

NBER TECHNICAL WORKING PAPER SERIES

FILTERING AND FORECASTING
WITH MISSPECIFIED ARCH MODELS II:
MAKING THE RIGHT FORECAST
WITH THE WRONG MODEL

Daniel B. Nelson

Dean P. Foster

Technical Working Paper No. 132

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 1992

We thank Arnold Zellner and two anonymous referees for helpful comments. This material is based on work supported by the National Science Foundation under grant #SES9110131. We thank the Graduate School of Business of the University of Chicago, the Center for Research in Security Prices, and the William S. Fishman Research Scholarship for additional research support. This paper is part of NBER's research program in Asset Pricing. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

FILTERING AND FORECASTING
WITH MISSPECIFIED ARCH MODELS II:
MAKING THE RIGHT FORECAST
WITH THE WRONG MODEL

ABSTRACT

A companion paper (Nelson (1992)) showed that in data observed at high frequencies, an ARCH model may do a good job at estimating conditional variances, even when the ARCH model is severely misspecified. While such models may perform reasonably well at *filtering* (i.e., at estimating unobserved instantaneous conditional variances) they may perform disastrously at medium and long term *forecasting*. In this paper, we develop conditions under which a misspecified ARCH model successfully performs both tasks, filtering and forecasting. The key requirement (in addition to the conditions for consistent filtering) is that the ARCH model correctly specifies the functional form of the first two conditional moments of all state variables. We apply these results to a diffusion model employed in the options pricing literature, the stochastic volatility model of Hull and White (1987), Scott (1987), and Wiggins (1987).

Daniel B. Nelson
The University of Chicago
Graduate School of Business
1101 East 58th Street
Chicago, IL 60637
and NBER

Dean P. Foster
The Wharton School
University of Pennsylvania
Steinberg-Dietrich Hall
3620 Locust Walk
Philadelphia, PA 19104

1. Introduction

Since their introduction by Engle (1982), ARCH models have been widely (and quite successfully) applied in modeling financial time series; see, for example, the survey paper of Bollerslev, Chou, and Kroner (1992). What accounts for the success of these models? A companion paper (Nelson (1992)) suggested one reason: under fairly mild conditions,¹ high frequency data contain a great deal of information about conditional variances, and as a continuous time limit is approached, the sample information about conditional variances increases without bound. This allows simple volatility estimates formed, for example, by taking a distributed lag of squared residuals (as in a GARCH model), to consistently estimate conditional variances as the time interval between observations goes to zero. It is not surprising, therefore, that as continuous time is approached, a sequence of GARCH models can consistently estimate the underlying conditional variance of a diffusion, even when the GARCH models are not the correct data generating process.

As shown in Nelson (1992), this continuous time consistency holds not only for GARCH, but for many other ARCH models as well, and is unaffected by a wide variety of misspecifications. For example, when considered as a data generating process, a given ARCH model might provide nonsensical forecasts--for example, by forecasting a negative conditional variance with positive probability, incorrectly forecasting explosions in the state variables, ignoring unobservable state variables, or by misspecifying the conditional mean. Nevertheless, in the limit as continuous time is approached, such a model can provide a consistent estimate of the instantaneous conditional variance in the underlying data

1) i.e., if the observed data is generated by a diffusion or near-diffusion (i.e., by a stochastic process imbedded in a sequence of processes converging weakly to a diffusion).

generating process. (Our use of the term "estimation" corresponds to its use in the filtering literature rather in the statistics literature--i.e., the ARCH model "estimates" the conditional variance in the sense that a Kalman filter estimates unobserved state variables. See, e.g., Anderson and Moore (1979, Chapter 2), or Arnold (1973, Chapter 12).) In other words, while a misspecified ARCH model may perform disastrously in medium or long-term forecasting, it may perform well at filtering.

In this paper, we show that under suitable conditions, a sequence of misspecified ARCH models may not only be successful at filtering, but at forecasting as well. That is, as a continuous time limit is approached, not only do the conditional covariance matrices generated by the sequence of ARCH models approach the true conditional covariance matrix, but the forecasts generated by these models converge in probability to the forecasts generated by the true data generating process.

The conditions for consistent estimation of the conditional distribution are considerably stricter than the conditions for consistent filtering; for example, *all* unobservable state variables must be consistently estimated and we must correctly specify the conditional mean and covariance of *all* state variables as continuous time is approached. These conditions are developed in Section 2.

While the conditions in Section 2 are stricter than the consistent filtering conditions developed in Nelson (1992), they are broad enough to accommodate a number of interesting cases. In Section 3, we provide a detailed example, using a stochastic volatility model familiar in the options pricing literature (see, e.g., Wiggins (1987), Hull and White (1987), Scott (1987), and Melino and Turnbull (1990)). In this model, a stock price and its

instantaneous returns volatility follow a diffusion process. While the stock price is observable at discrete intervals of length h , the instantaneous volatility is unobservable. We show that a suitably constructed sequence of ARCH models (in particular, AR(1) EGARCH models) can consistently estimate the instantaneous volatility and generate appropriate forecasts of the stock price and volatility processes as $h \downarrow 0$.

2. Main results

In this section, we develop our basic results on consistent estimation of forecast distributions. We begin by taking the nonnegative real line, chopping it up into pieces of length h , and considering, for each h , a stochastic processes $\{{}_hX_t, {}_hU_t\}$ which is a step function with jumps only (with probability one) at integer multiples of h --i.e., at times h , $2h$, $3h$, and so on. We interpret $\{{}_hX_t\}$ as the $n \times 1$ *discretely observable* component of the process and $\{{}_hU_t\}$ as the $m \times 1$ *unobservable* component. Associated with the true data generating process is a probability measure P_h , which we define below.

The misspecified ARCH model produces an estimate ${}_h\hat{U}_t$ of the true unobservable state variables ${}_hU_t$. Associated with the ARCH model is a (misspecified) probability measure \hat{P}_h for $\{{}_hX_t, {}_hU_t, {}_h\hat{U}_t\}$. The ARCH model incorporates the false assumption that ${}_hU_t = {}_h\hat{U}_t$ for all t almost surely. Our interest is in comparing forecasts made using information at a time T with the *incorrect* probability measure \hat{P}_h to those made using the *correct* measure P_h . Specifically, how can we characterize the forecasts generated by P_h and \hat{P}_h as $h \downarrow 0$? Under what circumstances do they become "close" as $h \downarrow 0$? Theorems 2.3 and 2.4 below compare the conditional forecast *distributions* generated by P_h and \hat{P}_h , while

Theorem 2.5 compares the conditional forecast *moments*. Finally, Theorem 2.6 compares the very long term forecasts (i.e., the forecast stationary distributions) generated by P_h and \hat{P}_h .

The Formal Setup

Let $D([0, \infty), \mathbb{R}^n \times \mathbb{R}^{2m})$ be the space of functions from $[0, \infty)$ into $\mathbb{R}^n \times \mathbb{R}^{2m}$ that are continuous from the right with finite left limits. D is a metric space when endowed with the Skorohod metric (see Ethier and Kurtz (1986, Chapter 3) for formal definitions). For each $h > 0$, let \mathcal{F}_t be the σ -algebra generated by ${}_h\hat{U}_0, {}_hX_0, {}_hX_h, {}_hX_{2h}, \dots, {}_hX_{kh},$ and ${}_hU_0, {}_hU_h, {}_hU_{2h}, \dots, {}_hU_{kh}$ for all $kh \leq t$. In our notation, curly brackets indicate a stochastic process--e.g., $\{{}_hX_t, {}_hU_t\}_{[0, T]}$ is the sample path of ${}_hX_t$ and ${}_hU_t$ as (random) functions of time on the interval $0 \leq t \leq T$. We refer to values of a process at a *particular* time t by omitting the curly brackets--e.g., ${}_hX_\tau$ and $({}_hX_\tau, {}_hU_\tau)$ are, respectively, the (random) values taken by the $\{{}_hX_t\}$ and $\{{}_hX_t, {}_hU_t\}$ processes at time τ .

Next, let $\mathcal{B}(E)$ denote the Borel sets on a metric space E , and let ν_h and $\hat{\nu}_h$ be probability measures on $(\mathbb{R}^{n+2m}, \mathcal{B}(\mathbb{R}^n \times \mathbb{R}^{2m}))$. Below, we will take ν_h and $\hat{\nu}_h$ to be, respectively, probability measures for the starting values $({}_hX_0, {}_hU_0, {}_h\hat{U}_0)$ under the true data generating process and under the misspecified ARCH model.

The functional limit theorems we employ will require $\{{}_hX_{kh}, {}_hU_{kh}, {}_h\hat{U}_{kh}\}_{k=0, \infty}$ to have a first order Markov structure under both P_h and \hat{P}_h . Accordingly, our next step is to introduce the Markov transition probabilities associated with the true data generating process and the misspecified ARCH model respectively: For each $h > 0$, let $\Pi_h(x, u, \hat{u}, \cdot)$ and

$\hat{\Pi}_h(x, u, \hat{u}, \cdot)$ be transition functions on $R^n \times R^{2m}$, i.e., $\Pi_h(x, u, \hat{u}, \cdot)$ and $\hat{\Pi}_h(x, u, \hat{u}, \cdot)$ are probability measures on $(R^n \times R^{2m}, \mathcal{B}(R^n \times R^{2m}))$ for all $(x, u, \hat{u}) \in R^n \times R^{2m}$, and $\Pi_h(\cdot, \cdot, \cdot, \Gamma)$ and $\hat{\Pi}_h(\cdot, \cdot, \cdot, \Gamma)$ are $\mathcal{B}(R^n \times R^{2m})$ measurable for all $\Gamma \in \mathcal{B}(R^n \times R^{2m})$. Expectations evaluated under P_h and \hat{P}_h are respectively denoted $E_h[\cdot]$ and $\hat{E}_h[\cdot]$.

In our examples, $\{\hat{U}_{kh}\}$ consists either of the unique elements of the estimated conditional covariance matrix generated by an ARCH model or some invertible function of these conditional covariances. In ARCH models, the conditional covariance matrix is a function of past values of ${}_hX_{kh}$ along with a startup value for the estimated conditional covariance matrix. In particular, we assume that ${}_h\hat{U}_{kh}$ is a measurable function of ${}_h\hat{U}_{(k-1)h}$, ${}_hX_{(k-1)h}$, and ${}_hX_{kh}$. This in turn implies that given the observed data ${}_hX_0, {}_hX_h, {}_hX_{2h}, \dots, {}_hX_{kh}$ and the startup value ${}_h\hat{U}_0$, we can derive the whole sequence $\{{}_h\hat{U}_{kh}\}_{j=1,k}$. Formally, we require that for each $h > 0$, there is a measurable function U_h from R^{2n+m} into R^m such that for all (x, u, \hat{u}) ,

$$\int_{\{\hat{u}^* = U_h(x^*, x, \hat{u})\}} \hat{\Pi}_h(x, u, \hat{u}, d(x^*, u^*, \hat{u}^*)) = \int_{\{\hat{u}^* = U_h(x^*, x, \hat{u})\}} \Pi_h(x, u, \hat{u}, d(x^*, u^*, \hat{u}^*)) = 1. \quad (1)$$

We also assume that the misspecified ARCH model generated by $\hat{\nu}_h$ and $\hat{\Pi}_h$ treats u and \hat{u} as being identical almost surely--i.e., for all $h > 0$, and all $(x, u, \hat{u}) \in R^{n+2m}$

$$\int_{\{u^* = \hat{u}^*\}} \hat{\Pi}_h(x, u, \hat{u}, d(x^*, u^*, \hat{u}^*)) = 1, \quad (2)$$

and

$$\int_{\{u^* = \hat{u}^*\}} \hat{\nu}_h(d(x^*, u^*, \hat{u}^*)) = 1. \quad (3)$$

Finally, we assume that there is no feedback from the ARCH estimate $\{{}_h\hat{U}_{kh}\}_{k=0,1,2,\dots}$ into

$\{ {}_hX_{kh}, {}_hU_{kh} \}_{k=0,1,2,\dots}$ —i.e., given ${}_hX_{kh}$ and ${}_hU_{kh}$, ${}_hX_{(k+1)h}$ and ${}_hU_{(k+1)h}$ are independent of ${}_h\hat{U}_{kh}$ under P_h . Formally, for any $\Gamma \in \mathcal{B}(R^{n+m})$ and all $h > 0$, we require that

$$\Pi_h(x, u, \hat{u}, \Gamma \times R^m) = \Pi_h(x, u, 0_{m \times 1}, \Gamma \times R^m), \quad (4)$$

where “ $0_{m \times 1}$ ” is an $m \times 1$ vector of zeros. Now let P_h and \hat{P}_h be the probability measures on $D([0, \infty), R^n \times R^{2m})$ such that

$$P_h[({}_hX_0, {}_hU_0, {}_h\hat{U}_0) \in \Gamma] = \nu_h(\Gamma) \text{ for any } \Gamma \in \mathcal{B}(R^n \times R^{2m}), \quad (5)$$

$$\hat{P}_h[({}_hX_0, {}_hU_0, {}_h\hat{U}_0) \in \Gamma] = \hat{\nu}_h(\Gamma) \text{ for any } \Gamma \in \mathcal{B}(R^n \times R^{2m}), \quad (5')$$

$$P_h[({}_hX_t, {}_hU_t, {}_h\hat{U}_t) = ({}_hX_{kh}, {}_hU_{kh}, {}_h\hat{U}_{kh}), kh \leq t < (k+1)h] = 1, \quad (6)$$

$$\hat{P}_h[({}_hX_t, {}_hU_t, {}_h\hat{U}_t) = ({}_hX_{kh}, {}_hU_{kh}, {}_h\hat{U}_{kh}), kh \leq t < (k+1)h] = 1, \quad (6')$$

and finally, for all $k \geq 0$ and $\Gamma \in \mathcal{B}(R^n \times R^{2m})$,

$$P_h[({}_hX_{(k+1)h}, {}_hU_{(k+1)h}, {}_h\hat{U}_{(k+1)h}) \in \Gamma \mid \mathcal{F}_{kh}] = \Pi_h({}_hX_{kh}, {}_hU_{kh}, {}_h\hat{U}_{kh}, \Gamma) \quad (7)$$

almost surely under P_h , and

$$\hat{P}_h[({}_hX_{(k+1)h}, {}_hU_{(k+1)h}, {}_h\hat{U}_{(k+1)h}) \in \Gamma \mid \mathcal{F}_{kh}] = \hat{\Pi}_h({}_hX_{kh}, {}_hU_{kh}, {}_h\hat{U}_{kh}, \Gamma) \quad (7')$$

almost surely under \hat{P}_h .

For each $h > 0$, (5) specifies the distributions of the starting point $({}_hX_0, {}_hU_0, {}_h\hat{U}_0)$ under P_h . (6) forms the continuous time process $\{ {}_hX_t, {}_hU_t, {}_h\hat{U}_t \}_{0 \leq t}$ as a step function with jumps at times $0, h, 2h, \dots$.² (7) specifies the transition probabilities for the jumps in

2) Our step function scheme follows Ethier and Kurtz (1986). Many other schemes would work just as well—for example, we could follow Stroock and Varadhan (1979, p. 267) and for $kh \leq t < (k+1)h$, set $({}_hX_t, {}_hU_t, {}_h\hat{U}_t) \equiv ({}_hX_{kh}, {}_hU_{kh}, {}_h\hat{U}_{kh}) + h^{-1}(t-kh)[({}_hX_{(k+1)h}, {}_hU_{(k+1)h}, {}_h\hat{U}_{(k+1)h}) - ({}_hX_{kh}, {}_hU_{kh}, {}_h\hat{U}_{kh})]$. This scheme makes

$\{ {}_hX_t, {}_hU_t, {}_h\hat{U}_t \}_{0 \leq t}$ (5')-(7') play the same role for \hat{P}_h as (5)-(7) do for P_h .

$D([0, \infty), R^n \times R^{2m})$ is a complete, separable metric space when endowed with the Skorohod topology, and for each $h > 0$, $(D([0, \infty), R^n \times R^{2m}), \mathcal{B}(D([0, \infty), R^n \times R^{2m})), P_h)$ and $(D([0, \infty), R^n \times R^{2m}), \mathcal{B}(D([0, \infty), R^n \times R^{2m})), \hat{P}_h)$ are probability spaces. Our interest is in characterizing the forecasts generated by P_h and \hat{P}_h . In particular, if we use the misspecified probabilities \hat{P}_h to generate forecasts regarding the future path of $\{ {}_hX_t, {}_hU_t \}$, under what circumstances will our forecast be "close" to the forecast that would have been obtained using the (correct) conditional probability generated by P_h ?³

Now consider the time τ forecasts of the future path of $\{ {}_hX_t, {}_hU_t \}$ generated by P_h . In particular, let the set $A \in \mathcal{B}(D([0, \infty), R^n \times R^m))$. (Note: $R^n \times R^m$ not $R^n \times R^{2m}$!) The conditional probability under P_h that $\{ {}_hX_t, {}_hU_t \}_{\tau \leq t < \infty} \in A$ is given by

$$\begin{aligned} P_h[\{ {}_hX_t, {}_hU_t \}_{\tau \leq t < \infty} \in A | {}_h\mathcal{F}_\tau] &= P_h[\{ {}_hX_t, {}_hU_t \}_{\tau \leq t < \infty} \in A | {}_hX_\tau, {}_hU_\tau, {}_h\hat{U}_\tau] \text{ a.s. } [P_h] \\ &= P_h[\{ {}_hX_t, {}_hU_t \}_{\tau \leq t < \infty} \in A | {}_hX_\tau, {}_hU_\tau] \text{ a.s. } [P_h]. \end{aligned} \quad (8)$$

The first equality in (8) holds since $\{ {}_hX_{kh}, {}_hU_{kh}, {}_h\hat{U}_{kh} \}$ is a Markov chain under P_h . The second equality follows from (4), the "no feedback" assumption, which assures that $\{ {}_hX_{kh}, {}_hU_{kh} \}$ is also a Markov chain under P_h . This Markov structure allows us to think of the conditional probability that $\{ {}_hX_t, {}_hU_t \}_{\tau \leq t < \infty} \in A$ as a function of the time τ and the state

$\{ {}_hX_t, {}_hU_t, {}_h\hat{U}_t \}$ piecewise linear and continuous. Presumably splines should also work in the limit as $h \downarrow 0$, but for each $h > 0$ they are "forward looking"--i.e., $\{ {}_hX_{kh}, {}_hU_{kh}, {}_h\hat{U}_{kh} \}_{k=0, \infty}$ would still be Markov, but $\{ {}_hX_t, {}_hU_t, {}_h\hat{U}_t \}_{0 \leq t < \infty}$ would not be, since ${}_h\mathcal{F}_\tau$ contains information about $\{ {}_hX_t, {}_hU_t, {}_h\hat{U}_t \}_{\tau \leq t < \infty}$ beyond the information in $({}_hX_\tau, {}_hU_\tau, {}_h\hat{U}_\tau)$.

3) For proof that these conditional probabilities are well defined, see Stroock and Varadhan (1979, Theorem 1.1.6).

variables ${}_hX_\tau$ and ${}_hU_\tau$. Informally, define the *forecast function of the set A at time τ under P_h* by:

$$F_h(A, x, u, \tau) \equiv P_h[\{{}_hX_t, {}_hU_t\}_{\tau \leq t < \infty} \in A | ({}_hX_\tau, {}_hU_\tau) = (x, u)]. \quad (9)$$

Recall that under \hat{P}_h , $\{{}_hX_{kh}, {}_hU_{kh}, {}_h\hat{U}_{kh}\}$ is also Markov chain. Due to (1)-(3), so is $\{{}_hX_{kh}, {}_h\hat{U}_{kh}\}$ (i.e., the state variable ${}_hU_t$ is redundant under \hat{P}_h , since $\hat{P}_h[{}_hU_t = {}_h\hat{U}_t \text{ for all } t] = 1$). Accordingly,

$$\begin{aligned} \hat{P}_h[\{{}_hX_t, {}_hU_t\}_{\tau \leq t < \infty} \in A | {}_h\mathcal{F}_\tau] &= \hat{P}_h[\{{}_hX_t, {}_hU_t\}_{\tau \leq t < \infty} \in A | {}_hX_\tau, {}_hU_\tau, {}_h\hat{U}_\tau] \text{ a.s. } [\hat{P}_h] \\ &= \hat{P}_h[\{{}_hX_t, {}_hU_t\}_{\tau \leq t < \infty} \in A | {}_hX_\tau, {}_h\hat{U}_\tau] \text{ a.s. } [\hat{P}_h], \end{aligned} \quad (10)$$

so we (again, informally) define the forecast function

$$\hat{F}_h(A, x, u, \tau) \equiv \hat{P}_h[\{{}_hX_t, {}_hU_t\}_{\tau \leq t < \infty} \in A | ({}_hX_\tau, {}_hU_\tau) = (x, u)]. \quad (11)$$

Unfortunately, our informal definitions of $F_h(A, x, u, \tau)$ and $\hat{F}_h(A, x, u, \tau)$ are inadequate, since the conditional expectations on the right-hand sides of (9) and (11) are random variables which are unique only up to probability 0 transforms. Since we will be evaluating conditional probabilities under both P_h and \hat{P}_h , *which may well have different sets of probability zero*, we must provide more detail on the *version* of the conditional probabilities we are using; in particular, we must carefully state how we condition on events of the form " $({}_hX_\tau, {}_hU_\tau) = (x, u)$ ". We use the version of the conditional probabilities that follows naturally from the transition probabilities $\Pi_h(x, u, \hat{u}, \cdot)$ and $\hat{\Pi}_h(x, u, \hat{u}, \cdot)$: namely, if we are conditioning on the event " $({}_hX_\tau, {}_hU_\tau, {}_h\hat{U}_\tau) = (x, u, \hat{u})$," we evaluate conditional moments and conditional probabilities by "restarting" the $\{{}_hX_t, {}_hU_t, {}_h\hat{U}_t\}$ process at time τ with $({}_hX_\tau, {}_hU_\tau, {}_h\hat{U}_\tau)$ set equal to (x, u, \hat{u}) with probability one. Since our interest will be in forecasting $\{{}_hX_t, {}_hU_t\}$ rather than $\{{}_hX_t, {}_hU_t, {}_h\hat{U}_t\}$, we will use the Markov structure of

$\{ {}_hX_t, {}_hU_t \}$ under both P_h and \hat{P}_h to "drop" $\{ {}_h\hat{U}_t \}$. Formally, we proceed as follows: for every $(x, u, \hat{u}, \tau) \in R^{n+2m+1}$, define measures $P_{(h,x,u,\tau)}$ and $\hat{P}_{(h,x,u,\tau)}$ on $(D([\tau, \infty), R^n \times R^m), \mathcal{B}(D([\tau, \infty), R^n \times R^m)))$ by replacing P_h , \hat{P}_h and " $k \geq 0$ " in (6)-(7) and (6')-(7') with $P_{(h,x,u,\tau)}$, $\hat{P}_{(h,x,u,\tau)}$ and " $k \geq \text{Int}[\tau/h]$ ", and replacing (5)-(5') with

$$P_{(h,x,u,\tau)}[({}_hX_\tau, {}_hU_\tau, {}_h\hat{U}_\tau) = (x, u, u)] = 1,^4 \text{ and} \quad (5'')$$

$$\hat{P}_{(h,x,u,\tau)}[({}_hX_\tau, {}_hU_\tau, {}_h\hat{U}_\tau) = (x, u, u)] = 1. \quad (5''')$$

Then (formally) define $F_h(A, x, u, \tau)$ and $\hat{F}_h(A, x, u, \tau)$ by

$$F_h(A, x, u, \tau) \equiv P_{(h,x,u,\tau)}[\{ {}_hX_t, {}_hU_t \}_{\tau \leq t < \infty} \in A] \quad (9')$$

$$\hat{F}_h(A, x, u, \tau) \equiv \hat{P}_{(h,x,u,\tau)}[\{ {}_hX_t, {}_hU_t \}_{\tau \leq t < \infty} \in A]. \quad (11')$$

Note that $F_h(A, x, u, \tau)$ and $\hat{F}_h(A, x, u, \tau)$ are now functions of A , x , u , and τ , and are *not* random variables, since they are defined in terms of *unconditional* rather than *conditional* probabilities. (In essence we have constructed $F_h(A, x, u, \tau)$ and $\hat{F}_h(A, x, u, \tau)$ directly from the measures $\Pi_h(x, u, \hat{u}, \cdot)$ and $\hat{\Pi}_h(x, u, \hat{u}, \cdot)$ without resorting to a sample space.) $F_h(A, {}_hX_\tau, {}_hU_\tau, \tau)$ and $\hat{F}_h(A, {}_hX_\tau, {}_h\hat{U}_\tau, \tau)$, however, *are* random variables on the same sample space. A natural definition of "close" time τ forecasts generated by P_h and \hat{P}_h is that $F_h(A, {}_hX_\tau, {}_hU_\tau, \tau)$ and $\hat{F}_h(A, {}_hX_\tau, {}_h\hat{U}_\tau, \tau)$ are close with high probability--i.e., that for every $\epsilon > 0$ and every "well-behaved" set A , $P_h[|F_h(A, {}_hX_\tau, {}_hU_\tau, \tau) - \hat{F}_h(A, {}_hX_\tau, {}_h\hat{U}_\tau, \tau)| > \epsilon] \rightarrow 0$ as $h \downarrow 0$ --that is, the difference between the forecasts generated by the two models converges to zero in probability under P_h . We postpone our formal definition until the admissible sets A are characterized.

4) It is without loss of generality in (5'') that ${}_hU_\tau = {}_h\hat{U}_\tau = u$, since the forecasts generated using these probabilities will regard only the future paths of $\{ {}_hX_t, {}_hU_t \}$, and under (1)-(4) once ${}_hU_\tau$ is fixed the value of ${}_h\hat{U}_\tau$ is irrelevant.

There are three natural steps in proving that \hat{P}_h consistently estimates the forecast distribution for P_h over the interval $[\tau, \infty)$: first, we show that given ${}_h\hat{U}_\tau = {}_hU_\tau$, the forecasts generated at time τ by P_h and \hat{P}_h become arbitrarily close as $h \downarrow 0$ --i.e., for $\tau > 0$ and for every $(x, u) \in \mathbb{R}^{n+m}$, $F_h(A, x, u, \tau) - \hat{F}_h(A, x, u, \tau) \rightarrow 0$ as $h \downarrow 0$. This first step, considered in detail in Nelson (1990), is concerned with the continuous limit properties of ARCH models as data generating processes. The results are summarized in Lemma 2.1. The second step is to show that \hat{P}_h is a consistent filter for P_h at time τ --i.e., that ${}_h\hat{U}_\tau - {}_hU_\tau \rightarrow 0$ in probability under P_h as $h \downarrow 0$. This second step is concerned with the *filtering* properties of misspecified ARCH models, considered in Nelson (1992). These results are summarized in Lemma 2.2. The third step is to show that the forecasts generated by the ARCH model are "smooth" in the underlying state variables, so that as ${}_h\hat{U}_\tau - {}_hU_\tau$ approaches zero, the forecasts generated by the ARCH model approach the forecasts generated by the correct model--i.e., $F_h(A, x, u, \tau) - \hat{F}_h(A, x, \hat{u}, \tau) \rightarrow 0$ as $u \rightarrow \hat{u}$ and $h \downarrow 0$. The three steps, taken together, yield Theorems 2.3-2.4, the main results of this paper. Theorems 2.5 and 2.6 extend these results.

Step One: Convergence as data generating processes

The first set of assumptions assures that $\{{}_hX_t, {}_hU_t\}_{[0, T]}$ and $\{{}_hX_t, {}_h\hat{U}_t\}_{[0, T]}$ converge weakly to limit diffusions $\{X_t, U_t\}_{[0, T]}$ and $\{X_t, \hat{U}_t\}_{[0, T]}$ under P_h and \hat{P}_h respectively as $h \downarrow 0$, where $\{X_t, U_t\}_{[0, T]}$ and $\{X_t, \hat{U}_t\}_{[0, T]}$ are generated by the stochastic integral equations

$$(X_t', U_t')' = (X_0', U_0')' + \int_0^t \mu(X_s, U_s) ds + \int_0^t \Omega^{1/2}(X_s, U_s) dW_s, \text{ and} \quad (12)$$

$$(X_t', \hat{U}_t')' = (X_0', \hat{U}_0')' + \int_0^t \hat{\mu}(X_s, \hat{U}_s) ds + \int_0^t \hat{\Omega}^{1/2}(X_s, \hat{U}_s) dW_s. \quad (12')$$

In (12), $\{W_t\}$ is an $(n+m) \times 1$ standard Brownian motion, $\mu(X_t, U_t)$ is the $(n+m) \times 1$ instantaneous drift per unit of time in $\{X_t, U_t\}$, and $\Omega(X_t, U_t)$ is the $(n+m) \times (n+m)$ instantaneous conditional covariance matrix per unit of time of the increments in $\{X_t, U_t\}$. $[X_0, U_0]$ is taken to be random with a distribution π . In (12'), $\hat{\mu}$, $\hat{\Omega}$, \hat{U}_t , and $\hat{\pi}$ replace μ , Ω , U_t and π . We call the probability measures on $D([0, \infty), R^n \times R^m)$ generated by (12) and (12') P_0 and \hat{P}_0 .

Under certain regularity conditions, four things completely characterize the distribution of $\{X_t, U_t\}$ in (12): the functions μ , and Ω , the distribution π , and the almost sure continuity of the sample path of $\{X_t, U_t\}$. We achieve weak convergence of $\{hX_t, hU_t\}$ to $\{X_t, U_t\}$ matching these four things in the limit as $h \downarrow 0$ by making the drift, conditional variance, and time zero distribution of $\{hX_t, hU_t\}$ converge to μ , Ω , and π respectively, and by making the sizes of the jumps in $\{hX_t, hU_t\}$ converge in probability to zero at an appropriate rate. This holds for (X_t, \hat{U}_t) in (12') as well, making suitable substitutions for μ , Ω , U_t and π . For further discussion and interpretation of these assumptions and relevant references to the probability literature, see Nelson (1990, 1992).

Assumption 1. Under P_h , $({}_hX_0, {}_hU_0, {}_h\hat{U}_0) \Rightarrow (X_0, U_0, \hat{U}_0)$ as $h \downarrow 0$, where (X_0, U_0, \hat{U}_0) has probability measure ν_0 . Under \hat{P}_h , $({}_hX_0, {}_hU_0, {}_h\hat{U}_0) \Rightarrow (X_0, U_0, \hat{U}_0)$ as $h \downarrow 0$, where (X_0, U_0, \hat{U}_0) has probability measure $\hat{\nu}_0$.

Assumption 2. There exists an $\epsilon > 0$ such that for every $R > 0$ and every $k > 0$,⁵

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R, \|u\| \leq R} h^{-1} E_{(h,x,u,kh)} [|{}_h X_{i,(k+1)h} - {}_h X_{i,kh} |^{4+\epsilon}] = 0 \quad (13)$$

for $i = 1$ to n , and

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R, \|u\| \leq R} h^{-1} E_{(h,x,u,kh)} [|{}_h U_{i,(k+1)h} - {}_h U_{i,kh} |^{4+\epsilon}] = 0 \quad (14)$$

a.s. $[P_h]$ for all $i = 1$ to m . Further, (13)-(14) continue to hold when E_h is replaced by \hat{E}_h .

Assumption 1 says that the distributions of the random starting points $({}_h X_0, {}_h U_0, {}_h \hat{U}_0)$ converge to (perhaps distinct) limit measures under P_h and \hat{P}_h as $h \downarrow 0$. Assumption 2 contains conditional moment restrictions guaranteeing that the jump sizes in $\{{}_h X_t, {}_h U_t, {}_h \hat{U}_t\}$ vanish to zero at an appropriate rate as $h \downarrow 0$. This is necessary because the sample paths of the limit diffusion are continuous with probability one.

Next, define the first and second conditional moments

$$\begin{aligned} \mu_h(x,u) &\equiv h^{-1} E_{(h,x,u,kh)} \left[\begin{bmatrix} {}_h X_{(k+1)h} - {}_h X_{kh} \\ {}_h U_{(k+1)h} - {}_h U_{kh} \end{bmatrix} \right], \\ \Omega_h(x,u) &\equiv h^{-1} E_{(h,x,u,kh)} \left[\begin{bmatrix} {}_h X_{(k+1)h} - {}_h X_{kh} \\ {}_h U_{(k+1)h} - {}_h U_{kh} \end{bmatrix} \begin{bmatrix} {}_h X_{(k+1)h} - {}_h X_{kh} \\ {}_h U_{(k+1)h} - {}_h U_{kh} \end{bmatrix}^T \right], \end{aligned} \quad (15)$$

5) More notation: $X_{i,t}$ is the i^{th} element of the vector X_t , and B_{ij} is the i - j^{th} element of the matrix B . " \Rightarrow " denotes weak convergence. $\|B\|$ is a norm of the $q \times r$ real matrix B defined by $\|B\| \equiv [\sum_{i=1,q} \sum_{j=1,r} B_{ij}^2]^{1/2}$. $E_{(h,x,u,kh)}[\cdot]$ is an expectation taken under $P_{(h,x,u,\tau)}$, and $\hat{E}_{(h,x,u,kh)}[\cdot]$ is an expectation taken under $\hat{P}_{(h,x,u,\tau)}$. By (7) and (7'), the expectations in (13)-(14) do not depend on τ , and neither do the moments on the right-hand side of (15)-(15') below.

$$\begin{aligned}\hat{\mu}_h(x,u) &\equiv h^{-1} \hat{E}_{(h,x,\hat{u},kh)} \left[\begin{bmatrix} {}_hX_{(k+1)h} - {}_hX_{kh} \\ {}_h\hat{U}_{(k+1)h} - {}_h\hat{U}_{kh} \end{bmatrix} \right], \\ \hat{\Omega}_h(x,u) &\equiv h^{-1} \hat{E}_{(h,x,\hat{u},kh)} \left[\begin{bmatrix} {}_hX_{(k+1)h} - {}_hX_{kh} \\ {}_h\hat{U}_{(k+1)h} - {}_h\hat{U}_{kh} \end{bmatrix} \begin{bmatrix} {}_hX_{(k+1)h} - {}_hX_{kh} \\ {}_h\hat{U}_{(k+1)h} - {}_h\hat{U}_{kh} \end{bmatrix}^T \right].\end{aligned}\quad (15')$$

Note that (1)-(4) allow us to ignore ${}_h\hat{U}_{kh}$ in the definitions of μ_h and Ω_h and ${}_hU_{kh}$ in the definitions of $\hat{\mu}_h$ and $\hat{\Omega}_h$. μ_h and Ω_h are the conditional drift and second moment matrix of the $\{{}_hX_t, {}_hU_t\}$ processes under P_h . Each is normalized by dividing by h . $\hat{\mu}_h$ and $\hat{\Omega}_h$ are the conditional drift and second moment matrix under \hat{P}_h . Our next assumption requires them to converge, uniformly on compact subsets of the state space, to appropriate limits:

Assumption 3. There exist continuous functions $\mu(x,u)$, $\hat{\mu}(x,u)$, $\Omega(x,u)$, and $\hat{\Omega}(x,u)$ such that for every $R > 0$,

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R, \|u\| \leq R} \| \mu_h(x,u) - \mu(x,u) \| = 0, \quad (16)$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R, \|u\| \leq R} \| \Omega_h(x,u) - \Omega(x,u) \| = 0,$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R, \|u\| \leq R} \| \hat{\mu}_h(x,u) - \hat{\mu}(x,u) \| = 0, \text{ and} \quad (16')$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R, \|u\| \leq R} \| \hat{\Omega}_h(x,u) - \hat{\Omega}(x,u) \| = 0.$$

We also require that $\mu(x,u)$, $\hat{\mu}(x,u)$, $\Omega(x,u)$, $\hat{\Omega}(x,u)$, ν_0 and $\hat{\nu}_0$ completely characterize the distributions of the limit diffusions $\{X_t, U_t\}_{[0,T]}$ and $\{\hat{X}_t, \hat{U}_t\}_{[0,T]}$:

Assumption 4. For any choice of π_0 and $\hat{\pi}_0$, distributionally unique solutions exist to

the stochastic integral equations (12) and (12').⁶

Lemma 2.1 (Stroock-Varadhan (1979)). Under Assumptions 1-4, $\{_h X_t, {}_h U_t\}_{[0, \infty)} \Rightarrow \{X_t, U_t\}_{[0, \infty)}$ under P_h as $h \downarrow 0$, where the initial distribution π is given by

$$\pi(\Gamma) \equiv \nu_0(\Gamma \times R^m) \quad (17)$$

for every $\Gamma \in \mathcal{B}(R^n \times R^m)$. If π and ν_0 in (17) are replaced with $\hat{\pi}$ and $\hat{\nu}_0$, $\{_h X_t, {}_h \hat{U}_t\}_{[0, \infty)} \Rightarrow \{X_t, \hat{U}_t\}_{[0, \infty)}$ under \hat{P}_h as $h \downarrow 0$.

Proof: See Nelson (1990, Theorem 2.2).

(12) and (12') define Markov processes. P_0 and \hat{P}_0 are the measures on $(D([0, \infty), R^n \times R^m), \mathcal{B}(D([0, \infty), R^n \times R^m)))$ generated by (12) and (12'). Consequently $(D([0, \infty), R^n \times R^m), \mathcal{B}(D([0, \infty), R^n \times R^m)), P_0)$ and $(D([0, \infty), R^n \times R^m), \mathcal{B}(D([0, \infty), R^n \times R^m)), \hat{P}_0)$ are probability spaces. (Note: P_h and \hat{P}_h used $D([0, \infty), R^n \times R^{2m})$ not $D([0, \infty), R^n \times R^m)$.) Accordingly, we define the forecast functions associated with P_0 and \hat{P}_0 :

$$F_0(A, x, u, \tau) \equiv P_{(0, x, u, \tau)}[\{X_t, U_t\}_{\tau \leq t < \infty} \in A] \text{ and} \quad (18)$$

$$\hat{F}_0(A, x, \hat{u}, \tau) \equiv \hat{P}_{(0, x, \hat{u}, \tau)}[\{X_t, U_t\}_{\tau \leq t < \infty} \in A], \quad (18')$$

where again $A \in \mathcal{B}(D([\tau, \infty), R^n \times R^m))$, and $P_{(0, x, u, \tau)}$ and $\hat{P}_{(0, x, \hat{u}, \tau)}$ are measures on $(D([\tau, \infty), R^n \times R^m), \mathcal{B}(D([\tau, \infty), R^n \times R^m)))$ constructed in the manner that $P_{(h, x, u, \tau)}$ and $\hat{P}_{(h, x, \hat{u}, \tau)}$ were--i.e., by starting the diffusions (12) and (12') at time τ with $(X_\tau, U_\tau) = (x, u)$ with probability one.

6) See 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 (1990).

Lemma 2.1 gives conditions under which P_h and \hat{P}_h are associated with well-behaved limit diffusions for $\{ {}_hX_t, {}_hU_t \}$ and $\{ {}_hX_t, {}_h\hat{U}_t \}$ as $h \downarrow 0$. Consistent estimation of the forecast distribution will also require that the drifts and conditional covariances of these diffusions are the same--i.e., we require

Assumption 5. For all $(x,u) \in R^{n+m}$, $\hat{\mu}(x,u) = \mu(x,u)$ and $\hat{\Omega}(x,u) = \Omega(x,u)$.

Assumption 5 says that the (misspecified) ARCH model generating \hat{P}_h correctly specifies the functional form of the first two conditional moments of ${}_hX_t$ and ${}_hU_t$. It is the most important additional assumption required to move from consistent filtering to consistent estimation of the forecast distribution. We have kept Assumption 5 separate from the first four assumptions, however, since we are also interested in characterizing the forecasts generated by \hat{P}_h when Assumption 5 is not satisfied (see Theorem 2.4 below).

The conditions of Lemma 2.1 (with the addition of Assumption 5) accomplish step 1--i.e., if ${}_hU_\tau = {}_h\hat{U}_\tau$, and if $\hat{\mu} = \mu$ and $\hat{\Omega} = \Omega$, the forecast distributions generated by P_h and \hat{P}_h at time τ become close (and both become close to the forecast distribution generated by the limit diffusion P_0) as $h \downarrow 0$.

Step Two: Consistent Filtering

Definition: We say that $\{ {}_h\hat{U}_t \}$ (or, equivalently \hat{P}_h) is a consistent filter for $\{ {}_hU_t \}$ at time τ under $\{ P_h \}_{h \downarrow 0}$, if for all $\varepsilon > 0$, $\lim_{h \downarrow 0} P_h[\| {}_h\hat{U}_{\tau-h}U_\tau \| > \varepsilon] = 0$.

In addition to Assumptions 1-4, three (quite technical) regularity conditions are required for consistent filtering. These are reproduced in the appendix as Assumptions 6-8.

Detailed discussion of these conditions can be found in Nelson (1992, section 2).

Lemma 2.2 accomplishes step 2:

Lemma 2.2 (Ethier-Nagylaki (1988)). Let Assumptions 1-4 and 6-8 hold. Then for every

$\tau, 0 < \tau < \infty, \{\hat{U}_t\}$ is a consistent filter for $\{U_t\}$ at time τ under $\{P_h\}_{h \downarrow 0}$.

Proof: See Ethier and Nagylaki (1988) or Nelson (1992, Theorem 2.2).

Step 3: Consistent Estimation of Forecast Distributions

Definition: Let $\{X_t, U_t\}_{[0, \infty)} \Rightarrow \{X_t, U_t\}_{[0, \infty)}$ under P_h as $h \downarrow 0$, where $\{X_t, U_t\}_{[0, \infty)}$ is generated by the diffusion (12). Let ∂A be the boundary of the set A —i.e., the set of all points in $D([\tau, \infty), \mathbb{R}^n \times \mathbb{R}^m)$ which are limit points both of A and of its complement. Let \mathcal{M}_τ be the collection of sets A such that $A \in \mathcal{B}(D([\tau, \infty), \mathbb{R}^n \times \mathbb{R}^m))$ and $P_{(0, x, u, \tau)}[\{X_t, U_t\}_{\tau \leq t < \infty} \in \partial A] = 0$ for all starting points (x, u) . We say that $\{X_t, \hat{U}_t\}_{[\tau, \infty)}$ consistently estimates the forecast distribution of $\{X_t, U_t\}_{[\tau, \infty)}$ at time τ if for every $A \in \mathcal{M}_\tau$ and every $\epsilon > 0$, $P_h[|F_h(A, X_\tau, U_\tau) - \hat{F}_h(A, X_\tau, \hat{U}_\tau)| > \epsilon] \rightarrow 0$ as $h \downarrow 0$.

Theorem 2.3. If Assumptions 1-8 are satisfied, then for every τ with $0 < \tau < \infty$, $\{X_t, \hat{U}_t\}_{[\tau, \infty)}$ consistently estimates the forecast distribution of $\{X_t, U_t\}_{[\tau, \infty)}$ as $h \downarrow 0$. In addition, for every $\epsilon > 0$ and for every $A \in \mathcal{M}_\tau$

$$P_h[|F_h(A, X_\tau, U_\tau) - F_0(A, X_\tau, U_\tau)| > \epsilon] \rightarrow 0 \text{ and} \quad (19)$$

$$P_h[|\hat{F}_h(A, X_\tau, \hat{U}_\tau) - \hat{F}_0(A, X_\tau, \hat{U}_\tau)| > \epsilon] \rightarrow 0 \text{ as } h \downarrow 0. \quad (20)$$

Proof: See Appendix.

Theorem 2.3 gives conditions for the difference between forecast distributions

generated by the sequence of misspecified ARCH models and the forecasts generated by the "correct" model to vanish (in probability) at time τ as $h \downarrow 0$. In addition, both the true and ARCH forecast distributions approach the forecast distribution generated by the limit diffusion P_0 . This holds for forecasts involving any of the state variables, whether observable or unobservable. As indicated earlier, the proof of Theorem 2.3 consists of Lemmas 2.1 and 2.2, plus "smoothness" of the forecast distributions as functions of the unobserved state variables. See the Appendix for details.

Together, the conditions of Theorem 2.3 ensure that the probabilities \hat{P}_h generated by the misspecified ARCH model provide both consistent filtering (for all $t > 0$) and consistent estimation of the forecast distribution for all time intervals $[\tau, \infty)$ for $0 < \tau < \infty$. In addition, they insure that the time τ conditional probabilities generated by \hat{P}_h and P_h are well-approximated by the time τ conditional probability generated by the limit diffusion P_0 .

What can we say about the forecasts generated by misspecified ARCH models when the conditions of Theorem 2.3 fail? Assumptions 1-4 and 6-8 allow a weaker characterization of the forecasts produced by the ARCH model:

Theorem 2.4. Let the conditions of Theorem 3 hold, except for Assumption 5. Then (19)-(20) still hold.

Proof: See Appendix.

Theorem 2.4 is easily summarized: when we drop Assumption 5, the ARCH model provides a *consistent filter* but uses the *wrong* limit diffusion (i.e., (12') instead of (12)) to

form the forecast.

Forecast Moments and Stationary Distributions

While Theorems 2.3 and 2.4 give convergence results for conditional distributions, they do not guarantee convergence of either forecast moments or of forecast stationary distributions. Theorem 2.5 considers convergence in probability of forecast moments of well-behaved functions of $\{h_t X_t, h_t U_t\}_{[\tau, \infty)}$:

Theorem 2.5. Let $0 < \tau < \infty$, and let Assumptions 1-4 and 6-8 be satisfied. Let g be a functional mapping $D([\tau, \infty), R^n \times R^m)$ into R^1 . Let $g^d \subset D([\tau, \infty), R^n \times R^m)$ be the set of discontinuity points in g . For every $(x, u) \in R^{n+m}$, let

$$P_{(0,x,u,\tau)}[\{X_t, U_t\}_{[\tau, \infty)} \in g^d] = 0.$$

In addition, let g satisfy the following uniform integrability condition: for every bounded $\Lambda \subset R^{n+m}$,

$$\lim_{K \rightarrow \infty} \limsup_{h \downarrow 0} \sup_{(x,u) \in \Lambda} \hat{E}_{(h,x,u,\tau)}[|g(\{h_t X_t, h_t U_t\}_{[\tau, \infty)})| \cdot I(|g(\{h_t X_t, h_t U_t\}_{[\tau, \infty)})| > K)] = 0 \quad (21)$$

and

$$\lim_{K \rightarrow \infty} \limsup_{h \downarrow 0} \sup_{(x,u) \in \Lambda} E_{(h,x,u,\tau)}[|g(\{h_t X_t, h_t U_t\}_{[\tau, \infty)})| \cdot I(|g(\{h_t X_t, h_t U_t\}_{[\tau, \infty)})| > K)] = 0, \quad (22)$$

where $I(\cdot)$ is an indicator function (i.e., $I(x \in H) = 1$ if $x \in H$ and $I(x \in H) = 0$ otherwise),

and $E_{(h,x,u,\tau)}[\cdot]$ and $\hat{E}_{(h,x,u,\tau)}[\cdot]$ are expectations taken under $P_{(h,x,u,\tau)}$ and $\hat{P}_{(h,x,u,\tau)}[\cdot]$. Define the forecast functions

$$G_h(g, x, u, \tau) \equiv E_{(h,x,u,\tau)}[g(\{h_t X_t, h_t U_t\}_{[\tau, \infty)})], \quad (23)$$

$$G_0(g, x, u, \tau) \equiv E_{(0,x,u,\tau)}[g(\{X_t, U_t\}_{[\tau, \infty)})], \quad (24)$$

$$\hat{G}_h(g, x, u, \tau) \equiv \hat{E}_{(h, x, u, \tau)}[g(\{X_t, \hat{U}_t\}_{\tau, \infty})], \text{ and} \quad (25)$$

$$\hat{G}_0(g, x, u, \tau) \equiv \hat{E}_{(0, x, u, \tau)}[g(\{X_t, \hat{U}_t\}_{\tau, \infty})]. \quad (26)$$

Then for every $\tau > 0$ and every $\epsilon > 0$,

$$P_h[|\hat{G}_h(g, X_\tau, \hat{U}_\tau, \tau) - \hat{G}_0(g, X_\tau, U_\tau, \tau)| > \epsilon] \rightarrow 0, \text{ and} \quad (27)$$

$$P_h[|G_h(g, X_\tau, U_\tau, \tau) - G_0(g, X_\tau, U_\tau, \tau)| > \epsilon] \rightarrow 0 \quad (28)$$

as $h \downarrow 0$. If Assumption 5 is also satisfied, then for every $\tau > 0$ and every $\epsilon > 0$,

$$P_h[|\hat{G}_h(g, X_\tau, \hat{U}_\tau, \tau) - G_h(g, X_\tau, U_\tau, \tau)| > \epsilon] \rightarrow 0 \text{ as } h \downarrow 0. \quad (29)$$

A sufficient condition for (21)-(22) is that there exists an $\epsilon > 0$ such that for every bounded $\Lambda \subset R^{n+m}$,

$$\limsup_{h \downarrow 0} \sup_{(x, u) \in \Lambda} \hat{E}_{(h, x, u, \tau)}[|g(\{X_t, \hat{U}_t\}_{\tau, \infty})|^{1+\epsilon}] < \infty, \text{ and} \quad (21')$$

$$\limsup_{h \downarrow 0} \sup_{(x, u) \in \Lambda} E_{(h, x, u, \tau)}[|g(\{X_t, U_t\}_{\tau, \infty})|^{1+\epsilon}] < \infty. \quad (22')$$

Proof: See Appendix.

The uniform integrability conditions described in Theorem 2.5 are similar to standard conditions allowing integration to the limit (see, e.g., Billingsley (1986, pp. 347-348)).

The next result, due to Kushner (1984), gives conditions for convergence of forecast stationary distributions. It also gives conditions for moment boundedness, which sometimes can be used to verify (21')-(22').

Theorem 2.6 (Kushner). Let Assumptions 1-4 be satisfied, and let the diffusion (12) admit a unique invariant measure for (X_t, U_t) --i.e., there exists a random vector (X_∞, U_∞) on

R^{n+m} such that for any probability measure $\pi(\cdot)$ for (X_0, U_0) in (12), $(X_t, U_t) \Rightarrow (X_\infty, U_\infty)$ as $t \rightarrow \infty$ under P_0 . For each sufficiently small $h > 0$, let P_h also admit a unique invariant measure--i.e., for any probability measure $\nu_h(\cdot)$ for $({}_hX_0, {}_hU_0)$ in (5), $({}_hX_t, {}_hU_t) \Rightarrow ({}_hX_\infty, {}_hU_\infty)$ as $t \rightarrow \infty$ under P_h . Let there be a twice-continuously differentiable Liapunov function $\zeta(x, u)$ and positive numbers h^* , Δ , and η such that for all h , $0 < h < h^*$,

$$\liminf_{\|x' u'\| \rightarrow \infty} \zeta(x, u) = \infty, \quad (30)$$

$$\min_{[x' u'] \in R^{n+m}} \zeta(x, u) = 0, \quad (31)$$

$$\sup_{0 < h \leq h^*} E_h[\zeta({}_hX_0, {}_hU_0)] < \infty, \quad (32)$$

and for all $(x, u) \in R^{n+m}$ and all h , $0 < h < h^*$

$$h^{-1} E_{(h, x, u, \tau)}[\zeta({}_hX_{(k+1)h}, {}_hU_{(k+1)h}) - \zeta(x, u)] < \Delta - \eta \cdot \zeta(x, u). \quad (33)$$

Then $({}_hX_t, {}_hU_t) \Rightarrow (X_\infty, U_\infty)$ under P_h as $t \rightarrow \infty$ and $h \downarrow 0$, and there exists a constant K and a sequence $\eta_h \rightarrow \eta$ as $h \downarrow 0$ such that for all $t \geq 0$,

$$E_h[\zeta({}_hX_t, {}_hU_t)] \leq \eta^{-1} K[\Delta + \exp(-\eta_h \cdot t) \cdot E_h[\zeta({}_hX_0, {}_hU_0)]]. \quad (34)$$

If a second Liapunov function $\hat{\zeta}(x, u)$ exists such that (30)-(34) hold when E_h , P_h , P_0 , $\zeta(x, u)$, $\nu_h(\cdot)$, ${}_hU_t$, and (12) are replaced by \hat{E}_h , \hat{P}_h , \hat{P}_0 , $\hat{\zeta}(x, u)$, $\hat{\nu}_h(\cdot)$, ${}_h\hat{U}_t$, and (12') respectively, then $({}_hX_t, {}_h\hat{U}_t) \Rightarrow (X_\infty, \hat{U}_\infty)$ under \hat{P}_h as $t \rightarrow \infty$ and $h \downarrow 0$, and there exists a constant \hat{K} and a sequence $\hat{\eta}_h \rightarrow \eta$ as $h \downarrow 0$ such that for all $t \geq 0$

$$\hat{E}_h[\hat{\zeta}({}_hX_t, {}_h\hat{U}_t)] \leq \eta^{-1} \hat{K}[\Delta + \exp(-\hat{\eta}_h \cdot t) \cdot \hat{E}_h[\hat{\zeta}({}_hX_0, {}_h\hat{U}_0)]]. \quad (35)$$

Proof: See Appendix.

In some applications, $\{{}_hX_t\}$ may be nonstationary even when $\{{}_hU_t\}$ is stationary. For

example, $\{ {}_hX_t \}$ may be the (nonstationary) cumulative return on a portfolio, while $\{ {}_hU_t \}$ may be the (stationary) instantaneous conditional variance of this return. Theorem 2.6 can be adapted to this case: If there is no feedback from $\{ {}_hX_t \}$ to $\{ {}_hU_t \}$ so that $\{ {}_hU_t \}$ is a stationary Markov process under P_h in its own right (i.e., without reference to ${}_hX_t$), replace " $({}_hX_t, {}_h\hat{U}_t)$ " and " $({}_hX_t, {}_hU_t)$ " with " $({}_h\hat{U}_t)$ " and " $({}_hU_t)$ " in the statement of Theorem 2.6, and it continues to hold.

The conditions of Theorem 2.6 are difficult to verify unless the innovations in $\{ {}_hX_t, {}_h\hat{U}_t \}$ are bounded either above or below: fortunately, in ARCH models they often are. For an ARCH example in which the conditions of Theorem 2.6 are verified (specifically, for GARCH(1,1), in which the innovations in the conditional variance process are bounded below), see the proof of Theorem 2.3 in Nelson (1990). Sometimes, as in the example in the next section, it is possible to verify convergence of steady state distributions by other means.

3. A stochastic volatility model

On first inspection, the assumptions underlying the results in Section 2 are quite forbidding. In practice, however, they can often be verified. In this section we provide an example, based on a stochastic volatility model employed in the options pricing literature. Variations of this model have been investigated by Wiggins (1987), Scott (1987), Hull and White (1987) and Melino and Turnbull (1990). We propose an ARCH discretization and show that it satisfies Lemmas 2.1 and 2.2 and Theorems 2.3, 2.5, and 2.6.

The Model:

Let S_t be the time t price of a non-dividend paying stock. σ_t is its instantaneous returns volatility. We assume that

$$d[\ln(S_t)] = (\mu - \sigma_t^2/2)dt + \sigma_t dW_{1,t}, \quad (36)$$

$$d[\ln(\sigma_t^2)] = -\beta[\ln(\sigma_t^2) - \alpha]dt + \Lambda \cdot dW_{2,t}. \quad (37)$$

$W_{1,t}$ and $W_{2,t}$ are standard Brownian motions with correlation ρ , i.e.,

$$\begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} \begin{bmatrix} dW_{1,t} & dW_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} dt. \quad (38)$$

μ , Λ , β , and α are constants with $\beta \geq 0$. If σ_t were constant, S_t would follow a geometric Brownian motion. (36)-(38) generalize this much-used model by allowing σ_t to vary randomly, with $\ln(\sigma_t^2)$ following an Ornstein-Uhlenbeck process.

We assume that we observe S_t at discrete intervals of length h , so for every t , ${}_hS_t \equiv S_{h\lfloor t/h \rfloor}$. In the notation of Section 2, ${}_hX_t \equiv \ln({}_hS_t)$, and ${}_hU_t \equiv \ln({}_h\sigma_t^2)$. We take $S_0 > 0$ and $\sigma_0^2 > 0$ to be nonrandom. We define ${}_h\hat{U}_t$ by (36')-(38') below and thus obtain the process $\{{}_hX_t, {}_hU_t, {}_h\hat{U}_t\}$.

We consider a discretization based on the AR(1) EGARCH model of Nelson (1991): specifically, we form our forecasts and our estimated conditional variances ${}_h\hat{\sigma}_{kh}^2$ assuming that the data are generated by the model

$$\ln({}_hS_{(k+1)h}) = \ln({}_hS_{kh}) + (\mu - {}_h\sigma_{kh}^2/2) \cdot h + h^{1/2} {}_h\sigma_{kh} \cdot {}_hZ_{kh}, \quad (36')$$

$$\ln({}_h\hat{\sigma}_{(k+1)h}^2) = \ln({}_h\hat{\sigma}_{kh}^2) - \beta \cdot h \cdot [\ln({}_h\hat{\sigma}_{kh}^2) - \alpha] + h^{1/2} \cdot g({}_hZ_{kh}), \quad (37')$$

where for all $h > 0$, ${}_h\hat{\sigma}_0^2 > 0$ is fixed, ${}_hZ_{kh} \sim$ i.i.d. $N(0,1)$, and

$$g(z) \equiv \theta \cdot z + \gamma [|z| - (2/\pi)^{1/2}]. \quad (38')$$

Though the residuals $\{ {}_hZ_{kh} \}$ in (36')-(38') are not directly observable, they can be obtained by rearranging (36'):

$${}_hZ_{kh} = h^{-1/2} \{ \ln({}_hS_{kh}) - \ln({}_hS_{(k-1)h}) - h \cdot \mu + h \cdot {}_h\hat{\sigma}_{kh}^2/2 \} / {}_h\hat{\sigma}_{kh}. \quad (39)$$

Given $\hat{\sigma}_0^2$ and $\{ {}_hS_{kh} \}_{k=0,K}$, (36')-(38') and (39) allow us to recursively compute $\{ {}_hZ_{kh} \}_{k=0,K}$ and $\{ {}_h\hat{\sigma}_k^2 \}_{k=0,K}$. Note that although the ARCH model (\hat{P}_h) assumes that $\hat{\sigma}_0^2 = \sigma_0^2$, this needn't be true under P_h . (This is why we required τ to be strictly positive in Section 2--clearly consistent filtering is impossible at time 0 if $\hat{\sigma}_0^2 \neq \sigma_0^2$!)

As in Section 2, we create the continuous time processes $\{ {}_hS_t, {}_h\hat{\sigma}_t^2 \}$ by interpolating the discrete processes $\{ {}_hS_{kh}, {}_h\hat{\sigma}_{kh}^2 \}$. While (36)-(38) (and (36')-(37')) and the $g(\cdot)$ of (38') for ${}_h\hat{U}_t$) generate the true probabilities P_h , the misspecified probabilities \hat{P}_h are generated by (36')-(38') and the requirement that ${}_h\sigma_t^2 = {}_h\hat{\sigma}_t^2$ for all t with probability one.

As indicated in (38'), the EGARCH model generating \hat{P}_h assumes that $\{ {}_hZ_{kh} \}$ is i.i.d. $N(0,1)$. When (36)-(38) generate the data and ${}_hZ_{kh}$ is recursively defined by (39), this will not be true. As we will see, however, this ARCH model, though misspecified, is able to provide consistent filtering and consistent estimation of forecast distributions for the true model (36)-(38) as $h \downarrow 0$. As shown in Nelson (1992), the main requirement for consistent filtering for this model is $\gamma > 0$. For consistent estimation of forecast distributions, however, we must match the first two conditional moments of the ARCH model considered as a data generating process to the corresponding moments of the true data generating process (36)-(38).

Under \hat{P}_h , $h^{1/2}g({}_hZ_{kh})$ is the innovation in $\ln({}_h\hat{\sigma}_{(k+1)h}^2)$, and has variance $h[\theta^2 + \gamma^2(1 - 2/\pi)]$. The instantaneous correlation of the increments in $\ln({}_hS_{kh})$ and

$\ln(\sigma_{kh}^2)$ under \hat{P}_h is $\theta/[\theta^2 + \gamma^2(1 - 2/\pi)]^{1/2}$. Matching these conditional second moments (under \hat{P}_h and P_h) requires that

$$\text{Var}[g(hZ_{kh})] = \theta^2 + \gamma^2(1 - 2/\pi) = \Lambda^2, \text{ and} \quad (40)$$

$$\rho = \theta/[\theta^2 + \gamma^2(1 - 2/\pi)]^{1/2}, \quad (41)$$

which is easily accomplished by setting $\theta = \rho \cdot \Lambda$ and $\gamma = |\Lambda|(1 - \rho^2)^{1/2}/(1 - 2/\pi)^{1/2}$. The drifts of $\ln(S_t)$ and $\ln(\sigma_t^2)$ in (36)-(37) are $(\mu - \sigma^2/2)$ and $-\beta[\ln(\sigma_t^2) - \alpha]$ respectively. (36') implies that the drift of $\ln(\sigma_{kh})$ is also $(\mu - \sigma^2/2)$, and sets the drift in $\ln(\hat{\sigma}_{kh}^2)$ equal to $-\beta[\ln(\hat{\sigma}_{kh}^2) - \alpha]$.

Theorem 3.1. For each $h > 0$, let $\{_h S_t, {}_h \sigma_t^2\}$ be generated by (36)-(38) observed at discrete intervals of length h . Let $\{_h \hat{\sigma}_t^2\}$ be generated by the EGARCH model (36')-(38') and (40)-(41). Then

(a) *For every τ , $0 < \tau < \infty$, $\{_h \hat{\sigma}_\tau^2\}$ is a consistent filter for $\{_h \sigma_\tau^2\}$.*

(b) *For every τ , $0 < \tau < \infty$, $\{_h S_t, {}_h \hat{\sigma}_t^2\}_{t \in [\tau, \infty)}$ consistently estimates the forecast distributions of $\{_h S_t, {}_h \sigma_t^2\}_{t \in [\tau, \infty)}$. The other conclusions of Theorem 2.3 also hold.*

(c) *Let $G(s_1, s_2, \sigma_1, \sigma_2)$ be a continuous function from R^4 into R^1 satisfying*

$$|G(s_1, s_2, \sigma_1, \sigma_2)| < A + B \cdot |s_1|^a \cdot |s_2|^b \cdot |\sigma_1|^c \cdot |\sigma_2|^d \quad (42)$$

for finite, non-negative A, B, a, b, c , and d . Then the conditions of Theorem 2.5 are satisfied for $g(\{_h S_t, {}_h \sigma_t^2\}_{t \in [\tau, \infty)}) \equiv G((\ln({}_h S_t), \ln({}_h S_T), {}_h \sigma_t, {}_h \sigma_T))$ for every τ, t , and T , with $0 < \tau < t < T < \infty$.

(d) *If $\beta > 0$, the stationary distributions of σ_t^2 and ${}_h \hat{\sigma}_t^2$ for each $h > 0$ exist--i.e., $\sigma_t^2 \Rightarrow \sigma_\infty^2$ as $t \rightarrow \infty$, and for each $h > 0$, ${}_h \hat{\sigma}_t^2 \Rightarrow {}_h \hat{\sigma}_\infty^2$ as $t \rightarrow \infty$. Further, ${}_h \hat{\sigma}_\infty^2 \Rightarrow \sigma_\infty^2$ as $h \downarrow 0$.*

Proof: See Appendix.

(a) and (b) are applications of Lemma 2.2 and Theorem 2.3 respectively. (c) is an application of Theorem 2.5 for a class of $G(\cdot)$ functions encompassing moments and cross moments of forecast stock returns (continuously compounded) and conditional variances. (d) yields the convergence in distribution of ${}_h\sigma_\infty^2$ to σ_∞^2 . (${}_hS_\infty$ doesn't converge, since $\{S_t\}$ is nonstationary).

Non-uniqueness of consistent estimates of forecast distributions

It is important to note that there are many ARCH discretizations of the model (36)-(38) which achieve consistent estimation of the forecast distribution. For example, Nelson and Foster (1992) replace the $g(\cdot)$ of (38') with

$$g(z) \equiv \Lambda^{1/2}[\rho z + [(1-\rho^2)/2]^{1/2} \cdot [z^2 - 1]]. \quad (43)$$

Though (43) is not a model familiar in the ARCH literature, Nelson and Foster show that it has slightly better filtering properties than (38') when (36)-(38) generate the data (i.e., $({}_h\hat{\sigma}_t^2 - {}_h\sigma_t^2)$ approaches zero at a slightly faster rate as $h \downarrow 0$). Theorem 3.1 continues to hold for (43). We leave the verification to the reader.

Consistent filtering without consistent estimation of the forecast distribution

It is important to emphasize that the conditions in Theorem 3.1 are much stricter than would be required for consistent filtering alone. For example the second moment matching conditions (40)-(41) are not needed for consistent filtering. For example, suppose we select (θ, γ) with $\gamma > 0$ but not satisfying (40)-(41). In this case, the conditions of lemmas 2.1-2.2 and theorem 2.4 are satisfied, but the conditions of theorem 2.3 are not. That is, the

unobservable state U_τ is recovered as $h \downarrow 0$, but the time τ forecasts generated by \hat{P}_h approach the forecasts generated by the diffusion

$$d[\ln(S_t)] = (\mu - \sigma_t^2/2)dt + \sigma_t dW_{1,t}, \quad (36'')$$

$$d[\ln(\sigma_t^2)] = -\beta[\ln(\sigma_t^2) - \alpha]dt + \Lambda^* \cdot dW_{2,t}, \quad (37'')$$

$$\begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} \begin{bmatrix} dW_{1,t} & dW_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & \rho^* \\ \rho^* & 1 \end{bmatrix} dt, \quad (38'')$$

where

$$\Lambda^* \equiv [\theta^2 + \gamma^2(1 - 2/\pi)]^{1/2}, \text{ and} \quad (40')$$

$$\rho^* \equiv \theta/[\theta^2 + \gamma^2(1 - 2/\pi)]^{1/2}. \quad (41')$$

We could also allow the β 's and α 's in (37) and (37') to differ. Again, this would not affect consistent filtering or the results of lemmas 2.1 and 2.2, but it would prevent consistent estimation of the forecast distribution. Alternatively, suppose GARCH(1,1) was used for filtering and forecasting (36)-(38). Under weak conditions on the GARCH coefficients, GARCH achieves consistent filtering (see Nelson (1992)), but the time τ forecasts generated by \hat{P}_h would approach the forecasts generated by the diffusion limit of GARCH(1,1) (given in Nelson (1990, (2.39)-(2.40))) rather than the forecasts generated by the true data generating process (36)-(36).

ARCH models as discretizations of diffusions

In empirical applications of stochastic volatility models such as (36)-(38) in options pricing, researchers have often adopted the following rather cumbersome procedure:⁷ first,

7) See, e.g., Scott (1987), Wiggins (1987), Nelson (1988, Chapter 1), Melino and Turnbull (1990), and Ruiz (1992).

approximate (36)-(38) with a discrete time stochastic volatility model such as

$$\ln({}_hS_{(k+1)h}) = \ln({}_hS_{kh}) + (\mu - {}_h\sigma_{kh}^2/2) \cdot h + h^{1/2} \cdot {}_h\sigma_{kh} \cdot {}_hZ_{1,kh} \quad (36''')$$

$$\ln({}_h\sigma_{(k+1)h}^2) = \ln({}_h\sigma_{kh}^2) - \beta \cdot h \cdot [\ln({}_h\sigma_{kh}^2) - \alpha] + \Lambda \cdot h^{1/2} \cdot {}_hZ_{2,kh} \quad (37''')$$

where $\{{}_hZ_{1,kh}, {}_hZ_{2,kh}\}$ are i.i.d. bivariate normal with

$$E[{}_hZ_{1,kh}, {}_hZ_{2,kh}] = [0, 0] \text{ and } \text{Cov}[{}_hZ_{1,kh}, {}_hZ_{2,kh}] = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \quad (38''')$$

(36''')-(38''') is the standard Euler approximation to (36)-(38) (see, e.g., Pardoux and Talay (1985)). Next, proceed with estimation as if this discretization were the true data generating process. Although it is easy to show using Lemma 2.1 that (36''')-(38''') converges weakly (i.e., as a data generating process) to (36)-(38), $\{{}_h\sigma_t^2\}$ is unobservable, so maximum likelihood estimation is intractable. Therefore, employ an alternative procedure such as the generalized method of moments (GMM) to estimate $\mu, \beta, \alpha, \Lambda$, and ρ . With the estimated parameters and some nonlinear filtering procedure (e.g., the extended Kalman filter), estimate the unobservable volatility process $\{{}_h\sigma_t^2\}$. Finally, price options using these estimated conditional variances ${}_h\hat{\sigma}_t^2$, the observed values of ${}_hS_t$, the estimated parameters, and a continuous-time options pricing model.

There are at least three potential problems with this procedure: first, the model that is being estimated is not the true data generating process, but a weakly converging approximation. Second, the regularity conditions yielding consistency and asymptotic normality of the GMM parameter estimates have not been successfully verified for this model, so considerable caution is required in using these parameter estimates in options pricing. There may also be considerable loss of efficiency in going from maximum likelihood to GMM estimation. Third, in the discretization (36''')-(38'''), $\{{}_h\sigma_t^2\}$ is

unobservable, so potentially cumbersome procedures such as extended Kalman filters must be employed to estimate conditional variances.⁸

The EGARCH discretization (36')-(38') and (40)-(41) eliminates the third problem, but not the first two: a researcher using either of the ARCH discretizations proposed in this section would estimate the coefficients by quasi-maximum likelihood, acting as if the (weakly converging) ARCH discretization were the true data generating process. The regularity conditions justifying this procedure have not been verified either, so the first two problems remain. The third problem vanishes, however, since under the ARCH discretization, $\{h_t \hat{\sigma}_t^2\}$ is observable,⁹ rendering the filtering step trivial. That is, one virtue of ARCH discretizations is that they are "self-filtering."¹⁰ For application of this ARCH

8) $\{h_t \hat{\sigma}_t^2\}$ in the extended Kalman filter is produced by a Markov updating rule just as the ARCH models of Section 2 are (see e.g., Anderson and Moore (1979)). In fact, the extended Kalman filter can itself be regarded as an ARCH model, except that an ARCH model would typically call the resulting $\{h_t \hat{\sigma}_t^2\}$ the "true" conditional variance process rather than explicitly acknowledging it as a noisy estimate. (This is probably a virtue of the Kalman filter.) An extended Kalman filter may or may not provide the most convenient, efficient, or tractable ARCH model in a given application. The extended Kalman filter can also be used for smoothing, but so can ARCH models (of course, this requires that the ARCH model is explicitly regarded as a filter). See Foster and Nelson (1991) for a two-sided version of GARCH.

9) That is, it is observable given the initial value $h_0 \sigma_0^2$. The effect of an incorrect initial guess for $h_0 \sigma_0^2$ vanishes quickly, however--see Nelson (1992).

10) At first glance, it is surprising that (36')-(38') can converge weakly to (36)-(38), since there are two noise terms-- $dW_{1,t}$ and $dW_{2,t}$ --driving (36)-(38) and two noise terms-- $h_t Z_{1,kh}$ and $h_t Z_{2,kh}$ --driving the Euler approximation (36'')-(38''), but only *one* noise term ($h_t Z_{kh}$) driving (36')-(38'). To grasp the intuition behind this, consider the following illustration: let η_k be iid $N(0,1)$, so $E|\eta_k| = (2/\pi)^{1/2}$. Although $[|\eta_k| - (2/\pi)^{1/2}]$ and η_k are certainly not *independent*, they are *uncorrelated*. By the central limit theorem therefore, $k^{-1/2} \sum_{j=1,k} \eta_j$ and $k^{-1/2} \sum_{j=1,k} [|\eta_j| - (2/\pi)^{1/2}]$ converge to *independent* normal random variables. What happens in the continuous time limit of (36')-(38') is similar: $h_t Z_{kh}$ and $g(h_t Z_{kh})$ are perfectly dependent but imperfectly correlated, so their partial sums converge weakly to the imperfectly correlated Brownian motions $W_{1,t}$ and

approach to options pricing see, e.g., Kuwahara and Marsh (1992) and Cao (1992).

A Note on Multivariate Models

As we have seen, approximately correct forecasts may be generated by misspecified ARCH models provided that (1) the first two conditional moments of all state variables, observable and unobservable, are correctly specified and (2) the data generating process is well approximated by a diffusion. (These are the most important conditions, though some other regularity conditions must also be satisfied.) Unfortunately, correctly specifying the first two conditional moments is highly nontrivial, especially in the multivariate case.

Let ${}_hX_t$ be $n \times 1$. There are $(n+1) \cdot n/2$ distinct elements of the conditional covariance matrix of ${}_hX_t$, so suppose ${}_hU_t$ is $m \times 1$, where $m \equiv (n+1) \cdot n/2$. Even when n is large, consistent filtering of ${}_hU_t$ can be achieved by a wide variety of (very simple) ARCH models. For example, let ${}_hU_t \equiv \text{vech}({}_h\Omega_{1:n,1:n,t})$, where "vech" is the operator which stacks the upper triangle of a symmetric matrix into a vector and ${}_h\Omega_{1:n,1:n,t}$ is the time t conditional covariance matrix of ${}_hX_t$ under P_h . For some $\alpha > 0$, define ${}_h\hat{U}_t$ by

$${}_h\hat{U}_{(k+1)h} = {}_h\hat{U}_{kh}(1 - h^{1/2}\alpha) + h^{-1/2}\alpha \cdot \text{vech}[({}_hX_{(k+1)h} - {}_hX_{kh})({}_hX_{(k+1)h} - {}_hX_{kh})']. \quad (44)$$

This ARCH model has only one parameter (which can be arbitrarily fixed) yet it provides a consistent filter under quite mild regularity conditions (see Foster and Nelson (1991) and Nelson (1992, pp. 71-73)).

Consistent estimation of the forecast distribution, on the other hand, requires in addition correct specification of the conditional mean of $\{{}_hX_t\}$ (n moments), the

$W_{2,t}$. For more detailed discussion see Nelson (1990, pp. 20-23).

conditional mean of $\{h_t U_t\}$ (m moments), the conditional covariance matrix of $\{h_t U_t\}$ ($[(m+1) \cdot m/2]$ moments), and the conditional covariances of $\{h_t U_t\}$ and $\{h_t X_t\}$, a grand total of $n \cdot (n^3 + 6n^2 + 11n + 14)/8$ moments. When $n=1$, as in the examples of this section, 4 moments must be matched. When $n=2$, 17 moments must be matched, progressing to 215 moments for $n=5$, 2155 moments for $n=10$, and more than thirteen million moments for $n=100$. By way of comparison, the Center for Research in Security Prices (CRSP) at the University of Chicago recorded daily returns during 1991 for 16,854 stocks, requiring the matching of more than 10^{16} moments. Clearly, consistent estimation of the forecast distribution is problematic unless either n is very small or a drastically simplified structure exists for the first two conditional moments, making m much less than $(n+1) \cdot n/2$.

For example, let $h_t X_t$ represent cumulative returns on n assets, and let $h_t \Omega_{1:n,1:n,t}$ be its conditional covariance matrix. In the Factor ARCH model of Engle, Ng, and Rothschild (1990), we have

$$h_t \Omega_{1:n,1:n,t} = \Omega + \sum_{i=1}^m \beta_i \beta_i' h_t \lambda_t^i \quad (45)$$

where $m < n$, Ω is a (constant) $n \times n$ nonnegative definite symmetric matrix, and for $i=1$ to m , β_i is an $n \times 1$ vector of constants and $h_t \lambda_t^i$ is the conditional variance of a portfolio formed from the n assets in $h_t X_t$. Here $h_t U_t \equiv [h_t \lambda_t^1, h_t \lambda_t^2, \dots, h_t \lambda_t^m]'$ is $m \times 1$, with m less than n and *much* less than $n \cdot (n+1)/2$.

Although it is straightforward to find a consistent filter for the portfolio variances $h_t \lambda_t^i$, (45) implies that for $s \neq t$, $h_t \Omega_{1:n,1:n,s} \sim h_t \Omega_{1:n,1:n,t}$ has a rank of at most $m < n$. If the model is misspecified, there is no reason to expect this to hold, so $h_t \hat{\Omega}_{1:n,1:n,t} \equiv \Omega + \sum_{i=1,m} \beta_i \beta_i' \hat{\lambda}_t^i$

need not approach ${}_h\Omega_{1:m,1:m,t}$ as $h \downarrow 0$. This illustrates the danger in assuming $m < n \cdot (n+1)/2$: not only can consistent estimation of the forecast distribution break down if the model is misspecified, even consistent filtering can be lost. It would be useful to create classes of multivariate ARCH models which retain the parametric tightness and intuitive appeal of (45) while retaining the filtering robustness of (44).

4. Conclusion

Researchers using ARCH models have focused their energies on modelling the first two conditional moments of time series. If the true data generating processes are near-diffusions, this emphasis is appropriate, since the first two conditional moments largely determine the behavior of the process. Near-diffusions and ARCH models therefore seem natural companions.¹¹ This paper has shown that corresponding to many diffusion and near-diffusion models, there is a sequence of ARCH models that produce forecasts that are "close" to the forecasts generated by the true model. An ARCH model can produce reasonable forecasts in many cases, even when misspecified, *if the ARCH model correctly specifies the first two conditional moments, including the first two conditional moments of the $\{{}_hX_t, {}_hU_t\}$ process and if the data are generated by a near-diffusion.* This may be part of the reason for the broad empirical success of ARCH.

An important limitation of our results is our assumption that the sequence of ARCH

11) Of course, this is untrue if the near-diffusion assumption is invalid. See the discussion in Nelson (1992, Section 4). Even if the near-diffusion assumption is valid, the entire instantaneous conditional distribution (i.e., not just the first two conditional moments) is important for the *efficiency* of an ARCH model in estimating conditional variances and covariances--see Nelson and Foster (1992).

models is selected a priori (and hence nonrandomly) rather than by an estimation procedure. Establishing consistency and asymptotic normality of maximum likelihood parameter estimates has proven notoriously difficult even for well-specified ARCH models,¹² and is not likely to be any easier for misspecified ARCH models. Monte Carlo evidence reported in Schwartz, Nelson, and Foster (1992) suggests that in large samples the filtering properties of misspecified ARCH models are affected little by parameter estimation. To our knowledge, no Monte Carlo studies of the forecasting properties of misspecified ARCH models with estimated parameters have been done.

Another important limitation of our results is that we have been concerned only with consistency, not with efficiency--i.e., we have had nothing to say about the *rates* at which the convergence in forecast distributions and moments takes place, or about the relative efficiency of different ARCH models in approximating forecast distributions. By analogy with the results of Section 2, this question appears to have two parts: first, at what rate does ${}_h\hat{U}_{t-h}U_t$ converge to zero? Second, at what rate does the weak convergence of $\{{}_hX_t, {}_hU_t\}$ to $\{X_t, U_t\}$ take place? Nelson and Foster (1992) consider the first question for the scalar case. They find a close connection between the efficiency of an ARCH model as a filter and its forecasting properties: apart from a few pathological cases, "optimal" filters also consistently estimate forecast distributions. Unfortunately, no results appear available on the second question except in very special cases, so results on the rate of convergence of forecast distributions must be left for future work.

12) Though there has been some recent progress in the GARCH(1,1) case--see Lumsdaine (1991).

Appendix

Additional Assumptions Required for Lemma 2.2

To keep the notation in line with Nelson (1992), it is useful to define the measurement error process $\{ {}_h Y_t \}$: for all $h > 0$ and all $t > 0$, define ${}_h Y_t \equiv {}_h \hat{U}_t - {}_h U_t$. The following assumptions are adapted from assumptions 5-7 in Nelson (1992):

Assumption 6. For every $h > 0$ and $\delta > 0$ and every $(x, u, \hat{u}) \in R^{n+2m}$, the following are well-defined and finite:

$$c_{h,\delta}(x, u, \hat{u}) \equiv h^{-\delta} E_h [({}_h Y_{(k+1)h} - {}_h Y_{kh}) | {}_h X_{kh} = x, {}_h U_{kh} = u, {}_h \hat{U}_{kh} = \hat{u}], \text{ and}$$

$$d_{h,\delta}(x, u, \hat{u}) \equiv h^{-\delta} E_h [({}_h Y_{(k+1)h} - {}_h Y_{kh})({}_h Y_{(k+1)h} - {}_h Y_{kh})' | {}_h X_{kh} = x, {}_h U_{kh} = u, {}_h \hat{U}_{kh} = \hat{u}].$$

Further, there exists a function $c(x, u, \hat{u})$ with $c(x, u, \hat{u}) = 0$ whenever $u = \hat{u}$, such that for some δ , $0 < \delta < 1$, and for every $R > 0$,

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R, \|u\| \leq R, \|\hat{u}\| \leq R} \|c_{h,\delta}(x, u, \hat{u}) - c(x, u, \hat{u})\| = 0, \text{ and}$$

$$\lim_{h \downarrow 0} \sup_{\|x\| \leq R, \|u\| \leq R, \|\hat{u}\| \leq R} \|d_{h,\delta}(x, u, \hat{u})\| = 0.$$

Assumption 7. For each $(x, u, \hat{u}) \in R^{n+2m}$, define the ordinary differential equation

$$dY(t, x, u, \hat{u})/dt = c(x, u, [Y(t, x, u, \hat{u}) + u]),$$

with initial condition $Y(0, x, u, \hat{u}) = \hat{u} - u$. Then $0_{m \times 1}$ (an $m \times 1$ vector of zeros) is a globally asymptotically stable solution for bounded values of (x, u, \hat{u}) --i.e., for every $R \geq 0$,

$$\lim_{t \rightarrow \infty} \sup_{\|x, u, \hat{u}\| \leq R} \|Y(t, x, u, \hat{u})\| = 0_{m \times 1}.$$

Assumption 8. There exists a non-negative, twice differentiable function $\rho(x,y,h)$ and a positive function $\lambda(R,h)$ such that

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

$$\limsup_{R \rightarrow \infty} \limsup_{h \downarrow 0} \lambda(R,h) < \infty, \text{ and}$$

$$\limsup_{h \downarrow 0} E_h [\rho({}_hX_0, {}_hY_0h)] < \infty,$$

and for every $R > 0$ and $h > 0$,

$$\sup_{\|x' - u' - \hat{u}'\| \leq R} h^{-1} E_h [\rho({}_hX_{(k+1)h}, {}_hY_{(k+1)h}h) \cdot \rho(x, \hat{\mu} \cdot \mu, h) | {}_hX_{kh} = x, {}_hU_{kh} = u, {}_h\hat{U}_{kh} = \hat{u}] \\ - \lambda(R,h) \rho(x, \hat{u} \cdot u, h) \leq 0.$$

Lemmas needed in the proof of Theorem 2.3

Lemma A.1. For any $\tau \geq 0$, define $\mu_{h,\tau}$ to be the probability measure for $({}_hX_\tau, {}_hU_\tau)$ generated by P_h . For any $\tau < \infty$, there exists an $h^* > 0$ such that $\{\mu_{h,\tau}\}_{0 \leq h \leq h^*}$ is uniformly tight--i.e., for any $\delta > 0$, there exists a compact $\Lambda(\delta) \subset R^{n+2m}$ such that for all h , $0 < h \leq h^*$, $P_h[({}_hX_\tau, {}_hU_\tau) \in \Lambda(\delta)] > 1 - \delta$.

Proof of Lemma A.1. Follows directly from Lemma 2.1 and Dudley (1989, Proposition 9.3.4). ■

Lemma A.2. Let Assumptions 1-4 be satisfied. Let $(x_h, u_h) \rightarrow (x, u)$ as $h \downarrow 0$. Then $P_{(0,x_h,u_h,\tau)} \Rightarrow P_{(0,x,u,\tau)} P_{(h,x_h,u_h,\tau)} \Rightarrow P_{(0,x,u,\tau)} \hat{P}_{(0,x_h,u_h,\tau)} \Rightarrow \hat{P}_{(0,x,u,\tau)} \hat{P}_{(h,x_h,u_h,\tau)} \Rightarrow \hat{P}_{(0,x,u,\tau)}$ as $h \downarrow 0$. If Assumption 5 is also satisfied, then $\hat{P}_{(h,x_h,u_h,\tau)} \Rightarrow P_{(0,x,u,\tau)}$ as $h \downarrow 0$. In each case, this

convergence is uniform on bounded subsets of R^{n+m} (i.e., for bounded (x,u) sets).

Proof of Lemma A.2. See the proof of Stroock and Varadhan (1979, Theorem 11.2.3). Their requirement that "the martingale problem . . . has exactly one solution . . ." is implied by our Assumption 4--see Ethier and Kurtz (1986, Chapter 5, Corollary 3.4).

■

Proof of Theorem 2.3. The Theorem is equivalent to the following: for every $A \in \mathcal{M}_T$, every $\delta > 0$, and every $\epsilon > 0$, there exists an $h^* > 0$ such that for every h , $0 < h \leq h^*$,

$$P_h[|F_h(A, {}_hX_\tau, {}_hU_\tau, \tau) - \hat{F}_h(A, {}_hX_\tau, {}_h\hat{U}_\tau, \tau)| > \epsilon] \leq \delta, \quad (A.1)$$

$$P_h[|F_h(A, {}_hX_\tau, {}_hU_\tau, \tau) - F_0(A, {}_hX_\tau, {}_hU_\tau, \tau)| > \epsilon] \leq \delta, \text{ and} \quad (A.2)$$

$$P_h[|\hat{F}_h(A, {}_hX_\tau, {}_h\hat{U}_\tau, \tau) - \hat{F}_0(A, {}_hX_\tau, {}_h\hat{U}_\tau, \tau)| > \epsilon] \leq \delta. \quad (A.3)$$

We prove (A.1) first. By Lemma A.1 there exists a compact $\Lambda(\delta)$ and an $h^{**}(\delta) > 0$ such that $P_h[({}_hX_\tau, {}_hU_\tau) \in \Lambda(\delta)] > 1 - \delta/2$ when $h \leq h^{**}(\delta)$. By Lemma 2.2, ${}_hU_\tau - {}_h\hat{U}_\tau \rightarrow 0$ in probability under P_h as $h \downarrow 0$. Therefore for every $\eta > 0$ and $\delta > 0$, there exists an $h^{***}(\eta, \delta) > 0$ such that $P_h[\|{}_hU_\tau - {}_h\hat{U}_\tau\| > \eta] < \delta/2$ for $h < h^{***}(\eta, \delta)$. We then have for all $h \leq \min\{h^{**}(\delta), h^{***}(\eta, \delta)\}$,

$$P_h[|F_h(A, {}_hX_\tau, {}_hU_\tau, \tau) - \hat{F}_h(A, {}_hX_\tau, {}_h\hat{U}_\tau, \tau)| > \epsilon] < \delta/2 + \delta/2 + \sup_{(x,u) \in \Lambda(\delta), \|u - \hat{u}\| \leq \eta} I[|F_h(A, x, u, \tau) - \hat{F}_h(A, x, \hat{u}, \tau)| > \epsilon]. \quad (A.4)$$

By lemma A.2, there exists an $\eta(\delta) > 0$ and an $h^{****}(\delta, \epsilon) > 0$ such that $|F_h(A, x, u, \tau) - \hat{F}_h(A, x, \hat{u}, \tau)| \leq \epsilon$ whenever $(x, u) \in \Lambda(\delta)$, $\|u - \hat{u}\| \leq \eta(\delta)$, and $h \leq h^{****}(\delta, \epsilon)$. The bound in (A.4) is therefore achieved whenever $h \leq h^*(\delta) \equiv \min\{h^{**}(\delta), h^{***}(\eta(\delta), \delta), h^{****}(\delta, \epsilon)\}$.

Next we prove (A.2). Define $\Lambda(\delta)$ and $h^{**}(\delta)$ as above. For $h \leq h^{**}(\delta)$,

$$P_h[|F_h(A, {}_hX_\tau, {}_hU_\tau, \tau) - F_0(A, {}_hX_\tau, {}_hU_\tau, \tau)| > \epsilon] < \delta/2 + \sup_{(x,u) \in \Lambda(\delta)} I[|F_h(A, x, u, \tau) - F_0(A, x, u, \tau)| > \epsilon]. \quad (A.5)$$

By Lemma A.2, the convergence of $P_{(h,x,u,\tau)} \Rightarrow P_{(0,x,u,\tau)}$ is uniform on compacts. Therefore the second term in (A.5), which restricts (x,u) to lie in $\Lambda(\delta)$ for all $h \leq h^*(\delta)$, vanishes uniformly in h , proving (A.2). Under Assumption 5 and Lemma 2.2, $P_{(0,x,u,\tau)}$ and $\hat{P}_{(0,x,u,\tau)}$ are identical, so (A.3) follows from (A.1) and (A.2), proving the theorem. ■

Proof of Theorem 2.4. The proof is nearly identical to the proof of Theorem 2.3. The details are left to the reader. ■

Lemmas needed in the Proof of Theorem 2.5

Lemma A.3. Let g be as in Theorem 2.5, but require in addition that g is bounded. Then the conclusion of Theorem 2.5 holds.

Proof of Lemma A.3. Let $\{Z_n\}_{n \rightarrow \infty}$ be a sequence of random variables on some probability space, with $\{Z_n\}_{n \rightarrow \infty} \Rightarrow Z$, and let f be a bounded functional mapping the space into R^1 . Let I^d be the set of discontinuity points of f . Weak convergence implies (see, e.g., Billingsley (1986, Theorem 25.8)) that if $P[Z \in I^d] = 0$, then $E[f(Z_n)] \rightarrow E[f(Z)]$ as $n \rightarrow \infty$. The proof of Lemma A.3 is essentially identical to the proof of Theorem 2.3, substituting $G_h(g, x, u, \tau)$ for $F_h(A, x, u, \tau)$, $G_0(g, x, u, \tau)$ for $F_0(A, x, u, \tau)$, $\hat{G}_h(g, x, u, \tau)$ for $\hat{F}_h(A, x, u, \tau)$, and $\hat{G}_0(g, x, u, \tau)$ for $\hat{F}_0(A, x, u, \tau)$. The details are left to the reader. ■

Lemma A.4. Let the conditions of Lemma 2.1 hold. Define g as in Theorem 2.5. Then

(22) implies that for every bounded $\Lambda \subset \mathbb{R}^{n+m}$,

$$\lim_{K \rightarrow \infty} \sup_{(x,u) \in \Lambda} E_{(0,x,u,\tau)}[|g(\{X_t, U_t\}_{[\tau,\infty)})| \cdot I(|g(\{X_t, U_t\}_{[\tau,\infty)})| > K)] = 0. \quad (A.6)$$

Proof of Lemma A.4. Let \mathcal{A} be a metric space and let f be a mapping from $\mathcal{A} \times \mathbb{R}^1$ into \mathbb{R}^1 . Then clearly

$$\lim_{h \downarrow 0} \sup_{x \in \mathcal{A}} f(x, h) \geq \sup_{x \in \mathcal{A}} \lim_{h \downarrow 0} f(x, h). \quad (A.7)$$

Now by (22) and (A.7), we have

$$\begin{aligned} 0 &= \lim_{K \rightarrow \infty} \lim_{h \downarrow 0} \sup_{(x,u) \in \Lambda} E_{(h,x,u,\tau)}[|g(\{X_t, U_t\}_{[\tau,\infty)})| \cdot I(|g(\{X_t, U_t\}_{[\tau,\infty)})| > K)] \\ &\geq \lim_{K \rightarrow \infty} \sup_{(x,u) \in \Lambda} \lim_{h \downarrow 0} E_{(h,x,u,\tau)}[|g(\{X_t, U_t\}_{[\tau,\infty)})| \cdot I(|g(\{X_t, U_t\}_{[\tau,\infty)})| > K)]. \end{aligned} \quad (A.8)$$

By Fatou's Lemma (Billingsley (1986, Theorem 25.11))

$$\geq \lim_{K \rightarrow \infty} \sup_{(x,u) \in \Lambda} E_{(0,x,u,\tau)}[|g(\{X_t, U_t\}_{[\tau,\infty)})| \cdot I(|g(\{X_t, U_t\}_{[\tau,\infty)})| > K)] \quad (A.9)$$

≥ 0 . ■

Proof of Theorem 2.5. The theorem is equivalent to the following: under (21)-(22) and Assumptions 1-4 and 6-8, there exists, for every $\delta > 0$ and $\epsilon > 0$, an $h^* > 0$ such that for all h , $0 < h < h^*$,

$$P_h[|\hat{G}_h(g, hX_\tau, h\hat{U}_\tau, \tau) - \hat{G}_0(g, hX_\tau, hU_\tau, \tau)| > \epsilon] \leq \delta \quad (27')$$

$$P_h[|G_h(g, hX_\tau, hU_\tau, \tau) - G_0(g, hX_\tau, hU_\tau, \tau)| > \epsilon] \leq \delta \quad (28')$$

and if assumption 5 is also satisfied,

$$P_h[|\hat{G}_h(g, hX_\tau, h\hat{U}_\tau, \tau) - G_h(g, hX_\tau, hU_\tau, \tau)| > \epsilon] \leq \delta. \quad (29')$$

First consider (29'). For any $K > 0$ we can rewrite g as

$$g = (g+K) \cdot I(g < -K) + (g-K) \cdot I(g > K) + \max(-K, \min\{g, K\}). \quad (A.10)$$

Define $\Lambda(\delta)$, η , h^{**} , and h^{***} as in the proof of Theorem 2.3. Then

$$P_h[|\hat{G}_h(g, {}_hX_\tau, {}_h\hat{U}_\tau, \tau) - G_h(g, {}_hX_\tau, {}_hU_\tau, \tau)| > \epsilon] < \delta/2 + \delta/2 + \quad (A.11)$$

$$\sup_{(x,u) \in \Lambda(\delta), \|u - \hat{u}\| \leq \eta} I[|G_h(g, x, u, \tau) - \hat{G}_h(g, x, \hat{u}, \tau)| > \epsilon] \\ \leq \delta/2 + \delta/2$$

$$+ \sup_{(x,u) \in \Lambda(\delta), \|u - \hat{u}\| \leq \eta} I[|G_h((g+K) \cdot I(g < -K), x, u, \tau) - \hat{G}_h((g+K) \cdot I(g < -K), x, \hat{u}, \tau)| > \epsilon/3] \\ + \sup_{(x,u) \in \Lambda(\delta), \|u - \hat{u}\| \leq \eta} I[|G_h((g-K) \cdot I(g > K), x, u, \tau) - \hat{G}_h((g-K) \cdot I(g > K), x, \hat{u}, \tau)| > \epsilon/3] \\ + \sup_{(x,u) \in \Lambda(\delta), \|u - \hat{u}\| \leq \eta} I[|G_h(\max(-K, \min\{g, K\}), x, u, \tau) - \hat{G}_h(\max(-K, \min\{g, K\}), x, \hat{u}, \tau)| > \epsilon/3]. \quad (A.12)$$

By (21) and (22), for any $\delta > 0$ and $\epsilon > 0$, $K(\delta, \epsilon)$ can be selected to make $|G_h((g+K(\delta, \epsilon)) \cdot I(g < -K(\delta, \epsilon)), x, u, \tau) - \hat{G}_h((g+K(\delta, \epsilon)) \cdot I(g < -K(\delta, \epsilon)), x, \hat{u}, \tau)| \leq \epsilon/3$ for $(x, u) \in \Lambda(\delta)$, $\|u - \hat{u}\| \leq \eta$, and all sufficiently small h . The same is true of the $|G_h((g-K(\delta, \epsilon)) \cdot I(g > K(\delta, \epsilon)), x, u, \tau) - \hat{G}_h((g-K(\delta, \epsilon)) \cdot I(g > K(\delta, \epsilon)), x, \hat{u}, \tau)|$ term. For a given $K(\delta, \epsilon)$, $\max(-K(\delta, \epsilon), \min\{g, K(\delta, \epsilon)\})$ is bounded and almost surely continuous for all (x, u, \hat{u}) , so by Lemma A.3, the last term in (A.12) vanishes for sufficiently small h . This bounds (A.12) above by δ , proving (29').

Next consider (28'). For $h < h^{**}$, we have

$$P_h[|G_h(g, {}_hX_\tau, {}_hU_\tau, \tau) - G_0(g, {}_hX_\tau, {}_hU_\tau, \tau)| > \epsilon] < \delta/2 + \sup_{(x,u) \in \Lambda(\delta)} I[|G_h(g, x, u, \tau) - G_0(g, x, u, \tau)| > \epsilon]$$

$$\begin{aligned}
& \leq \sup_{(x,u) \in \Lambda(\delta)} I[|G_h((g+K) \cdot I(g < -K), x, u, \tau)| + |G_0((g+K) \cdot I(g < -K), x, u, \tau)| > \epsilon/3] \\
& + \sup_{(x,u) \in \Lambda(\delta)} I[|G_h((g-K) \cdot I(g > K), x, u, \tau)| + |G_0((g-K) \cdot I(g > K), x, u, \tau)| > \epsilon/3] \quad (A.13) \\
& + \delta/2 + \sup_{(x,u) \in \Lambda(\delta)} I[|G_h(\max(-K, \min\{g, K\}), x, u, \tau) - G_0(\max(-K, \min\{g, K\}), x, u, \tau)| > \epsilon/3].
\end{aligned}$$

By (22) and Lemma A.4, for any $\delta > 0$ and any $\epsilon > 0$, there is a $K(\delta, \epsilon)$ such that

$|G_h((g+K(\delta, \epsilon)) \cdot I(g < -K(\delta, \epsilon)), x, u, \tau)| + |G_0((g+K(\delta, \epsilon)) \cdot I(g < -K(\delta, \epsilon)), x, u, \tau)| \leq \epsilon/3$ whenever $(x, u) \in \Lambda(\delta)$ and h is sufficiently small. The same is true of the

$|G_h((g-K(\delta, \epsilon)) \cdot I(g > K(\delta, \epsilon)), x, u, \tau)| + |G_0((g-K(\delta, \epsilon)) \cdot I(g > K(\delta, \epsilon)), x, u, \tau)|$ term. By Lemma A.3 and the boundedness and almost sure continuity of $\max(-K(\delta, \epsilon), \min\{g, K(\delta, \epsilon)\})$, the last term on the right side of (A.13) also vanishes for sufficiently small h , completing the proof of (28').

The proof of (27') is similar to the proofs of (28')-(29'). (22') and (23') follow by Billingsley (1986, 25.18). ■

Proof of Theorem 2.6. The theorem follows directly as a special case of Theorems 3 and 6 in Chapter 6 of Kushner (1984): both Kushner's Liapunov function $V(x, u)$ and his perturbed Liapunov function $V^\epsilon(x, u)$ are set equal to our $\zeta(x, u)$ and to $\hat{\zeta}(x, u)$ in (19)-(21). This makes Kushner's $\delta_\epsilon \equiv 0$ in both cases. We set Kushner's ψ^ϵ equal to the constant Δ . The moment boundedness conditions (23) and (24) imply, respectively, the tightness of $\{hX_t, {}_hU_t\}$ under P_h and $\{hX_t, {}_h\hat{U}_t\}$ under \hat{P}_h as $t \rightarrow \infty$ and $h \downarrow 0$. (See the discussion on Kushner (1984, pp. 151 and 157).) The proof then follows by Theorem 6 in Chapter 6 of Kushner. ■

Proof of Theorem 3.1. (a) To apply Lemma 2.2, we must verify Assumptions 1-4 and 6-8. It is convenient to do a change of variables in (36)-(37) by defining $X_t \equiv \ln(S_t)$ and $U_t \equiv \ln(\sigma_t^2)$. We then have

$$\begin{aligned} dX_t &= (\mu - \exp(U_t)/2)dt + \exp(U_t/2) \cdot dW_{1,t}, \text{ and} \\ dU_t &= -\beta[U_t - \alpha]dt + \Lambda \cdot dW_{2,t}. \end{aligned} \quad (\text{A.14})$$

Assumption 1 is immediate, since ${}_h S_0$, ${}_h \sigma_0^2$, and ${}_h \hat{\sigma}_0^2$ are constant for all h . We next verify Assumption 2. (14) is immediate for any $\epsilon > 0$, since U_{t+h} is Gaussian with moments continuous in U_t (Arnold (1973, Section 8.3)). By (A.14),

$$(X_t - X_0 - \mu \cdot t - (1/2) \int_0^t \exp(U_s) ds) = \int_0^t \exp(U_s/2) \cdot dW_{1,s}.$$

By Karatzas and Shreve (1988 p. 163, Exercise 3.25),

$$\begin{aligned} h^{-1} E_{(h,x,u,\tau)}[|X_{t+h} - X_t - \mu \cdot h - (1/2) \int_t^{t+h} \exp(U_s) \cdot ds|^6] \\ \leq h \cdot 15^3 \cdot E_{(h,x,u,\tau)} \int_t^{t+h} \exp(3 \cdot U_s) \cdot ds. \end{aligned} \quad (\text{A.15})$$

Fubini's Theorem (e.g., Dudley (1989, Theorem 4.4.5)), allows us to interchange the expectations and integral in the right hand side of (A.15). $\exp(3 \cdot U_s)$ is lognormal and has a bounded expectation on $[t, t+h]$, bounding the expectation on the left. Since $\exp(U_s)$ has arbitrary finite moments (which are continuous in U_t), Assumption 2 follows.

We next check Assumption 3: (16) is immediate from the definitions of μ and Ω in (36)-(37) and Assumption 2 (see, e.g., Arnold (1973, p. 40.)). (12') follows as in the proof of Nelson (1990, Theorem 3.3), verifying Assumption 3. Assumption 4 also follows as in the proof of Nelson (1990, Theorem 3.3).

Assumptions 6, 7, and 8 are verified in Nelson (1992, Theorem 3.2) under Assumptions 1-4 and the following additional conditions:

First, there is an $\varepsilon > 0$ such that for every $\eta > 0$,

$$\lim_{h \downarrow 0} \sup_{\|x, u\| \leq \eta} h^{-1} E_{(h, x, u, \tau)} [\|h U_{(k+1)h} - u\|^{2+\varepsilon}] = 0. \quad (\text{A.16})$$

Second, there is a twice differentiable, non-negative function $w(x, u)$ and a $\Theta > 0$ such that for every $\eta > 0$,

$$\liminf_{\|x, u\| \rightarrow \infty} w(x, u) = \infty, \quad (\text{A.17})$$

$$\lim_{h \downarrow 0} \sup_{\|x, u\| \leq \eta} h^{-1} E_{(h, x, u, \tau)} ([|w(h X_{(k+1)h}, h U_{(k+1)h}) - w(x, u)|^{1+\Theta}] < \infty, \quad (\text{A.18})$$

and there is a $\lambda > 0$ such that for all $(x, u) \in \mathbb{R}^2$,

$$\begin{aligned} & (\mu \cdot e^u / 2) \frac{\partial w(x, u)}{\partial x} - \beta [u - \alpha] \frac{\partial w(x, u)}{\partial u} + \Lambda^2 \frac{\partial^2 w(x, u)}{2 \partial u^2} + e^u \frac{\partial^2 w(x, u)}{2 \partial x^2} + \rho \Lambda e^{u/2} \frac{\partial^2 w(x, u)}{\partial x \partial u} \\ & \leq \lambda w(x, u). \end{aligned} \quad (\text{A.19})$$

(A.16) is immediate, since U_t is Gaussian and therefore has arbitrary finite conditional moments. To verify (A.17)-(A.18), we set

$$w(x, u) \equiv 1 + (1 - \exp(-x^2)) \cdot |x| + \exp(u) + \exp(-u).$$

(A.17) is trivial. It is easy to verify that $w(x, u)$ is arbitrarily differentiable on \mathbb{R}^2 (including when $x = 0$) and that $-2 < \partial w / \partial x < 2$ and $-2 < \partial^2 w / \partial x^2 < 2$ and $\partial^2 w(x, u) / \partial x \partial u = 0$ for all (x, u) . This bounds the left hand side of (A.19) from above by

$$2|\mu| + 2 \exp(u) - \beta [\exp(u) - \exp(-u)] \cdot (u - \alpha) + \Lambda^2 [\exp(u) + \exp(-u)] / 2, \quad (\text{A.20})$$

which we require to be bounded above by the right-hand side of (A.19), which equals $\lambda(1 + |x| \cdot (1 - \exp(-x^2)) + \exp(u) + \exp(-u))$. Since $\beta \geq 0$, this holds for all $(x, u) \in \mathbb{R}^2$ if we choose a sufficiently large λ . To verify (A.18), we choose $\Theta = 1$. The finiteness of the expected increments in X_t^2 follows from Assumption 2. Further, since $\pm U_t$ is normal

(conditionally and unconditionally), $\exp(\pm U_t)$ is lognormal and its increments have finite expectation. (A.18) follows. Assumptions 6, 7, and 8 now follow from Nelson (1992, Theorem 3.2), concluding the proof of (a).

(b) Assumption 5 is verified in Nelson (1990, Section 3.3). Theorem 2.3 immediately follows, implying Theorem 3.1 (b).

(c) By (21')-(22') and the results of parts (a) and (b) of Theorem 2.5, it is sufficient that for every $\tau > 0$, every bounded $\Lambda \subset \mathbb{R}^2$, and for all bounded nonnegative A, B, a, b, c , and d ,

$$\limsup_{h \downarrow 0} \sup_{(\ln(s), \ln(\sigma)) \in \Lambda} \hat{E}_{(h,s,\sigma,\tau)}[|G((\ln(hS_t), \ln(hS_T), {}_h\sigma_t, {}_h\sigma_T))|^{1+\epsilon}] < \infty, \text{ and} \quad (\text{A.21})$$

$$\limsup_{h \downarrow 0} \sup_{(\ln(s), \ln(\sigma)) \in \Lambda} E_{(h,s,\sigma,\tau)}[|G((\ln(hS_t), \ln(hS_T), {}_h\sigma_t, {}_h\sigma_T))|^{1+\epsilon}] < \infty. \quad (\text{A.22})$$

We prove (A.22) first. This relation clearly holds when

$$\limsup_{h \downarrow 0} \sup_{(\ln(s), \ln(\sigma)) \in \Lambda} E_{(h,s,\sigma,\tau)}[|\ln(hS_t)|^a \cdot |\ln(hS_T)|^b \cdot |{}_h\sigma_t|^c \cdot |{}_h\sigma_T|^d] < \infty. \quad (\text{A.23})$$

for arbitrary nonnegative τ, t, T, a, b, c, d (we can ignore the $1+\epsilon$ term since a, b, c , and d can be made *arbitrarily* large). Because of the Markov structure of $({}_hS_{kh}, {}_h\sigma_{kh}^2)$ and the arbitrariness of t and T , τ is irrelevant, so we set it equal to zero. Applying Hölder's inequality and the arbitrariness of a, b, c , and d , (A.22) holds if for arbitrary nonnegative a, b , and $t \geq 0$,

$$\limsup_{h \downarrow 0} \sup_{(\ln(s), \ln(\sigma)) \in \Lambda} E_{(h,s,\sigma,0)}[|\ln(hS_t)|^a] < \infty, \text{ and} \quad (\text{A.24})$$

$$\limsup_{h \downarrow 0} \sup_{(\ln(s), \ln(\sigma)) \in \Lambda} E_{(h,s,\sigma,0)}[{}_h\sigma_t^b] < \infty. \quad (\text{A.25})$$

The lognormality of σ_t^2 implies (A.25). Using Fubini's theorem and Karatzas and Shreve

(1988, p. 163, Exercise 3.25), (A.24) follows by the argument used to establish (A.15) above, proving (A.22).

We now turn to (A.21). By the same argument as in (A.25)-(A.25), (A.21) follows if for arbitrary nonnegative a , b , and $t \geq 0$,

$$\limsup_{h \downarrow 0} \sup_{(s, \ln(\hat{\sigma})) \in \Lambda} \hat{E}_{(h, s, \hat{\sigma}, 0)}[|\ln(h S_t)|^a] < \infty, \text{ and} \quad (\text{A.26})$$

$$\limsup_{h \downarrow 0} \sup_{(s, \ln(\hat{\sigma})) \in \Lambda} \hat{E}_{(h, s, \hat{\sigma}, 0)}[h \hat{\sigma}_t^b] < \infty. \quad (\text{A.27})$$

We consider (A.20) first. Using (25') and (27)-(31), we write the expectation in (A.27) as

$$\hat{\sigma}^{b(1-\beta h)^{k+1}} \cdot f(\alpha b, \beta h, k) \cdot \prod_{i=0, k} \hat{E}_h[\exp(bh^{1/2}(1-\beta h)^i g(h Z_{(k-i)h})/2)], \quad (\text{A.28})$$

where $k \equiv [t/h]$, and

$$f(\alpha b, \beta h, k) \equiv \exp[b\alpha(1 - (1-\beta h)^{k+1})/2]. \quad (\text{A.29})$$

By Nelson (1991, Theorem A1.1), we have, for any real λ ,

$$E[e^{g(z)\lambda}] = \{ \Phi(\gamma\lambda + \theta\lambda)e^{\lambda^2(\theta+\gamma)^2/2} + \Phi(\gamma\lambda - \theta\lambda)e^{\lambda^2(\gamma-\theta)^2/2} \} \cdot e^{-\lambda\gamma(2/\pi)^{1/2}} \quad (\text{A.30})$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal, and γ and θ are as in (40)-(41). Considered as a function of $\lambda\gamma$ and $\lambda\theta$, it is easy to check that $E[e^{g(z)\lambda}] \leq E[e^Y]$, where $Y \sim N(0, \lambda^2(\gamma + |\theta|)^2)$. Substituting into (A.28) yields

$$\hat{E}_{(h, s, \hat{\sigma}, 0)}[h \hat{\sigma}_t^b] \leq \hat{\sigma}^{b(1-\beta h)^{k+1}} \cdot f(\alpha b, \beta h, k) \cdot \exp[b^2 h (\gamma + |\theta|)^2 \sum_{i=0, k} (1-\beta h)^{2i}/8].$$

Applying Taylor series and binomial expansions (e.g., Silverman (1972 p. 202)) to $f(\cdot, \cdot, \cdot)$ and $(1-\beta h)^k$ and evaluating the sum $\sum_{i=1, k} (1-\beta h)^{2i}$, we obtain

$$\hat{E}_{(h, s, \hat{\sigma}, 0)}[h \hat{\sigma}_t^b] \leq (1+O(h)) \cdot \hat{\sigma}^b \cdot e^{\beta t} \cdot \exp(-b\beta(1-e^{-\beta t})/2) \cdot \exp[(b^2(\gamma + |\theta|)^2 t/8)]. \quad (\text{A.31})$$

The bound (A.31) is uniform for bounded $(s, \hat{\sigma}, t)$ and holds for arbitrary nonnegative b . (A.27) follows.

Finally, we prove (A.26). Here, it helps to define the $\{ {}_h Z_{kh} \}$ in (27) by ${}_h Z_{kh} \equiv h^{-1/2} \cdot [W_{kh} - W_{(k-1)h}]$, where W_t is a standard Brownian motion on $[-h, \infty)$. This allows us to write for each positive h and t

$$\ln({}_h S_t) - \ln({}_h S_0) - h[t/h]\mu - \int_0^{h[t/h]} {}_h \hat{\sigma}_s^2/2 \cdot ds = \int_0^{h[t/h]} {}_h \hat{\sigma}_s \cdot dW_s \quad (A.32)$$

under \hat{P}_h . Since ${}_h \hat{\sigma}_s^2$ has arbitrary finite moments, the moments of $\ln({}_h S_t)$ will be bounded if and only if the moments on the right hand side of (A.32) are. (A.26) will therefore be proved if we can show that for arbitrary positive a and t and arbitrary bounded $\Lambda \subset \mathbb{R}^2$,

$$\limsup_{h \downarrow 0} \sup_{(s, \ln(\sigma)) \in \Lambda} \hat{E}_{(h, s, \sigma, 0)} [|\int_0^{h[t/h]} {}_h \hat{\sigma}_t \cdot dW_s|^a] < \infty. \quad (A.33)$$

By Karatzas and Shreve (1988 p. 163, Exercise 3.25), (A.33) is bounded above by

$$\limsup_{h \downarrow 0} \sup_{(s, \ln(\sigma)) \in \Lambda} [a(a-1)/2]^{a/2} \cdot t^{(a/2 - 1)} \hat{E}_{(h, s, \sigma, 0)} \int_0^{h[t/h]} {}_h \hat{\sigma}_t^a \cdot ds \quad (A.34)$$

Since ${}_h \hat{\sigma}_t$ is a step function, we can rewrite (A.34) as

$$\limsup_{h \downarrow 0} \sup_{(s, \ln(\sigma)) \in \Lambda} [a(a-1)/2]^{a/2} \cdot t^{(a/2 - 1)} \hat{E}_{(h, s, \sigma, 0)} [\sum_{i=0, \{t/h\}} {}_h \hat{\sigma}_{ih}^a \cdot h] \quad (A.35)$$

But by (A.31), the terms in the summation in (A.35) are uniformly bounded for bounded (s, σ, t) , proving (A.26) and completing the proof of (c).

(d) When $|\gamma| \geq |\theta|$, Theorem 2.6 can be employed to prove (d) (see the very similar application in the proof of Nelson (1990, Theorem 2.3)). To prove (d) in the general case, it is easier to employ the Lyapunov central limit theorem for triangular arrays (see e.g., Billingsley (1986, Theorem 27.3)). When $\beta > 0$, the stationary distribution of $\ln(\sigma_t^2)$ is $N[\alpha, \Lambda^2/2\beta]$ (see Arnold (1973, Section 8.3)). When the AR(1) EGARCH model is the data generating process we have

$$\ln(\sigma_{kh}^2) = \alpha + (1-h\beta)^k \cdot [\ln(\sigma_0^2) - \alpha] + \sum_{j=1,k} (1-h\beta)^{j-1} \cdot h^{1/2} \cdot g(Z_{(k-j)h}).$$

If $|1 - \beta h| < 1$, we let $k \rightarrow \infty$ to obtain the stationary distribution of $\ln(\sigma_{kh}^2)$ as the distribution of

$$\ln(\sigma_{\infty}^2) \equiv \alpha + \sum_{i=0,\infty} (1-h\beta)^i \cdot h^{1/2} \cdot g(Z_i),$$

where $\{Z_i\} \sim \text{i.i.d. } N(0,1)$. From the normality of Z_i and the definition of $g(\cdot)$, it is clear that $g(Z_i)$ possesses finite moments of arbitrary finite order. We also have

$$\text{Var}[\ln(\sigma_{\infty}^2)] = \text{Var}[g(Z_i)]/[2\beta - \beta^2 \cdot h], \text{ and} \quad (\text{A.36})$$

$$E[(1-h\beta)^i \cdot h^{1/2} \cdot g(Z_i)]^4 = h^2 E[g(Z_i)]^4 (1-h\beta)^{4i}. \quad (\text{A.37})$$

The Lyapunov Condition is satisfied if

$$0 = \lim_{h \downarrow 0} \sum_{i=0,\infty} E[(1-h\beta)^i \cdot h^{1/2} \cdot g(Z_i)]^4 / [\text{Var}[\ln(\sigma_{\infty}^2)]^2].$$

But from (A.36)-(A.37),

$$= \lim_{h \downarrow 0} \frac{h \cdot [4\beta^2 + O(h)] \cdot E[g(Z_i)]^4}{[\text{Var}(g(Z_i))]^2 \cdot [4\beta + O(h)]} = 0.$$

Applying the Lyapunov central limit theorem for triangular arrays, we have

$$\ln(\sigma_{\infty}^2) \Rightarrow N(\alpha, \text{Var}[g(Z_i)]/2\beta).$$

We complete the proof by invoking (40). ■

References

- Anderson, Brian D. O., and John B. Moore, 1979, *Optimal Filtering* (Prentice-Hall, Inc.: New York).
- Arnold, Ludwig, 1973, *Stochastic Differential Equations: Theory and Applications* (John Wiley and Sons: New York).
- Billingsley, Patrick, 1986, *Probability and Measure*, Second edition, (John Wiley and Sons: New York).
- Bollerslev, Tim, Ray Y. Chou, and Kenneth F. Kroner, 1992, ARCH modeling in finance: A review of the theory and empirical evidence, *Journal of Econometrics*, 52, 5-60.
- Cao, Charles Q., 1992, Pricing options with stochastic volatility: A general equilibrium approach. Mimeo, University of Chicago GSB.
- Dudley, Richard M., 1989, *Real Analysis and Probability* (Wadsworth & Brooks/Cole: Pacific Grove, CA).
- Engle, Robert F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987-1008.
- 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, Stewart N. and Thomas G. Kurtz, 1986, *Markov Processes: Characterization and Convergence*, (John Wiley and Sons: New York).
- Ethier, Stewart N. and Thomas Nagylaki, 1988, *Diffusion approximations of Markov chains*

- with two time scales and applications to population genetics, II, *Advances in Applied Probability*, 20, 525-545.
- Foster, Dean P. and Daniel B. Nelson, 1991, Rolling regressions, working paper, (University of Chicago Graduate School of Business: Chicago, IL).
- Hull, John and Allan White, 1987, The pricing of options on assets with stochastic volatilities, *Journal of Finance*, 42, 281-300.
- Karatzas, Ioannis and Steven E. Shreve, 1988, *Brownian Motion and Stochastic Calculus* (Springer Verlag: New York).
- Kuwahara, Hiroto, and Terry A. Marsh, 1992, The pricing of Japanese equity warrants, forthcoming, *Management Science*.
- Kushner, Harold J., 1984, *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory* (M.I.T. Press: Cambridge, MA).
- Lumsdaine, Robin L., 1991, Asymptotic properties of the quasi-maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models. Unpublished manuscript, Princeton University.
- Melino, Angelo and Stuart M. Turnbull, 1990, Pricing foreign currency options with stochastic volatility, *Journal of Econometrics*, 45, 239-267.
- Nelson, Daniel B., 1988, The time series behavior of stock market volatility and returns. Unpublished doctoral dissertation, M.I.T. Economics Department.
- Nelson, Daniel B., 1990, ARCH models as diffusion approximations, *Journal of Econometrics*, 45, 7-38.

- Nelson, Daniel B., 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347-370.
- Nelson, Daniel B., 1992, Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model, *Journal of Econometrics* 52, 61-90.
- Nelson, Daniel B., and Dean P. Foster, 1992, Asymptotic Filtering Theory for Univariate ARCH Models, working paper (University of Chicago Graduate School of Business: Chicago, IL).
- Pardoux, E. and D. Talay, 1985, Discretization and simulation of stochastic differential equations, *Acta Applicandae Mathematicae*, 3, 23-47.
- Ruiz, Esther, 1992, Quasi-maximum likelihood estimation of stochastic variance models. Unpublished manuscript, London School of Economics.
- Schwartz, Boaz A., Daniel B. Nelson, and Dean P. Foster, 1992, Variance filtering with ARCH models: A monte carlo investigation. Unpublished manuscript, University of Chicago GSB.
- Scott, Louis O., 1987, Option pricing when the variance changes randomly: Theory, estimation, and an application, *Journal of Financial and Quantitative Analysis*, 22, 419-438.
- Silverman, Richard A., 1972, *Introductory Complex Analysis*, New York: Dover.
- Stroock, Daniel W. and S. R. S. Varadhan, 1979, *Multidimensional Diffusion Processes* (Springer Verlag: Berlin).
- Wiggins, James B., 1987, Option values under stochastic volatility: Theory and empirical estimates, *Journal of Financial Economics*, 19, 351-372.