)$$

$$m \equiv E[|h^{-1/2} {}_h z_{kh}|] = (2/\pi)^{1/2}. \quad (3.25)$$

¹²This distinction between the behavior of a model as a filter and its behavior as a data-generating process is found in linear models as well. For example, a stationary series that has been exponentially smoothed remains stationary. As a data-generating process, however, an exponential smoothing model corresponds to an ARIMA(0, 1, 1) model, which is nonstationary [Granger and Newbold (1986)].

Nelson (1990a) shows that if $\mu_h \rightarrow \mu$ uniformly on compacts,

$$\begin{aligned}\beta_h &= \beta + o(1), & \alpha_h &= \alpha + o(1), \\ \theta_h &= \theta + o(1), & \gamma_h &= \gamma + o(1),\end{aligned}\tag{3.26}$$

if the starting values ${}_hX_0$ and $\ln({}_h\sigma_0^2)$ converge in distribution, and if μ satisfies fairly mild regularity conditions, then as $h \downarrow 0$, the sequence of step function processes $\{{}_hX_{1,t}, {}_h\sigma_t^2\}_{h \downarrow 0}$ generated by (3.22)–(3.26) and (2.2) converge weakly to the diffusion¹³

$$dX_{1,t} = \mu(X_t, \sigma_t^2, t) dt + \sigma_t dW_{1,t},\tag{3.27}$$

$$d\ln(\sigma_t^2) = -\beta(\ln(\sigma_t^2) - \alpha) dt + dW_{2,t},\tag{3.28}$$

where $W_{1,t}$ and $W_{2,t}$ are one-dimensional Brownian motions with no drift and with

$$\begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} [dW_{1,t} \quad dW_{2,t}] = \begin{bmatrix} 1 & \theta \\ \theta & \theta^2 + (1 - m^2)\gamma^2 \end{bmatrix} dt.\tag{3.29}$$

But suppose, once again, that the data are generated by (1.1) and that a misspecified EGARCH model is used to produce an estimate of the true underlying conditional variance process $\{\Omega_{1,t}(X_t)\}$. We generate fitted $\{{}_h\hat{\sigma}_{kh}^2\}$ recursively by (3.23) with some ${}_h\hat{\sigma}_0^2$, and generate $\{{}_h\hat{z}_{kh}\}$ by

$${}_h\hat{z}_{kh} \equiv \left[{}_hX_{1,kh} - {}_hX_{1,(k-1)h} - h \cdot \hat{\mu}_h({}_hX_{(k-1)h}, {}_h\hat{\sigma}_{kh}^2, kh) \right] / {}_h\hat{\sigma}_{kh}.\tag{3.30}$$

When the model is misspecified, $\{{}_h\hat{z}_{kh}\}$ is neither i.i.d. nor $N(0, h)$. The requirements we place on $\hat{\mu}_h$, β_h , α_h , θ_h , and γ_h to achieve consistent filtering are much weaker than (3.26): for some δ , $0 < \delta < 1$, we require

$$\hat{\mu}_h = o(h^{-1/2})\tag{3.31}$$

$$\beta_h = o(h^{\delta-1}),\tag{3.32}$$

$$\alpha_h\beta_h = o(h^{\delta-1}),\tag{3.33}$$

$$\theta_h = o(h^{(\delta-1)/2}),\tag{3.34}$$

$$\gamma_h = \gamma \cdot h^{\delta-1/2} + o(h^{\delta-1/2}) \quad \text{where } \gamma > 0.\tag{3.35}$$

¹³This diffusion is employed in the options pricing models of Hull and White (1987) and Wiggins (1987).

Defining the measurement error

$${}_h Y_t \equiv \left[\ln({}_h \hat{\sigma}_t^2) - \ln({}_h \Omega_{1,1,t}) \right], \quad (3.36)$$

we have

$$\begin{aligned} c(x, y) &\equiv \lim_{h \downarrow 0} h^{-\delta} \mathbb{E} \left[{}_h Y_{(k+1)h} - {}_h Y_{kh} \mid {}_h X_{kh} = x, {}_h Y_{kh} = y \right] \\ &= \gamma \cdot m \cdot [\exp(-y/2) - 1]. \end{aligned} \quad (3.37)$$

The drift per unit of time in $\{{}_h Y_{kh}\}$ is approximately $h^{\delta-1} \cdot c(x, y)$. By (3.37), this drift explodes to $+\infty$ ($-\infty$) whenever ${}_h Y_{kh}$ is below (above) zero, causing the measurement error to mean-revert to zero with increasing speed and vanish in the limit as $h \downarrow 0$. The differential equation of Assumption 6 is

$$dY/dt = \gamma \cdot m \cdot [\exp(-Y/2) - 1], \quad (3.38)$$

which is globally asymptotically stable when $\gamma > 0$. Condition 3.6 replaces Condition 3.4:

Condition 3.6. For every $\eta > 0$, there is an $\varepsilon > 0$ such that

$$\begin{aligned} \lim_{h \downarrow 0} \sup_{\|x\| \leq \eta} h^{-1} \mathbb{E} \left[\left| \ln(\Omega_{1,1}({}_h X_{(k+1)h})) - \ln(\Omega_{1,1}(x)) \right|^{2+\varepsilon} \mid {}_h X_{kh} = x \right] \\ = 0. \end{aligned} \quad (3.39)$$

Theorem 3.2. Let (3.31)–(3.36) and Conditions 3.1–3.3 and 3.6 hold. Then ${}_h Y_t \rightarrow 0$ in probability for every $t > 0$ as $h \downarrow 0$.

As was the case with GARCH(1,1), the effect of misspecification in the conditional mean washes out of the conditional variance as $h \downarrow 0$. This is in accord with experience from estimated ARCH models. For example, Nelson (1991) fit an EGARCH model to daily stock index returns computed two different ways: first, using capital gains, and second, using including dividends and a (crude) adjustment for the riskless interest rate. Since dividends and riskless interest rates are almost entirely predictable one day ahead, their inclusion in the returns series amounts to a perturbation in the conditional mean of the capital gains series. The inclusion or exclusion of dividends and riskless rates made virtually no difference in the fitted conditional variances of the two series, which had nearly identical in-sample means and variances and a correlation of approximately 0.9996. Considering our results, this is not

surprising.¹⁴ Similarly, Schwert (1990) found that neglecting dividends made little difference in stock volatility estimates even in monthly data. Note also that (3.23) and (3.32) imply an approach to a unit root in $\{\ln({}_h\sigma_{kh}^2)\}$ as $h \downarrow 0$. As in the GARCH(1, 1) case, this holds for the model both as a filter and as a data-generating process and does not necessarily imply nonstationarity.¹⁵

3.3. The Taylor–Schwert variant

Univariate GARCH(1, 1) creates its estimate of the conditional variance as a distributed lag of squared residuals. An equally natural procedure is to estimate the conditional standard deviation using a distributed lag of absolute residuals. Taylor (1986) employed this method to estimate conditional variances for several financial time series. Schwert (1989) used a closely related procedure, generating conditional standard deviations with a 12th-order moving average of absolute residuals. We follow Taylor (1986) by using an AR(1). As a data-generating process, the system is

$${}_hX_{1,kh} = {}_hX_{1,(k-1)h} + h \cdot \mu_h({}_hX_{1,(k-1)h}, {}_h\sigma_{kh}, kh) + {}_hz_{kh} \cdot {}_h\sigma_{kh}, \quad (3.40)$$

$${}_h\sigma_{(k+1)h} = \omega_h + \beta_h \cdot {}_h\sigma_{kh} + h^{-1/2} \alpha_h |{}_hz_{kh}| {}_h\sigma_{kh}, \quad (3.41)$$

$${}_hz_{kh} \sim \text{i.i.d. } N(0, h) \quad (3.42)$$

where ω_h , β_h , and α_h are nonnegative, and the distribution $\nu_h(x, \sigma)$ of $\{{}_hX_{1,0}, {}_h\sigma_0\}$ converges to a limit $\nu_0(x, \sigma)$ as $h \downarrow 0$. The $\{{}_hX_{1,t}, {}_h\sigma_t\}$ are formed using (2.2).¹⁶

The properties of the procedure as a filter are similar to GARCH(1, 1). Let ${}_hX_{1,t}$ consist of the first element of ${}_hX_t$: as in the EGARCH case, define an initial value ${}_h\hat{\sigma}_0$ and define $\{{}_h\hat{z}_{kh}\}$ and $\{{}_h\hat{\sigma}_{kh}\}$ by (3.30) and (3.41). Again,

¹⁴Merton (1980) demonstrated the near-irrelevance of the mean as $h \downarrow 0$ when the variance is constant.

¹⁵Once again, this is in accord with the empirical literature – compare, for example, the AR parameters generated in Nelson (1991) using daily data and in Glosten, Jagannathan, and Runkle (1989) using monthly data. See also Priestley (1981, section 3.7.2) for a related discussion of passage to continuous time in the linear AR(1) model. Sims (1984) uses a related result to explain the martingale-like behavior of asset prices and interest rates.

¹⁶Higgins and Bera (1989) nest the Taylor–Schwert model and GARCH in a class of ‘NARCH’ (nonlinear ARCH) models. They make $\sigma_t^{2\delta}$ a distributed lag of absolute residuals each raised to the 2δ power. The Taylor–Schwert model corresponds to $\delta = \frac{1}{2}$, while univariate GARCH sets $\delta = 1$.

define $m \equiv (2/\pi)^{1/2}$. For some δ with $0 < \delta < 1$, let

$$\omega_h = o(h^\delta), \quad (3.43)$$

$$\alpha_h = h^\delta \alpha + o(h^\delta) \quad \text{with} \quad \alpha > 0, \quad (3.44)$$

$$\beta_h = 1 - h^\delta \alpha \cdot m + o(h^\delta), \quad (3.45)$$

$$\hat{\mu}_h = o(h^{-1/2}). \quad (3.46)$$

Defining

$${}_h Y_t \equiv ({}_h \hat{\sigma}_t - {}_h \Omega_{1,1,t}^{1/2}), \quad (3.47)$$

the differential equation (2.15) is

$$dY = -\alpha \cdot m \cdot Y dt, \quad (3.48)$$

which satisfies Assumption 6.¹⁷

Condition 3.7. For every $\eta > 0$, there is an $\varepsilon > 0$ such that

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq \eta} h^{-1} \mathbb{E} \left[\left| \Omega_{1,1}^{1/2}({}_h X_{(k+1)h}) - \Omega_{1,1}^{1/2}(x) \right|^{2+\varepsilon} \middle| {}_h X_{kh} = x \right] = 0. \quad (3.49)$$

Theorem 3.3. Under (3.43)–(3.46) and Conditions 3.1–3.3 and 3.7, ${}_h Y_t \rightarrow 0$ in probability as $h \downarrow 0$ for every $t > 0$.

3.4. A generalization: From diffusions to near-diffusions

Thus far we have generated $\{{}_h X_t\}$ by the stochastic integral equation (1.1) observed at intervals of length h . This can be generalized: it is not necessary that $\{{}_h X_t\}$ is generated by a diffusion, though it is essential that the sequence

¹⁷Considered as a data-generating process, the model's diffusion limit is also similar to GARCH(1,1)'s. An application of Theorem 2.1 shows that if $\nu_h \rightarrow \nu_0$ as $h \downarrow 0$, $\omega_h = h\omega$, $\alpha_h = h^{1/2}\alpha$, $\beta_h = 1 - h^{1/2}\alpha \cdot m - \theta h$, with ω and α nonnegative, and if $\mu(\cdot)$ satisfies minimal regularity conditions, $\{{}_h X_{1,t}, {}_h \sigma_t\}$ converges weakly to the solution of

$$dX_{1,t} = \mu(X_{1,t}, \sigma_t, t) dt + \sigma_t dW_{1,t},$$

$$d\sigma_t = (\omega - \theta\sigma_t) dt + \alpha(1 - m^2)^{1/2} \sigma_t dW_{2,t},$$

as $h \downarrow 0$, where $\{W_{1,t}\}$ and $\{W_{2,t}\}$ are independent Brownian motions.

of processes $\{ {}_hX_t \}_{h \downarrow 0}$ converges to the diffusion (1.1) as $h \downarrow 0$. We call a process $\{ {}_hX_t \}$ embedded in such a sequence a *near-diffusion*.

For example $\{ {}_hX_t \}$ may be generated by a stochastic volatility model [e.g., Melino and Turnbull (1990)] or by an ARCH model. This requires minor modifications to Theorems 3.1–3.3:

Condition 3.1'. The sequence of processes $\{ {}_hX_t \}_{h \downarrow 0}$ is generated as in section 2 and satisfies Assumptions 1–4.

Condition 3.3'. There exists a twice differentiable, nonnegative function $\omega(x)$ such that

$$\lim_{\|x\| \rightarrow \infty} \omega(x) = \infty, \quad (3.50)$$

$$\limsup_{h \downarrow 0} E[\omega({}_hX_0)] < \infty, \quad (3.51)$$

and there exists a $\lambda > 0$ such that for all $x \in R^n$,

$$\limsup_{h \downarrow 0} h^{-1} E[\omega({}_hX_{(k+1)h}) - \omega(x) | {}_hX_{kh} = x] - \lambda \omega(x) \leq 0. \quad (3.52)$$

Theorem 3.4. Let Conditions 3.1' and 3.3' replace Conditions 3.1 and 3.3 in the statement of Theorem 3.1. Then the conclusions of Theorems 3.1 hold. Next, let the function

$$m(h, x) = E\left[h^{-1/2} \Omega_{1,1}(x)^{-1/2} | {}_hX_{1,(k+1)h} - x_1 | | {}_hX_{kh} = x \right] \quad (3.53)$$

be well-defined for all $x \in R^n$, and let

$$m(h, x) = m + o(1), \quad (3.54)$$

where m is a constant,¹⁸ and the $o(1)$ term in (3.54) vanishes uniformly on bounded x sets as $h \downarrow 0$. Let Conditions 3.1' and 3.3' replace Conditions 3.1 and 3.3 and (3.54) replace (3.25) in the statements of Theorems 3.2–3.3. Then the conclusions of Theorems 3.2–3.3 hold.

¹⁸When the increments in $\{X_{1,t}\}$ are conditionally normal, $m = (2/\pi)^{1/2}$. In general, however, it is not. When $\{ {}_hX_t \}$ is generated by (1.1), the increments in $\{ {}_hX_t \}$ are approximately normal as $h \downarrow 0$, so $m = (2/\pi)^{1/2}$.

4. Inconsistent ARCH filters

These consistency results do not, alas, apply to all ARCH models and all data-generating processes. There are at least three important cases in which consistency breaks down:

4.1. Not using enough lagged residuals

All of the consistent filters discussed in the paper achieved consistency by (in effect) employing a law of large numbers to extract covariance estimates from an infinite number of lagged residuals. This suggests that consistent filtering will break down if the number of lagged residuals used to form the covariance estimates remain bounded as $h \downarrow 0$. For example, the ARCH(p) model of Engle (1982) and the seminonparametric (SNP) model of Gallant, Rossi, and Tauchen (1990) employ only a finite number of residuals, p , in forming the conditional densities. If $p \rightarrow \infty$ at a suitable rate as $h \downarrow 0$, then consistency could no doubt be achieved. This may be difficult to prove formally, however: the convergence theorems of section 2 allowed only a finite number of state variables. Both for consistent filtering and for parsimony therefore, GARCH(p, q) may be preferable to ARCH(p) in applications where the data are observed at high frequencies and are generated by a near-diffusion. Similarly, it may be useful to extend the SNP model to allow infinite (geometrically declining) lags of residuals to appear in the conditional distribution of the observable variables.

4.2. Restrictions on the functional form of ${}_h\hat{\Omega}_{1:q, 1:q, t}$

The Factor ARCH model of Engle, Ng, and Rothschild (1990) illustrates another way in which consistency may break down. This model imposes the condition that there is a $p < q$, and, for each $h > 0$, $q \times 1$ vectors β_h^i , $i = 1, \dots, p$, a $q \times q$ matrix Ω_h , and nonnegative scalar random processes $\{\lambda_h^i\}$, $i = 1, \dots, p$, such that for all $t \geq 0$,

$${}_h\hat{\Omega}_{1:q, 1:q, t} = \Omega_h + \sum_{i=1}^p \beta_h^i \beta_h^{i'} \lambda_h^i. \quad (4.1)$$

It is easy to see that (4.1) implies that for any $s \geq 0$ and $t \geq 0$, ${}_h\hat{\Omega}_{1:q, 1:q, t} - {}_h\hat{\Omega}_{1:q, 1:q, s}$ is singular. This generally prevents consistent filtering, since the difference between arbitrary $\Omega_{1:q, 1:q, t}$ and $\Omega_{1:q, 1:q, s}$ need not be singular.

The model of Schwert and Seguin (1990) imposes similar restrictions on the estimated conditional covariance matrix:

$${}_h\hat{\Omega}_{1:q, 1:q, t} = \Omega_h + \Sigma_h \cdot {}_h\sigma_t + \Lambda_h \cdot {}_h\sigma_t^2, \quad (4.2)$$

where Ω_h , Σ_h , and Λ_h are $q \times q$ symmetric matrices and $\{_h\sigma_t^2\}$ is the conditional variance of returns on a market index. (4.2) constrains the $q(q+1)/2$ distinct elements of $_h\hat{\Omega}_{1:q,1:q,t}$ to lie in a one-dimensional subspace of $R^{q(q+1)/2}$ for all t . When $q > 1$, consistency breaks down if the model is misspecified.

Restrictions on conditional covariances matrices such as those imposed by Engle, Ng, and Rothschild (1990) and Schwert and Seguin (1990) are useful because they achieve parsimony. This considerable virtue comes with a price, namely the loss of consistency when the model is misspecified.

4.3. Failure of the near-diffusion assumption

Consider again the $\hat{\Omega}_{1,1,T}(\Delta, M)$ discussed in the introduction, but with $\{X_{1,t}\}$ generated by a continuous time Poisson process instead of the diffusion (1.1). The sample paths of $\{X_{1,t}\}$ are (almost surely) step functions with a finite number of steps over any finite time interval. For almost every t therefore, $\hat{\Omega}_{1,1,T}(\Delta, M) \rightarrow 0$ in probability as $\Delta \downarrow 0$ and $M \rightarrow \infty$. Yet the instantaneous conditional variance in the increments in $\{X_{1,t}\}$ is nonnegative, so $\hat{\Omega}_{1,1,T}(\Delta, M)$ does not consistently estimate $\Omega_{1,1,T}$. Similarly, if $\{X_t\}$ is a jump-diffusion, consistency for $\hat{\Omega}_{1:q,1:q,T}$ generated by misspecified ARCH models breaks down, since except in the immediate aftermath of a jump, the ARCH models effectively ignore the contribution of the jump component to the instantaneous covariance matrix.¹⁹

The near-diffusion assumption is crucial for two reasons: first, it guarantees that the increments in $\{X_t\}$ are small over small time intervals (i.e., they are not too thick-tailed). Second, it guarantees that Ω_t changes by only small amounts over small time intervals. These two conditions effectively allowed us to apply laws of large numbers in estimating Ω_t . When these two assumptions break down, so, in general, does consistency. Unfortunately, so too will consistency for the other procedures commonly used for estimating time-varying conditional variances and covariances [e.g., Beckers (1983), Cox and Rubinstein (1985, sect. 6.1), French, Schwert, and Stambaugh (1987), Fama and MacBeth (1973), Schwert (1989), Taylor (1986)].²⁰

¹⁹The situation is not quite so bleak if the jumps occur in $\Omega_{1:q,1:q,t}$ but not in $X_{1:q,t}$. Recall that our consistency results did not assume $_h\hat{\Omega}_{1:q,1:q,0} = _h\Omega_{1:q,1:q,0}$, which is why we had $_h\hat{\Omega}_{1:q,1:q,t} - _h\Omega_{1:q,1:q,t} \rightarrow 0_{q \times q}$ in probability for positive t . As far as the filtering problem is concerned, a jump in $\Omega_{1:q,1:q,t}$ essentially resets the clock back to time 0, when $_h\hat{\Omega}_{1:q,1:q,0}$ and $_h\Omega_{1:q,1:q,0}$ were not equal. If the probability of a jump in $_h\Omega_{1:q,1:q,t}$ at any given t equals zero, then the consistency results of sections 2 and 3 should still hold – i.e., $_h\hat{\Omega}_{1:q,1:q,t} - _h\Omega_{1:q,1:q,t} \rightarrow 0_{q \times q}$ in probability (again, pointwise) at t . Of course, $_h\hat{\Omega}_{1:q,1:q,t}$ will not converge to $_h\Omega_{1:q,1:q,t}$ when t is a jump point, but the probability of this at any fixed t equals zero.

²⁰The method in Bates (1991) may be an exception.

The possibility of large jumps in asset prices plays an important role in some asset pricing theories. For example, in a regime of fixed exchange rates, there is typically a large probability of no change in the exchange rate over a small interval and a small probability of a large move due to a currency revaluation. [See, for example, Krasker's (1980) discussion of the 'peso problem'.] While the diffusion assumption is commonly made in asset pricing models, it is not without its dangers even apart from the peso problem: we know, for example, that asset prices cannot *literally* follow diffusion processes: markets are not always open (they are typically closed on evenings, weekends, and holidays), stocks do not trade at every instant even when the markets are open (leading to discrete price changes when they do trade), prices change in discrete units ($\frac{1}{8}$ th's for most stocks), stocks in portfolios trade nonsynchronously, and prices are observed sometimes at the bid and sometimes at the asked. These and other facts of market microstructure make the diffusion assumption, if pushed too far, unrealistic.²¹ Even if we could correct for such market microstructure considerations, it might be that 'true' underlying asset prices do not change smoothly – as in the peso problem. Information may arrive in a lumpy manner better characterized by a mixed jump-diffusion process.

Our results are relevant to the extent that the estimation error becomes small for h large enough that market microstructure problems haven't destroyed the approximate validity of our diffusion assumption. Whether this is actually true is an open question. The usual large-sample asymptotics employed in time series analysis suffer from a similar defect: it is almost always unrealistic to assume that parameters (or hyperparameters in a random coefficients model) remain fixed *forever* – i.e., as the sample period and sample size T goes to infinity. If taken literally, the notion of consistency breaks down. Nevertheless, large- T consistency is a useful concept, partly because it imposes discipline in the choice of parameter estimators. Similarly, small- h consistency may help impose a useful discipline on the choice of ARCH models.

5. Conclusion

Traditionally, econometricians have regarded heteroskedasticity mainly as an obstacle to the efficient estimation of regression parameters. In recent years, however, financial economists have stressed that the behavior of conditional variances and covariances is important in and of itself. Unfortu-

²¹Similar warnings are in order when diffusion models are applied in almost any field. For example, Brownian motion is often used by physicists to model particle motion, even though a particle cannot *literally* follow a Brownian motion: Brownian sample paths have unbounded variation almost surely, so a particle following a Brownian motion would, with probability one, travel an infinite distance in a finite time period, a physical impossibility.

nately, traditional linear time series models are simply not capable of modelling dynamic conditional variances and covariances in a natural way. ARCH models have greatly expanded the ability of econometricians to model heteroskedastic time series, particularly financial time series.

If the near-diffusion assumption is approximately valid for high-frequency financial time series, our results may explain some features of estimated ARCH models: first, the tendency to find 'persistent' volatility (e.g., IGARCH) when using high-frequency (e.g., daily) data, second, the small influence of misspecification in the conditional mean on the estimated conditional variance, and third, estimated models (especially in the multivariate case) that do not make sense as data-generating processes (i.e., in which the conditional covariance matrix loses nonnegative definiteness with positive probability).

Our results should be reassuring to users of ARCH models and suggest a reason for the forecasting success of ARCH: If the process generating prices is (approximately) a diffusion, then there is so much information about conditional second moments at high frequencies that even a misspecified model can be a consistent filter. This does not suggest, however, that the econometrician should be indifferent as to which ARCH model to use: a consistent filter may be a very poor long-term forecaster, and there may be considerable differences in the efficiency of different ARCH filters. These issues remain for future work.

Appendix

Proof of Theorem 2.1. See Ethier and Kurtz (1986, ch. 7, theorem 4.1) or Nelson (1990a, theorem 2.2).

Proof of Theorem 2.2. This theorem is a special case of Ethier and Nagylaki (1988, theorem 2.1). There are a few changes in notation: we have set ε_N , δ_N , $X^N(k)$, $Y^N(k)$, and $Z^N(k)$ in Ethier and Nagylaki equal to h , h^δ , ${}_hX_{kh}$, ${}_hY_{kh}$, $({}_hX_{kh}, {}_hY_{kh}, kh)$, respectively. By our (2.4), Ethier and Nagylaki's $h(x, y)$ function [see their (2.1)–(2.2)] is a vector of zeros. Our Theorem 2.2 now directly follows by Ethier and Nagylaki (1988, theorem 2.1). Q.E.D.

In proving the theorems in section 3, we need the following lemma:

Lemma A.1. Let $\{\zeta_t\}_{[0, T]}$ be generated by the stochastic integral equation

$$\zeta_t = \zeta_0 + \int_0^t m(\zeta_s, s) ds + \int_0^t A^{1/2}(\zeta_s, s) dW_s, \quad (\text{A.1})$$

where ζ_0 is fixed, $\{W_s\}$ is a $q \times 1$ standard Brownian motion, $m(\cdot, \cdot)$ and $\Lambda(\cdot, \cdot)$ are continuous, $(q \times 1)$ - and $(q \times q)$ -valued functions, respectively, and where (A.1) has a unique weak-sense solution. Next, let $g(t)$ be a $q \times 1$ function satisfying

$$g(t) = o(t^{1/2}) \quad \text{as } t \downarrow 0. \quad (\text{A.2})$$

Then,

$$t^{-1/2}(\zeta_t - g(t) - \zeta_0) \xrightarrow{d} N(0_{q \times 1}, \Lambda(\zeta_0, 0)) \quad \text{as } t \downarrow 0. \quad (\text{A.3})$$

Further, let $f(\zeta, t)$ be a continuous function from R^{q+1} into R^1 . Let there exist an $\varepsilon > 0$ such that

$$\limsup_{t \downarrow 0} E \left| f(t^{-1/2}(\zeta_t - g(t) - \zeta_0), t) \right|^{1+\varepsilon} < \infty. \quad (\text{A.4})$$

Then if $Z \sim N(0_{q \times 1}, \Lambda(\zeta_0, 0))$,

$$\lim_{t \downarrow 0} E \left[f(t^{-1/2}(\zeta_t - g(t) - \zeta_0), t) \right] = E[f(Z, 0)] < \infty. \quad (\text{A.5})$$

Proof of Lemma A.1. Once we prove (A.3), (A.5) follows directly by Billingsley (1986, theorem 25.7, corollary a and theorem 25.12, corollary). To prove (A.3), note that by (A.2), $t^{-1/2}g(t) \rightarrow 0$ as $t \downarrow 0$. Next, rewrite (A.1) as

$$\begin{aligned} t^{-1/2}(\zeta_t - \zeta_0) &= t^{-1/2}\Lambda^{1/2}(\zeta_0, 0)W_t + t^{-1/2} \int_0^t m(\zeta_s, s) ds \\ &\quad + t^{-1/2} \int_0^t [\Lambda^{1/2}(\zeta_s, s) - \Lambda^{1/2}(\zeta_0, 0)] dW_s. \end{aligned} \quad (\text{A.6})$$

Since $t^{-1/2}W_t \sim N(0, I)$ and $t^{-1/2}g(t) \rightarrow 0_{q \times 1}$ as $t \downarrow 0$, (A.3) follows if the last two terms on the right-hand side of (A.6) converge in probability to zero as $t \downarrow 0$. The middle term is

$$\left| t^{-1/2} \int_0^t m(\zeta_s, s) ds \right| \leq t^{1/2} \max_{0 \leq s \leq t} |m(\zeta_s, s)|, \quad (\text{A.7})$$

which converges to a vector of zeros in probability as $t \downarrow 0$. Next, consider the

last term. Define

$$Z_t \equiv \int_0^t [A^{1/2}(\xi_s, s) - A^{1/2}(\xi_0, 0)] dW_s. \quad (\text{A.8})$$

The proof of (A.3) is complete if $t^{-1/2}Z_t$ converges in probability to a vector of zeros as $t \downarrow 0$. Letting $I(\cdot)$ be an indicator function, we have

$$t^{-1}Z_t \cdot Z'_t = I(\|Z_t\| < 1) \cdot t^{-1}Z_t \cdot Z'_t + I(\|Z_t\| \geq 1) \cdot t^{-1}Z_t \cdot Z'_t. \quad (\text{A.9})$$

The second term on the right-hand side of (A.9) vanishes in probability as $t \downarrow 0$ since

$$\lim_{t \downarrow 0} E[t^{-1} \cdot I(\|Z_t\| \geq 1) | \xi_0] = 0_{q \times q} \quad (\text{A.10})$$

[Arnold (1973, p. 40, formula (a))]. Since A is continuous, the first term also vanishes [Arnold (1973, p. 40, formula (b))]. Q.E.D.

Proof of Theorem 3.1. We need only verify Assumptions 1–7. Condition 3.1 includes Assumptions 1, 3, and 4. Assumption 3 and (1.1) directly imply Assumption 2 [see Arnold (1973, p. 40)]. Define

$$\begin{aligned} h^{-1/2} {}_h\xi_{1:q, (k+1)h} &\equiv h^{-1/2} [{}_hX_{1:q, (k+1)h} - {}_hX_{1:q, kh} \\ &\quad - h \cdot \hat{\mu}({}_hX_{kh}, {}_h\hat{\Omega}_{1:q, 1:q, kh})]. \end{aligned} \quad (\text{A.11})$$

Applying Condition 3.4 and Lemma A.1 we have for all $x \in R^{n+m}$,

$$\left[h^{-1/2} {}_h\xi_{1:q, (k+1)h} | {}_hX_{kh} = x \right] \Rightarrow N(0, \Omega_{1:q, 1:q}(x)), \quad (\text{A.12})$$

as $h \downarrow 0$. By Condition 3.4 and Lemma A.1, the conditional absolute moments of $h^{-1/2} {}_h\xi_{1:q, kh}$ (given ${}_hX_{kh} = x$) of order less than $4 + \varepsilon$ converge to the corresponding moments of a $N(0, \Omega_{1:q, 1:q}(x))$. It is then easy to derive $c(y) = -A \cdot y$ and $d(y) = 0$, as required by Assumption 5. The differential equation of Assumption 6 is $dY/dt = -A \cdot y$ with unique solution $Y_t = \exp(-At)Y_0$ [see, e.g., Hochstadt (1975)]. The solution is globally asymptotically stable if the eigenvalues of A have strictly positive real parts. To verify

Assumption 7, define $\psi(y) \equiv y'y \cdot [1 - \exp(-y'y)]$ and

$$\rho(x, y) \equiv 1 + \omega(x) + \psi(y). \quad (\text{A.13})$$

(2.18) is immediate and Condition 3.2 implies (2.20). Condition 3.3 implies

$$\begin{aligned} \limsup_{h \downarrow 0} \sup_{\|x', y'\| \leq \eta} h^{-1} \mathbb{E} \left[\omega({}_h X_{(k+1)h}) - \omega(x) \mid {}_h X_{kh} = x, {}_h Y_{kh} = y \right] \\ - \lambda \omega(x) \leq 0, \end{aligned} \quad (\text{A.14})$$

so to verify (2.21) it suffices that there is a $\lambda(\eta, h)$ satisfying (2.19) such that for every $\eta > 0$,

$$\begin{aligned} \limsup_{h \downarrow 0} \sup_{\|x', y'\| \leq \eta} h^{-1} \mathbb{E} \left[\psi({}_h Y_{i, (k+1)h}) - \psi(y) \mid {}_h X_{kh} = x, {}_h Y_{kh} = y \right] \\ - \lambda(\eta, h) [1 + \psi(y)] \leq 0. \end{aligned} \quad (\text{A.15})$$

Using (1.1) and (3.1)–(3.2), we have

$${}_h Y_{(k+1)h} = {}_h Y_{kh} + {}_h D_{kh} + {}_h N_{kh}, \quad (\text{A.16})$$

where ${}_h D_{kh}$ and ${}_h N_{kh}$ are, respectively, drift and noise terms defined by

$${}_h D_{kh} \equiv \mathbb{E} [{}_h Y_{(k+1)h} - {}_h Y_{kh} \mid {}_h \mathcal{L}_{kh}], \quad (\text{A.17})$$

$${}_h N_{kh} \equiv {}_h Y_{(k+1)h} - {}_h Y_{kh} - {}_h D_{kh}. \quad (\text{A.18})$$

Condition 3.4 guarantees that ${}_h D_{kh}$ is well-defined and that the conditional variance of ${}_h N_{kh}$ is bounded in compact regions of the state space. By Conditions 3.1, 3.4, and 3.5,

$$\begin{aligned} \lim_{h \downarrow 0} \sup_{\|x', y'\| \leq \eta} h^{-1} \mathbb{E} \left[\psi({}_h Y_{i, (k+1)h}) - \psi(y) \mid {}_h X_{kh} = x, {}_h Y_{kh} = y \right] \\ = - \frac{\partial \psi h^{\delta-1}}{\partial y} (Ay + o(1)) + \frac{\partial^2 \psi}{2 \partial y \partial y'} \cdot [O(1) + O(h^{2\delta-1})]. \end{aligned} \quad (\text{A.19})$$

Since the eigenvalues of A have strictly positive real parts, $-(\partial\psi/\partial y)Ay$ is negative for $y \neq 0_{m \times 1}$. Hence, the first term on the right-hand side of (A.19) dominates the second term (since it diverges to $-\infty$) except in a shrinking neighborhood of $y = 0_{m \times 1}$ (i.e., in neighborhoods of the form $\|y\| \leq K \cdot h^\Delta$ for $K > 0$ and $0 < \Delta < \min\{1 - \delta, \delta\}$). In this shrinking neighborhood of $y = 0_{m \times 1}$, $\partial\psi/\partial y \rightarrow 0$ and $\partial^2\psi/\partial y \partial y' \rightarrow 0$, so the right-hand side of (A.19) is uniformly bounded on compacts, diverging to $-\infty$ except at $y = 0_{m \times 1}$. (A.15) can therefore be satisfied with a $\lambda(\eta, h)$ satisfying (2.19). (2.21) now follows directly. Q.E.D.

Proof of Theorem 3.2. Assumptions 1–4 follow directly from Conditions 3.1–3.3. The proof that Assumptions 5 and 7 hold closely follows the proof of Theorem 3.1, with Condition 3.6 taking the place of Condition 3.4. The differential equation of Assumption 6 is $dY/dt = \gamma \cdot m \cdot [\exp(-Y/2) - 1]$, with solution

$$Y_t = 2 \cdot \ln\left(1 - \exp[-\gamma m \cdot t/2] + \exp\left[\frac{1}{2}(Y_0 - \gamma m t)\right]\right), \quad (\text{A.20})$$

which satisfies Assumption 6 as long as $\gamma > 0$. Q.E.D.

Proofs of Theorems 3.3 and 3.4. Nearly identical to the proofs of Theorem 3.1 and Theorems 3.1–3.3, respectively. Q.E.D.

References

- Anderson, B.D.O. and J.B. Moore, 1979, *Optimal filtering* (Prentice Hall, Englewood Cliffs, NJ).
- Arnold, L., 1973, *Stochastic differential equations: Theory and applications* (Wiley, New York, NY).
- Baillie, R.T. and T. Bollerslev, 1989, The message in daily exchange rates: A conditional variance tale, *Journal of Business and Economic Statistics* 7, 297–305.
- Baillie, R.T. and T. Bollerslev, 1990, Intra-day and inter-market volatility in foreign exchange rates, *Review of Economic Studies* 58, 565–585.
- Bates, D.S., 1991, The crash of '87: Was it expected? The evidence from options markets, *Journal of Finance* 46, 1009–1044.
- Beckers, S., 1983, Variances of security price returns based on high, low, and closing prices, *Journal of Business* 56, 97–112.
- Billingsley, P., 1986, *Probability and measure*, 2nd ed. (Wiley, New York, NY).
- Black, F. and M. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637–654.
- Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307–327.
- Bollerslev, T., R.Y. Chou, and K. Kroner, 1992, ARCH modeling in finance: A review of the theory and empirical evidence, *Journal of Econometrics*, this issue.
- Bollerslev, T., R.F. Engle, and J.M. Wooldridge, 1988, A capital asset pricing model with time varying covariances, *Journal of Political Economy* 96, 116–131.
- Chan, K., K.C. Chan, and G.A. Karolyi, 1991, Intraday volatility in the stock index and stock index futures markets, *Review of Financial Studies*, forthcoming.

- Cox, J. and M. Rubinstein, 1985, Options markets (Prentice Hall, Englewood Cliffs, NJ).
- 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, 1986, Modelling the persistence of conditional variances, *Econometric Reviews* 5, 1–50.
- Engle, R.F., V. Ng, and M. Rothschild, 1990, Asset pricing with a factor ARCH covariance structure: Empirical estimates for Treasury bills, *Journal of Econometrics* 45, 213–238.
- Ethier, S.N. and T.G. Kurtz, 1986, Markov processes: Characterization and convergence (Wiley, New York, NY).
- Ethier, S.N. and T. Nagylaki, 1988, Diffusion approximations of markov chains with two time scales and applications to population genetics, II, *Advances in Applied Probability* 20, 525–545.
- Fama, E.F. and J.D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.
- French, K.R., G.W. Schwert, and R.F. Stambaugh, 1987, Expected stock returns and volatility, *Journal of Financial Economics* 19, 3–29.
- Gallant, A.R., P.E. Rossi, and G. Tauchen, 1990, Stock prices and volume, Manuscript (University of Chicago, Chicago, IL).
- Glosten, L.R., R. Jagannathan, and D. Runkle, 1989, Relationship between the expected value and the volatility of the nominal excess return on stocks, Manuscript (Northwestern University, Evanston, IL).
- Granger, C.W.J. and P. Newbold, 1986, Forecasting economic time series, 2nd ed. (Academic Press, New York, NY).
- Hardouvelis, G.A., 1990, Margin requirements, volatility, and the transitory component of stock prices, *American Economic Review* 80, 736–762.
- Harvey, C.R., 1989, Time-varying conditional covariances in tests of asset pricing models, *Journal of Financial Economics* 24, 289–318.
- Higgins, M.L. and A.K. Bera, 1989, A class of nonlinear ARCH models, *International Economic Review*, forthcoming.
- Hochstadt, H., 1975, Differential equations: A modern approach (Dover Press, New York, NY).
- Hsieh, D.A. and M.H. Miller, 1990, Margin regulation and stock market volatility, *Journal of Finance* 45, 3–30.
- Hull, J. and A. White, 1987, The pricing of options on assets with stochastic volatilities, *Journal of Finance* 42, 281–300.
- Krasker, W.S., 1980, The peso problem in testing for efficiency of foreign exchange markets, *Journal of Monetary Economics* 6, 269–276.
- Lintner, J., 1965, The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13–37.
- Melino, A. and S. Turnbull, 1990, Pricing foreign currency options with stochastic volatility, *Journal of Econometrics* 45, 239–266.
- Merton, R.C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.
- Merton, R.C., 1980, On estimating the expected return on the market, *Journal of Financial Economics* 8, 323–361.
- Nelson, D.B., 1990a, ARCH models as diffusion approximations, *Journal of Econometrics* 45, 7–39.
- Nelson, D.B., 1990b, Stationarity and persistence in the GARCH(1,1) model, *Econometric Theory* 6, 318–334.
- Nelson, D.B., 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347–370.
- Priestley, M.B., 1981, Spectral analysis and time series (Academic Press, London).
- Ross, S., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.
- Schwert, G.W., 1989, Why does stock market volatility change over time?, *Journal of Finance* 44, 1115–1154.
- Schwert, G.W., 1990, Indexes of U.S. stock prices, *Journal of Business* 63, 399–426.
- Schwert, G.W. and P.J. Seguin, 1990, Heteroskedasticity in stock returns, *Journal of Finance* 45, 1129–1156.

- Shanken, J., 1990, Intertemporal asset pricing: An empirical investigation, *Journal of Econometrics* 45, 99–120.
- Sharpe, W., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.
- Sims, C., 1984, Martingale-like behavior of asset prices and interest rates, Manuscript (University of Minnesota, Minneapolis, MN).
- Stroock, D.W. and S.R.S. Varadhan, 1979, *Multidimensional diffusion processes* (Springer Verlag, Berlin).
- Summers, L.H. and V.P. Summers, 1989, When financial markets work too well: A cautious case for a securities transactions tax, *Journal of Financial Services Research* 3, 261–286.
- Taylor, S., 1986, *Modeling financial time series* (Wiley, New York, NY).
- Wiggins, J.B., 1987, Option values under stochastic volatility: Theory and empirical estimates, *Journal of Financial Economics* 19, 351–372.