Econometrica, Vol. 63, No. 5 (September, 1995), 1113–1132

REGRESSION WITH NONSTATIONARY VOLATILITY

BY BRUCE E. HANSEN[1]

A new asymptotic theory of regression is introduced for possibly nonstationary time series. The regressors are assumed to be generated by a linear process with martingale difference innovations. The conditional variances of these martingale differences are specified as autoregressive stochastic volatility processes, with autoregressive roots which are local to unity.

We find conditions under which the least squares estimates are consistent and asymptotically normal. A simple adaptive estimator is proposed which achieves the same asymptotic distribution as the generalized least squares estimator, without requiring parametric assumptions for the stochastic volatility process.

KEYWORDS: Conditional heteroskedasticity, stochastic volatility, adaptive estimation, integrated processes, stochastic integrals.

1. INTRODUCTION

MANY ECONOMETRICIANS ARE BEGINNING TO SERIOUSLY ENTERTAIN the notion that some economic series might violate the assumption of covariance stationarity. Covariance stationarity is a very strong assumption, requiring time invariance of unconditional variances and autocovariances. Casual examination of plots of recursive and rolling estimates of variances for many series, however, suggests nonconstancy. Recent papers which formally discuss this phenomenon include DeLong and Summers (1986), Pagan and Schwert (1990a, 1990b), and Phillips and Loretan (1990).

The finding of covariance nonstationarity has implications for both economic and econometric theory. This paper is concerned exclusively with the second topic. Virtually all econometric theory (with the exception of the literature on unit roots and cointegration) assumes that the data are draws from stationary distributions (or asymptotically stationary distributions, such as mixing processes). The implicit assumption is that if the data are approximately stationary, then the use of the theory for stationary random variables is still useful. This view seems reasonable, if the departures from stationarity are minor. On the other hand, if the departures from stationarity are substantial then it seems clear that we need a new theory, and it is currently unknown what constitutes a "minor" or a "substantial" departure. This paper attempts to break new ground by developing a large-sample distribution theory for random variables with possible nonstationarity in the variance.

To handle the difficult concept of nonstationary variances, we work with a class of autoregressive *stochastic volatility* processes. In these models, the "variance" is a nonlinear transformation of a latent autoregressive process. This

[1]A preliminary version of this research was circulated as "Regression Theory When Variances Are Non-Stationary," RCER Working Paper No. 226. I thank two referees and the co-editor for helpful comments and suggestions and the NSF for research support.

is an increasingly popular class of models for time-varying conditional heteroskedasticity. Financial economists (see, e.g., Hull and White (1987), Wiggins (1987), and Andersen (1993)) have developed models with stochastic volatility, and recently Nelson (1990) and Nelson and Foster (1994) have shown that near-integrated GARCH models provide a good approximation to stochastic volatility processes, suggesting one explanation for the empirical success of GARCH models.

We allow for nonstationarity by letting the autoregressive root of the stochastic volatility process to be close to one. Technically, this allows for "nearly nonstationary" volatility processes as well as strictly nonstationary processes, but we will retain the nomenclature of nonstationarity for simplicity. Nonstationary (or near nonstationary) volatility seems a reasonable characterization in practice. Indeed, most of the theoretical literature has assumed that the stochastic volatility autoregressive root is unity, and most of the applied literature finds estimated values close to one.

Our question concerns the behavior of ordinary least squares (OLS), generalized least squares (GLS), and adaptive least squares (ALS) estimation of linear regression models. While a stochastic volatility structure could be used to compute quasi-MLE or GMM estimates, this would require a complete specification of the model. In contrast, our adaptive estimation procedure does not require knowledge of any functional forms other than the regression equation. The volatility is treated as a nuisance in the estimation of the regression parameters.

A large sample theory of inference is derived. We demonstrate consistency and asymptotic mixture normality, which is sufficient to justify construction of conventional confidence intervals and test statistics. We show that the feasible ALS estimator is asymptotically equivalent to the theoretical GLS estimator.

This paper may be seen as contributing to the econometric literature on estimation with nonergodic data. Early work included Robinson (1978) and Anderson and Taylor (1979), who considered linear estimation with nonconstant variances. More recently, Harvey and Robinson (1988) considered GLS estimation in the presence of deterministic nonstationary variances, and Wooldridge and White (1988) and Davidson (1992) studied central limit theory for processes with nonstationary, yet asymptotically constant, variances.

Section 2 introduces the model and assumptions. Section 3 examines linear estimation. Section 4 examines adaptive estimation. Section 5 contains a brief conclusion, and the Appendix contains the proofs of the theorems.

Throughout the paper $|\cdot|$ refers to the Euclidean norm $|A| = (\operatorname{tr}(A'A))^{1/2}$, $\|\cdot\|_p$ to the L_p-norm $\|A\|_p = (E|A|^p)^{1/p}$, and $[\cdot]$ to integer part. The symbol $\Rightarrow$ denotes weak convergence with respect to the uniform metric on $[0,1]$. All limits are taken as the sample size, n, diverges to positive infinity. For brevity, stochastic integrals such as $\int_0^1 X(r)\,dB(r)$ will be written as $\int_0^1 X\,dB$ and integrals with respect to Lebesgue measure $\int_0^1 X(r)Z(r)\,dr$ will be written as $\int_0^1 XZ$.

2. MODEL AND ASSUMPTIONS

2.1. *Regression Equation*

Let $\{y_{ni}, x_{ni}: 1 \leq i \leq n+1\}$ be a random array, where y_{ni} and x_{ni} are real-valued. The regression model of interest is the following linear relationship

$$(1) \qquad y_{ni} = \mu + \beta x_{n,i-1} + u_{ni}.$$

For some increasing array of sigma-fields $\{\Im_{ni}: i \leq n\}$ to which y_{ni} and x_{ni} are adapted, we assume that u_{ni} is a martingale difference array (MDA):

$$E(u_{ni}|\Im_{n,i-1}) = 0 \quad \text{a.s.}$$

and x_{ni} is a linear process with martingale difference innovations:

$$(2) \qquad x_{ni} = \mu_x + \sum_{k=0}^{\infty} a_k v_{n,i-k}, \qquad E(v_{ni}|\Im_{n,i-1}) = 0 \quad \text{a.s.}$$

where $a_0 = 1$.

2.2. *Conditional Variances*

Define

$$\sigma_{n,i-1}^2 = E(u_{ni}^2|\Im_{n,i-1}) \quad \text{a.s.}$$

and

$$\nu_{n,i-1}^2 = E(v_{ni}^2|\Im_{n,i-1}) \quad \text{a.s.}$$

Note that $\{\sigma_{ni}^2, \nu_{ni}^2\}$ are the conditional variances of $\{u_{ni}, v_{ni}\}$ with respect to the array $\{\Im_{ni}\}$. When these variances are unobservable (not functions of the past history of y_{ni} and x_{ni} alone) they are typically called *volatility* processes.

Define the normalized errors

$$\varepsilon_{ni} = u_{ni}/\sigma_{n,i-1}$$

and

$$\xi_{ni} = v_{ni}/\nu_{n,i-1}.$$

By construction, $\{\varepsilon_{ni}, \Im_{ni}\}$ and $\{\xi_{ni}, \Im_{ni}\}$ are martingale differences with unit conditional variances.

We assume that the model satisfies the following standard moment and decay conditions:

ASSUMPTION 1: *For some $p > 2$, $r > 2$, and $\bar{a} < \infty$,*

$$(\text{i}) \qquad E(|\varepsilon_{ni}|^p|\Im_{n,i-1}) \leq Q_\varepsilon < \infty \quad a.s.;$$

(ii) $E|\xi_{ni}|^p \le Q_\xi < \infty$;

(iii) $|a_m| \le \bar{a} m^{-r}$.

We assume that the volatilities are functions of first-order autoregressive processes.

ASSUMPTION 2: *The volatility processes* σ_{ni}^2 *and* ν_{ni}^2 *are generated by*

$$\sigma_{ni}^2 = h_1(\omega_1 + \tau_{n1} S_{n1i}),$$
$$\nu_{ni}^2 = h_2(\omega_2 + \tau_{n2} S_{n2i}),$$

where $h_1(\cdot) \ge b > 0$ *and* $h_2(\cdot) \ge b > 0$ *are real continuous functions, and*

$$S_{n1i} = \rho_{n1} S_{n1,i-1} + z_{n1i},$$
$$S_{n2i} = \rho_{n2} S_{n2,i-1} + z_{n2i}.$$

Setting $z_{ni} = (z_{n1i}, z_{n2i})$, $\{z_{ni}, \Im_{ni}\}$ *is a uniformly square integrable martingale difference array with conditional covariance matrix* $E(z_{ni} z_{ni}' | \Im_{n,i-1}) = \Omega_z$ *a.s., where* Ω_z *is normalized to have unit elements on the diagonal. The system is initialized by setting* $S_{n10} = S_{n20} = 0$. *The coefficients* τ_{n1}, τ_{n2}, ρ_{n1}, *and* ρ_{n2} *are given by*

(3) $$\tau_{n1} = \eta_1/\sqrt{n}, \qquad \tau_{n2} = \eta_2/\sqrt{n},$$

(4) $$\rho_{n1} = 1 - c_1/n, \quad \textit{and} \quad \rho_{n2} = 1 - c_2/n,$$

for finite constants η_1, η_2, c_1, c_2.

Stochastic volatility models as in Assumption 2 with $h(\cdot) = \exp(\cdot)$ have been proposed in the finance literature as discrete-time versions of continuous-time asset pricing models. See, e.g., Hull and White (1987), Wiggins (1987), Chesney and Scott (1989), and Andersen (1993). There have been several recent attempts to empirically estimate stochastic volatility models, including Melino and Turnbull (1990), Shephard (1994), Harvey, Ruiz, and Shephard (1994). Most of these papers either impose a unit root in the variance equation ($\rho_{n1} = \rho_{n2} = 1$) or estimate a value which is close to one. This motivates our local-to-unity parameterization (4). It allows for unit roots ($c_1 = c_2 = 0$) or for roots close to unity (including mildly explosive roots).

The assumption that the autoregressive stochastic volatility processes are first-order is not essential to the analysis. Indeed, higher-level dynamics could be allowed, but the first-order case is consistent with the stochastic volatility literature and illustrates the main points of the paper without needless generality.

The specification of the variance equation coefficients given in (3) and (4) have been selected to be able to derive a useful approximation to the finite sample distributions of relevant test statistics. The local-to-unity specification (4) has been found to provide good approximations for the finite sample distributions of test statistics constructed from time series with autoregressive roots close to unity. This formulation allows the distribution theory to be continuous as an autoregressive root approaches unity, rather than discontinuous. One interpretation of the local-to-unity specification (3) is that the stochastic component of volatility is small relative to the deterministic part. Another is that it is the only practical way to generate a sensible distribution theory. As we show in the Appendix, Assumption 2 implies that

$$\sigma_{n[ns]}^2 \Rightarrow h_1(\omega_1 + \eta_1 W_1^c(s)) \tag{5}$$

where $W_1^c(s)$ is a diffusion process. Finding a weak distributional limit for the variance process turns out to be an essential part of our regression theory. Are there any practical alternatives to the array framework of (3)–(4)? When $h_1(\cdot)$ takes the power form $h_1(x) = |x|^{\alpha}$ for some $\alpha > 0$ (Andersen (1993) calls this "polynomial stochastic autoregressive volatility"), one could set $\tau_{n1} = \tau_1$ (a constant) and then normalize σ_{ni}^2 by $n^{-1/2\alpha}$. In this event,

$$n^{-1/2\alpha}\sigma_{n[ns]}^2 = |n^{-1/2}\omega_1 + n^{-1/2}\tau_1 S_{n1[ns]}|^{\alpha} \Rightarrow |\tau_1 S_1^c(s)|^{\alpha}. \tag{6}$$

Note in contrast that our approximation (5) would be $|\omega_1 + \eta_1 W_1^c(s)|^{\alpha}$, which is superior to (6), as the latter essentially ignores the intercept ω_1. The situation is even more problematic when h_1 does not take a power form (for example, when $h_1(\cdot) = \exp(\cdot)$ as more commonly assumed in the stochastic volatility literature) for there is no normalization which generates a nondegenerate asymptotic theory.

3. LINEAR ESTIMATION

3.1. *Weighted Least Squares*

We consider weighted least squares (WLS) estimates of equation (1). For some array of weights $\{w_{ni}^2\}$, define the WLS estimator

$$\begin{pmatrix} \tilde{\mu}_n \\ \tilde{\beta}_n \end{pmatrix} = \left(\sum_{i=1}^{n} w_{ni}^2 x_{ni}^* x_{ni}^{*\prime} \right)^{-1} \left(\sum_{i=1}^{n} w_{ni}^2 x_{ni}^* y_{n,i+1} \right)$$

where $x_{ni}^* = (1 \;\; x_{ni})'$. We focus on the slope coefficient $\tilde{\beta}_n$. The class of WLS estimators is useful for it contains ordinary least squares (OLS), generalized least squares (GLS), and feasible generalized least squares (FGLS) as special cases, setting w_{ni}^2 equal to 1, σ_{ni}^{-2}, and $\tilde{\sigma}_{ni}^{-2}$, respectively, where $\tilde{\sigma}_{ni}^2$ is some estimate of σ_{ni}^2. We will denote the OLS estimator by $(\hat{\mu}_n, \hat{\beta}_n)'$.

ASSUMPTION 3: (i) w_{ni}^2 *is adapted to* $\Im_{ni}$; (ii) *for all* $n, i, 0 \leq w_{ni}^2 \leq H < \infty$; (iii) $w_{n[ns]}^2 \Rightarrow w^2(s)$, *where* $w^2(\cdot)$ *has continuous sample paths with probability one*; (iv) $\int_0^1 w^2 > 0$ *a.s.*

We now give the first main result of the paper.

THEOREM 1: *Under Assumptions* 1–3,

$$\sqrt{n}\,(\tilde{\beta}_n - \beta) \Rightarrow \frac{\int_0^1 w^2 \sigma v \, dW_0}{A \int_0^1 w^2 v^2}, \tag{7}$$

where

$$A^2 = \sum_{k=0}^{\infty} a_k^2,$$

$$\sigma^2(s) = h_1(\omega_1 + \tau_1 W_1^c(s)), \tag{8}$$

$$v^2(s) = h_2(\omega_2 + \tau_2 W_2^c(s)), \tag{9}$$

$$dW_1^c(s) = -c_1 W_1^c(s) + dW_1(s), \tag{10}$$

$$dW_2^c(s) = -c_2 W_2^c(s) + dW_2(s), \tag{11}$$

and $W(s) = (W_0(s), W_1(s), W_2(s))$ *is a vector Brownian motion with covariance matrix*

$$E(W(1)W(1)') = \begin{pmatrix} 1 & \kappa' \\ \kappa & \Omega_z \end{pmatrix}. \tag{12}$$

In (12), $\kappa = A^{-1} \sum_{k=0}^{\infty} a_k \kappa_k$, *where*

$$\kappa_k = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E(z_{n,i+1} \xi_{n,i-k} \varepsilon_{n,i+1}).$$

Theorem 1 shows that the asymptotic distribution of the WLS estimator has a stochastic integral representation. The form of the limiting theory resembles that found in the "unit root" and cointegration literature, e.g. Park and Phillips (1988). By itself, the limiting distribution (7) is not very informative, other than yielding the implication that $\tilde{\beta}$ is consistent for β at the rate $\sqrt{n}$.

3.2. *Ordinary Least Squares*

The first major implication of Theorem 1 concerns the OLS estimator, which obtains by setting $w_{ni}^2 = 1$.

COROLLARY 1: *Under Assumptions* 1–2, *then*

$$\sqrt{n}\,(\hat{\beta}_n - \beta) \Rightarrow \frac{\int_0^1 \sigma v \, dW_0}{A \int_0^1 v^2}.$$

It is illustrative to point out that conventional asymptotic theory emerges as a special case:

COROLLARY 2: *Under Assumptions* 1–2, *if* $\eta_1 = \eta_2 = 0$, *then*

$$\sqrt{n}\left(\hat{\beta}_n - \beta\right) \Rightarrow N\left(0, \frac{\sigma^2}{A^2\nu^2}\right)$$

where $\sigma^2 = h_1(\omega_1)$ *and* $\nu^2 = h_2(\omega_2)$.

Corollary 1 gives the asymptotic distribution theory for OLS applied to traditional time series models, where the conditional variances are either constant or approximately so.

The covariance κ plays an important role in the distribution theory for the OLS estimator, as the Brownian motion W_0 is independent of W_1 and W_2 when (and only when) $\kappa = 0$. This occurs when $\kappa_k = 0$ for all $k \geq 1$. This moment is indeed zero when the innovations are symmetrically distributed. It also occurs when $E(\varepsilon_{ni} z_{ni} | \Im_{n,i-1}) = \mu_{ni}$ is nonrandom, for then

$$\begin{aligned} E(\xi_{n,i-k}\varepsilon_{ni}z_{ni}) &= E(\xi_{n,i-k}E(\varepsilon_{ni}z_{ni}|\Im_{n,i-1})) = E(\xi_{n,i-k}\mu_{ni}) \\ &= E(\xi_{n,i-k})\mu_{ni} = 0. \end{aligned}$$

The latter condition trivially holds when the martingale differences are independent over time. We state these observations formally.

ASSUMPTION 4: *One of the following conditions holds*: (i) $\{\xi_{n,i-k}, \varepsilon_{ni}, z_{ni}\}$ *is jointly symmetrically distributed*; *or* (ii) $E(\varepsilon_{ni} z_{ni} | \Im_{n,i-1}) = \mu_{ni}$ *a.s.*, *where* $\mu_{ni} < \infty$ *is nonrandom*; *or* (iii) $\{\xi_{ni}, \varepsilon_{ni}, z_{ni}\}$ *is independently distributed across* i.

LEMMA 1: *If Assumption* 4 *holds*, *then* $\kappa = 0$.

Under Assumption 4, the covariance matrix (12) is block diagonal, so the Brownian motion W_0 is independent of W_1 and W_2, and hence of $\sigma^2(s)$ and $\nu^2(s)$. It is well known that in this case the stochastic integral (7) has a variance mixture of normals distribution.

COROLLARY 3: *Under Assumptions* 1–2 *and* 4,

$$\sqrt{n}\left(\hat{\beta}_n - \beta\right) \Rightarrow \int N(0,V)\, dP(V), \tag{13}$$

where P is the probability measure of th random variance

$$V = \frac{\int_0^1 \sigma^2 \nu^2}{A^2\left(\int_0^1 \nu^2\right)^2}. \tag{14}$$

Without knowing the distribution function P, (13) does not give a complete characterization of the asymptotic distribution of $\hat{\beta}_n$, only showing that it is a variance mixture of normals. Even so, this simple result is quite useful as we can conclude that the large sample distribution of $\hat{\beta}_n$ is symmetric about β and unimodal. The most important point is that knowledge of P is not necessary for inference about β. This is because test statistics will have standard asymptotic distributions (either normal or chi-square depending on the form of the statistic). See Basawa and Scott (1983), Park and Phillips (1988), and Phillips (1991) for discussions of mixtures of normals distributions.

It is not difficult to see that the random covariance matrix given by (14) will (in general) not be properly estimated by the standard OLS covariance matrix. Instead, an appropriate estimator will take the Eicker-White form (see, e.g., Eicker (1963) and White (1980)). A notable exception arises when the limiting variance of the regression error is constant, that is, $\sigma^2(s) = \sigma^2$. In this case, the random variance simplifies and the standard OLS covariance matrix will be appropriate for inferential purposes.

3.3. *Generalized Least Squares*

It is well known that when the regression error is conditionally heteroskedastic, ordinary least squares is not efficient. A more efficient estimator will utilize the information in the conditional variance. The generalized least squares (GLS) estimator, for example, weights the data in inverse proportion to the square root of the conditional variance by setting $w_{ni}^2 = \sigma_{ni}^{-2}$. Since $\sigma_{ni}^2 \geq b > 0$, we see that $\sigma_{ni}^{-2} \leq b^{-1} < \infty$ so the conditions of Assumption 3 hold with $H = b^{-1}$ and $w^2(s) = \sigma^{-2}(s)$. Let $\tilde{\beta}_n^*$ denote this GLS estimator of β.

COROLLARY 4: *Under Assumptions* 1–3,

$$\sqrt{n}\left(\tilde{\beta}_n^* - \beta\right) \Rightarrow \frac{\int_0^1 \phi \, dW_0}{A \int_0^1 \phi^2}, \tag{15}$$

where $\phi(s) = \nu(s)/\sigma(s)$.

Corollary 4 shows that as a general rule, the GLS estimator does not necessarily possess a normal or mixture normal asymptotic distribution. It will, however, in two important special cases. The first is when the asymptotic random variances $\sigma^2(s)$ and $\nu^2(s)$ are proportional. (One may loosely think of this as "cointegration in the variance," although it is quite distinct from the co-persistence idea of Bollerslev and Engle (1993).) In this case the ratio $\phi(s) = \nu(s)/\sigma(s)$ is constant and the distribution of (15) simplifies to a standard normal.

COROLLARY 5: *Under Assumptions* 1–3, *if* $v^2(s) = \omega^2\sigma^2(s)$ *a.s. for some* $\omega^2 > 0$, *then*

$$\sqrt{n}\left(\tilde{\beta}_n^* - \beta\right) \to_d N\left(0, (\omega^2 A^2)^{-1}\right).$$

Another case in which the asymptotic distribution of the GLS estimator is a variance mixture of normals is under the symmetry/independence condition of Assumption 4.

COROLLARY 6: *Under Assumptions* 1–4,

$$\sqrt{n}\left(\tilde{\beta}_n^* - \beta\right) \Rightarrow \int N(0,V)\,dP(V)$$

where P *is the probability measure of the random variance* $V = (A^2 \int_0^1 \phi^2)^{-1}$.

4. ADAPTIVE ESTIMATION

The GLS estimator of the previous section is not feasible, since it requires the use of the unobservable volatility process σ_{ni}^2. It is useful to see whether we can construct a feasible estimator which achieves the same asymptotic distribution as the GLS estimator. Such an estimator is called adaptive.

One approach would be to assume parametric forms for the variance functions h_1 and h_2, and the distributions of the errors. Then quasi-maximum-likelihood estimation could be performed using the Kalman filter. This approach has been proposed and pursued by other authors. The advantages are efficient estimation, under the assumption that the parametric forms are correct. A main disadvantage is the need for the parametric assumptions. In this section I discuss a nonparametric method, only requiring that $\sigma^2(s)$ be continuous almost surely (as implied by Assumption 2).

4.1. *Nonparametric Variance Estimation*

Let $\hat{u}_{ni} = y_{ni} - \hat{\mu}_n - x_{n,i-1}\hat{\beta}_n$ be the OLS residual. Our idea is to estimate σ_{ni}^2 by averaging the squared OLS residuals $\hat{u}_{n,i-j}^2$ for small j. Specifically, we use a nonparametric kernel of the form

$$\hat{\sigma}_{ni}^2 = \frac{\sum_{j=0}^{N} k(j/N)\hat{u}_{n,i-j}^2}{\sum_{j=0}^{N} k(j/N)}, \qquad i \geq N,$$

and

$$\hat{\sigma}_{ni}^2 = \hat{\sigma}_{nN}^2, \qquad i < N,$$

where $k(\cdot)$: $[0,1] \to [0,1]$ is a standard kernel function satisfying $\int_0^1 k(x)\,dx > 0$. For example, the rectangular, Bartlett, and Parzen kernels all satisfy these conditions. The integer N is a bandwidth number, controlling the degree of local smoothing.

ASSUMPTION 5: *For* $p > 2$ *defined in Assumption* 1, (i) $E|\varepsilon_{ni}|^{2p} \leq Q_{\varepsilon\varepsilon} < \infty$; (ii) $N = Bn^{\psi}$ *for some* $0 < B < \infty$ *and* $\psi \in (2/p, 1)$.

The bandwidth N is required to grow at a rate slower than sample size, but not too slowly. If p is not much greater than 2, then ψ needs to be close to 1. As p increases, ψ can be smaller. Intuitively, a smaller ψ implies less smoothing, and an attempt to estimate the variance array σ_{ni}^2 at a higher resolution. This is more difficult without the presence of higher moments in which case more smoothing is required (and hence a larger ψ) to estimate the variance array uniformly.

It is important in practice that the estimated variance sequence be bounded away from zero. To ensure this we suggest a trimmed estimator:

$$\tilde{\sigma}_{ni}^2 = \max(\hat{\sigma}_{ni}^2, b^*), \tag{16}$$

where $0 < b^* \leq b$.

We have our second main result:

THEOREM 2: *Under Assumptions* 1–3 *and* 5, $\tilde{\sigma}_{n[ns]}^2 \Rightarrow \sigma^2(s)$.

Theorem 2 shows that consistent estimation in $C[0,1]$ (in the sense of weak convergence) of the asymptotic variance process is possible by a simple nonparametric kernel technique. This result is a consequence of the assumption that $\sigma^2(s)$ is almost surely continuous, so local averaging reveals the underlying variance process.

In principle, the trimming recommended in (16) is not necessary, for as $n \to \infty$ we find for any $\varepsilon > 0$

$$P\left\{\inf_{i \leq n} \hat{\sigma}_{ni}^2 \leq b - \varepsilon\right\} \to 0.$$

In practice, however, sampling variation may produce values of $\hat{\sigma}_{ni}^2$ which are sufficiently small to distort $\hat{\beta}_n$. Trimming is inherently conservative, as it reduces the impact of observations where $\hat{\sigma}_{ni}^2$ is unusually small. To implement (16) properly, prior knowledge of b is necessary. Since b is typically unknown, a feasible approach is to set b^* equal to some prespecified fraction of the full sample average, i.e., $b^* = q\hat{\sigma}^2$, where $\hat{\sigma}^2 = (1/n)\sum_{i=1}^n \hat{u}_{ni}^2$, and q is some number between 0 and 1, for example 0.1. A particular choice is unfortunately *ad hoc*, and it is hard to see any convincing argument for one choice over another.

4.2. *Adaptive Least Squares*

The estimated array $\tilde{\sigma}_{ni}^2$ is adapted to $\Im_{ni}$ (for $i \geq N$), so the conditions for Theorem 1 are applicable if we set $w_{ni} = \tilde{\sigma}_{ni}^{-2}$. We now define an adaptive least squares (ALS) estimator

$$\tilde{\beta}_n^a = \left(\sum_{i=1}^{n} \tilde{\sigma}_{ni}^{-2} x_{ni}^* x_{ni}^{*\prime} \right)^{-1} \left(\sum_{i=1}^{n} \tilde{\sigma}_{ni}^{-2} x_{ni}^* y_{n,i+1} \right).$$

THEOREM 3: *Under Assumptions* 1–3 *and* 5,

$$\sqrt{n}\left(\tilde{\beta}_n^a - \beta \right) \Rightarrow \frac{\int_0^1 \phi \, dW_0}{A \int_0^1 \phi^2}.$$

Theorem 3 shows that the ALS estimator achieves the same asymptotic distribution as the GLS estimator. It follows as well that asymptotic mixture normality of the ALS estimator holds under the conditions given in Corollaries 5 and 6.

5. CONCLUSIONS

It is probably the case that most applied time series analysts pay insufficient attention to the long-run properties of the second moments of their data. Some attempt to reduce the extent of residual heteroskedasticity by data transformation, but few pay any attention to the second moment properties of their regressors. The implicit assumption, of course, has been that such properties do not really matter. As the distributional theory of this paper shows, the long-run properties of the second moment properties of both the regression error and the regressors can matter for the large sample distribution of estimators. Under the regularity conditions of Assumption 4, we have shown that the asymptotic distributions are variance mixtures of normals, so standard testing procedures can be used. Unfortunately, it is unclear how one might attempt to verify the conditions of Assumption 4.

When the regression error has a conditional variance which displays long-run nonstationarity, OLS estimation is not efficient. Feasible GLS techniques are available, one of which is outlined here. The ALS estimator we propose is very simple to implement, does not require any parametric assumptions, and achieves the same asymptotic distribution as the theoretical GLS estimator. In the presence of nonstationary volatility, the use of the ALS estimator can lead to major gains in estimation efficiency over OLS.

Dept. of Economics, Boston College, Chestnut Hill, MA 02167-3806, U.S.A.

Manuscript received October, 1992; final revision received October, 1994.

APPENDIX

Since the analysis of the paper focuses on the slope coefficient β, and the regression (1) contains an intercept, we can without loss of generality set $\mu_x = 0$, which we shall do throughout this section. The proofs of the main theorems rely on a set of intermediate results which we now present. Our first two results show that stochastic processes which are asymptotically continuous have uniformly small increments.

LEMMA A.1: *If $X_{n[ns]} \Rightarrow X(s)$, and X has continuous sample paths almost surely, and $N \leq Bn^{\alpha}$ where $B < \infty$ and $0 < \alpha < 1$, then*

$$\max_{i \leq n,\, j \leq N} |X_{n,i+j} - X_{ni}| \to_p 0.$$

PROOF: Since $X(s)$ lies in $C[0,1]$, X_n converges weakly in the uniform metric, and therefore must be tight in that metric. This implies that for all $\varepsilon > 0$ and $\eta > 0$, there exists a $\delta > 0$ and an integer n_0 such that for all $n \geq n_0$,

$$P\left(\max_{i \leq n,\, j \leq [\delta n]} |X_{n,i+j} - X_{ni}| > \varepsilon \right) \leq \eta.$$

See, for example, Billingsley (1968, p. 55). For any $0 < \delta < 1$, however, we can find a n_1 sufficiently large such that for all $n \geq n_1$, $N \leq Bn^{\alpha} \leq [\delta n]$ (since $\alpha < 1$). Thus for all $n \geq \max(n_0, n_1)$,

$$\max_{i \leq n,\, j \leq N} |X_{n,i+j} - X_{ni}| \leq \max_{i \leq n,\, j \leq [\delta n]} |X_{n,i+j} - X_{ni}|$$

and thus

$$P\left(\max_{i \leq n,\, j \leq N} |X_{n,i+j} - X_{ni}| > \varepsilon \right) \leq P\left(\max_{i \leq n,\, j \leq [\delta n]} |X_{n,i+j} - X_{ni}| > \varepsilon \right) \leq \eta.$$

Since ε and η are arbitrary, the proof is complete. *Q.E.D.*

The following is a direct implication of Lemma A.1, taking $N = 1$ and $X_{ni} = g(Z_{ni})$.

LEMMA A.2: *If $Z_{n[ns]} \Rightarrow Z(s)$, and Z has continuous sample paths almost surely, then for any continuous function $g(\cdot)$,*

$$\max_{i \leq n} |g(Z_{ni}) - g(Z_{n,i-1})| \to_p 0.$$

Our next result derives asymptotic representations for the conditional variances σ_{ni}^2 and ν_{ni}^2, as well as the partial sums of the product processes $\xi_{n,i-k}\varepsilon_{n,i+1}$.

LEMMA A.3: *Under Assumptions* 1–3,

$$\sigma_{n[ns]}^2 \Rightarrow \sigma^2(s), \tag{17}$$

$$\nu_{n[ns]}^2 \Rightarrow \nu^2(s), \tag{18}$$

and for all $k \geq 0$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{[ns]} \xi_{n,i-k}\varepsilon_{n,i+1} \Rightarrow B_k(s), \tag{19}$$

where $\sigma^2(s)$ and $\nu^2(s)$ are defined by (8)–(11) *in the statement of Theorem* 1, *and the $\{B_k\}$ are mutually independent standard Brownian motions. Setting $W_0 = A^{-1}\sum_{k=0}^{\infty} a_k B_k$ and $W = (W_0\ W_1\ W_2)$, then* (12) *holds.*

PROOF: Since the innovations z_{ni} are uniformly square integrable martingale differences, under the local-to-unity structure of Assumption 3, it is well known that $(1/\sqrt{n})S_{n1[ns]} \Rightarrow W_1^c(s)$ and $(1/\sqrt{n})S_{n2[ns]} \Rightarrow W_2^c(s)$, where W_1^c and W_2^c are defined by (10)–(11). By the continuous mapping theorem,

$$\sigma_{n[ns]}^2 = h_1\left(\omega_1 + \frac{\eta_1}{\sqrt{n}} S_{n1[ns]}\right) \Rightarrow h_1(\omega_1 + \eta_1 W_1^c(s)) = \sigma^2(s),$$

establishing (17). Equation (18) follows similarly. Next, note that the product innovation $\xi_{n,i-k}\varepsilon_{n,i+1}$ is a martingale difference with unit conditional variance. Furthermore, by the law of iterated expectations and Assumption 1,

$$E|\xi_{n,i-k}\varepsilon_{n,i+1}|^p = E[|\xi_{n,i-k}|^p E(|\varepsilon_{n,i+1}|^p|\Im_{ni})] \le E|\xi_{n,i-k}|^p Q_\varepsilon \le Q_\xi Q_\varepsilon < \infty,$$

so $\xi_{n,i-k}\varepsilon_{n,i+1}$ is uniformly square integrable. (19) follows by the invariance principle for martingale difference arrays. To see that the Brownian motions B_k are mutually independent, it is sufficient to note that $\xi_{n,i-k}\varepsilon_{n,i+1}$ is uncorrelated with $\xi_{n,i-j}\varepsilon_{n,i+1}$ for $j \ne k$.

To establish (12), first note that the covariance matrix of $(W_1\ W_2')$ is the same as z_{ni}, which is Ω_z. Second, note that

$$E\left(\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \cdot W_0\right) = A^{-1} \sum_{k=0}^{\infty} a_k E\left(\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \cdot B_k\right)$$
$$= A^{-1} \sum_{k=0}^{\infty} a_k \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E(z_{n,i+1}\xi_{n,i-k}\varepsilon_{n,i+1}) = A^{-1} \sum_{k=0}^{\infty} a_k \kappa_k = \kappa.$$

Finally, $EW_0^2 = A^{-2}\sum_{k=0}^{\infty} a_k^2 = 1$. *Q.E.D.*

The next two results give large sample representations for the first two weighted sample moments of the regressor x_{ni}.

LEMMA A.4: *Under Assumptions* 1–3,

$$\frac{1}{n}\sum_{i=1}^{n} w_{ni}^2 x_{ni} \to_p 0.$$

PROOF: Note that

$$x_{ni} = \sum_{k=0}^{\infty} a_k \nu_{n,i-k-1}\xi_{n,i-k} = \nu_{ni}\xi_{ni}^* + \sum_{k=0}^{\infty} a_k(\nu_{n,i-k-1} - \nu_{ni})\xi_{n,i-k} \tag{20}$$

where $\xi_{ni}^* = \sum_{k=0}^{\infty} a_k \xi_{n,i-k}$. Since as $m \to \infty$,

$$\sup_{n,i} E|E(\xi_{ni}^*|\Im_{n,i-m})| \le \sum_{k=m}^{\infty} |a_k| \sup_{n,i} E|\xi_{ni}| \to 0,$$

and $w_{n[ns]}^2 \nu_{n[ns]}^2 \Rightarrow w^2(s)\nu^2(s)$ which is almost surely continuous, an application of Theorem 3.3 of Hansen (1992) yields

$$\frac{1}{n}\sum_{i=1}^{n} w_{ni}^2 \nu_{ni}\xi_{ni}^* \to_p 0. \tag{21}$$

Next, set $\xi_{ni}^{**} = \sum_{k=0}^{\infty} k|a_k\xi_{n,i-k}|$ which is uniformly integrable, so $(1/n)\sum_{i=1}^{n} w_{ni}^2\xi_{ni}^{**} = O_p(1)$. Thus

$$\left|\frac{1}{n}\sum_{i=1}^{n} w_{ni}^2 \sum_{k=0}^{\infty} a_k(\nu_{n,i-k-1} - \nu_{ni})\xi_{n,i-k}\right| \le \frac{1}{n}\sum_{i=1}^{n} w_{ni}^2 \xi_{ni}^{**} \max_i |\nu_{n,i-1} - \nu_{ni}| \to_p 0 \tag{22}$$

by Lemma A.2. Equations (20), (21), and (22) together establish the result. *Q.E.D.*

LEMMA A.5: *Under Assumptions* 1–3,

$$\frac{1}{n}\sum_{i=1}^{n} w_{ni}^2 x_{ni}^2 \Rightarrow A^2\int_0^1 w^2\nu^2.$$

PROOF: Expanding (2) we find the standard decomposition

$$\frac{1}{n}\sum_{i=1}^{n} w_{ni}^2 x_{ni}^2 = \frac{1}{n}\sum_{i=1}^{n} w_{ni}^2 f_{ni} + \frac{1}{n}\sum_{i=1}^{n} w_{ni}^2 g_{ni}, \tag{23}$$

where

$$f_{ni} = \sum_{k=0}^{\infty} a_k^2 v_{ni-k}^2,$$

and

$$g_{ni} = \sum_{k=1}^{\infty}\sum_{j=0}^{\infty} a_j a_{j+k} v_{n,i-j} v_{n,i-k-j}.$$

We examine each sum on the right-hand-side of (23) in turn.

A few algebraic manipulations yield the equality

$$f_{ni} = A^2\nu_{ni}^2 + \nu_{ni}^2 f_{ni}^a + f_{ni}^b + f_{ni}^c \tag{24}$$

where

$$f_{ni}^a = \sum_{k=0}^{\infty} a_k^2(\xi_{n,i-k}^2 - 1),$$

$$f_{ni}^b = \sum_{k=0}^{\infty} a_k^2(\nu_{n,i-k-1}^2 - \nu_{ni}^2),$$

and

$$f_{ni}^c = \sum_{k=0}^{\infty} a_k^2(\nu_{n,i-k-1}^2 - \nu_{ni}^2)(\xi_{n,i-k}^2 - 1).$$

First, (18), Assumption 3, and the continuous mapping theorem yield

$$\frac{1}{n}\sum_{i=1}^{n} w_{ni}^2 A^2\nu_{ni}^2 \Rightarrow A^2\int_0^1 w^2\nu^2. \tag{25}$$

Second,

$$\sup_{n,i} E|E(f_{ni}^a|\Im_{n,i-m})| = \sup_{n,i} E\left|\sum_{k=m}^{\infty} a_k^2(\xi_{n,i-k}^2 - 1)\right| \le 2\sum_{k=m}^{\infty} a_k^2 \to 0$$

as $m \to \infty$, and so by Theorem 3.3 of Hansen (1992),

$$\frac{1}{n}\sum_{i=1}^{n} w_{ni}^2\nu_{ni}^2 f_{ni}^a \to_p 0. \tag{26}$$

Third, by the triangle inequality, the facts that $(1/n)\Sigma_{i=1}^n w_{ni}^2 = O_p(1)$ and $\Sigma_{k=0}^{\infty} k a_k^2 < \infty$, and Lemma A.2,

$$\left|\frac{1}{n}\sum_{i=1}^{n} w_{ni}^2 f_{ni}^b\right| = \left|\frac{1}{n}\sum_{i=1}^{n} w_{ni}^2 \sum_{k=0}^{\infty} a_k^2(\nu_{n,i-k-1}^2 - \nu_{ni}^2)\right| \le \frac{1}{n}\sum_{i=1}^{n} w_{ni}^2 \sum_{k=0}^{\infty} k a_k^2 \max_{i\le n}|\nu_{ni}^2 - \nu_{n,i-1}^2| \to_p 0. \tag{27}$$

Fourth, letting $e_{ni} = \sum_{k=0}^{\infty} k a_k^2 |\xi_{n,i-k}^2 - 1|$ which is uniformly integrable and thus $(1/n)\sum_{i=1}^{n} w_{ni}^2 e_{ni} = O_p(1)$, hence

$$(28) \qquad \left| \frac{1}{n} \sum_{i=1}^{n} w_{ni}^2 f_{ni}^c \right| \le \frac{1}{n} \sum_{i=1}^{n} w_{ni}^2 e_{ni} \max_{i \le n} |\nu_{ni}^2 - \nu_{n,i-1}^2| \to_p 0$$

by Lemma A.2. Equations (24), (25), (26), (27), and (28) yield

$$(29) \qquad \frac{1}{n} \sum_{i=1}^{n} w_{ni}^2 f_{ni} \Rightarrow A^2 \int_0^1 w^2 \nu^2.$$

Next, we turn to g_{ni}. Algebraic manipulations show that

$$g_{ni} = \nu_{ni}^2 g_{ni}^a + \nu_{ni} g_{ni}^b + \nu_{ni} g_{ni}^c + g_{ni}^d,$$

where

$$g_{ni}^a = \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} a_j a_{j+k} \xi_{n,i-j} \xi_{n,i-k-j},$$

$$g_{ni}^b = \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} a_j a_{j+k} (\nu_{n,i-k-j-1} - \nu_{ni}) \xi_{n,i-j} \xi_{n,i-k-j},$$

$$g_{ni}^c = \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} a_j a_{j+k} (\nu_{n,i-j-1} - \nu_{ni}) \xi_{n,i-j} \xi_{n,i-k-j},$$

and

$$g_{ni}^d = \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} a_j a_{j+k} (\nu_{n,i-j-1} - \nu_{ni})(\nu_{n,i-k-j-1} - \nu_{ni}) \xi_{n,i-j} \xi_{n,i-k-j}.$$

First,

$$\sup_{n,i} E|E(g_{ni}^a | \Im_{n,i-m})| \le \sum_{k=1}^{\infty} |a_k| \sum_{j=m}^{\infty} |a_j| Q_\xi \to 0$$

as $m \to \infty$; by Theorem 3.3 of Hansen (1992) we find

$$\frac{1}{n} \sum_{i=1}^{n} w_{ni}^2 \nu_{ni}^2 g_{ni}^a \to_p 0.$$

Second, letting $\lambda_{ni} = \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} j |a_j a_{j+k} \xi_{n,i-j} \xi_{n,i-k-j}|$, which is uniformly integrable, we find

$$\left| \frac{1}{n} \sum_{i=1}^{n} w_{ni}^2 \nu_{ni} g_{ni}^b \right| \le \frac{1}{n} \sum_{i=1}^{n} w_{ni}^2 \nu_{ni} \lambda_{ni} \max_{i \le n} |\nu_{ni}^2 - \nu_{n,i-1}^2| \to_p 0.$$

Similarly, the sums over g_{ni}^c and g_{ni}^d are also asymptotically negligible. Together, we have shown that

$$(30) \qquad \frac{1}{n} \sum_{i=1}^{n} w_{ni}^2 g_{ni} \to_p 0.$$

Equations (23), (29), and (30) together establish the result. *Q.E.D.*

Our final preliminary result is probably the most important. It gives a stochastic integral asymptotic representation for the "numerator" of the WLS estimator.

LEMMA A.6: *Under Assumptions* 1–3,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{ni}^2 x_{ni} u_{n,i+1} \Rightarrow A \int_0^1 w^2 \sigma \nu \, dW_0.$$

PROOF: Set

$$N_{nk} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{ni}^2 v_{n,i-k} u_{n,i+1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{ni}^2 \nu_{n,i-k-1} \sigma_{ni} \xi_{n,i-k} \varepsilon_{n,i+1}.$$

Under (17), (18), and (19), by Theorem 2.1 of Hansen (1992) (which is a special case of Theorem 4.6 of Kurtz and Protter (1991)), for any $k \geq 0$,

$$N_{nk} \Rightarrow \int_0^1 w^2 \nu \sigma \, dB_k = N_k. \tag{31}$$

Set $N_n = (N_{n0}, N_{n1}, \ldots)$ and $N = (N_0, N_1, \ldots)$, both of which are elements of R^∞.

We want to show that $N_n \Rightarrow N$. This follows from the finite dimensional convergence implied by (31), if the sequence of probability measures is tight in R^∞ (see, for example, Theorem 6.2 and the preceding discussion in Billingsley (1968)). For $e \in R^\infty$, define the metric $d(e) = \sum_{k=0}^\infty k^{-r}|e_k|$. For any $\varepsilon > 0$, set K_k to be a compact real interval so that $P\{N_{nk} \in K_k\} \geq 1 - \varepsilon k^{-r}/\sum_{k=0}^\infty k^{-r}$ for all n sufficiently large, which is possible under (31). Set $K = K_0 \times K_1 \cdots$, which can be shown to be a compact subset of R^∞. Thus for n sufficiently large,

$$P\{N_n \notin K\} = P\left\{ \bigcup_{k=0}^{\infty} \{N_{nk} \notin K_k\} \right\} \leq \sum_{k=0}^{\infty} P\{N_{nk} \notin K_k\} \leq \sum_{k=0}^{\infty} \varepsilon k^{-r} \Big/ \sum_{k=0}^{\infty} k^{-r} = \varepsilon,$$

from which it follows that the probability measures of the sequence N_n are tight, and hence $N_n \Rightarrow N$. From the continuous mapping theorem it follows that $f(N_n) \Rightarrow f(N)$ for any d-continuous f.

Take the function $f(e) = \sum_{k=0}^\infty a_k e_k$. To show that it is d-continuous, take any $\varepsilon > 0$ and any $e^1 \in R^\infty$ and $e^2 \in R^\infty$ such that $d(e^1 - e^2) = \sum_{k=0}^\infty k^{-r}|e_k^1 - e_k^2| \leq \varepsilon$. Set $\delta = \bar{a}\varepsilon$. Then

$$|f(e^1) - f(e^2)| = \left| \sum_{k=0}^{\infty} a_k (e_k^1 - e_k^2) \right| \leq \sum_{k=0}^{\infty} \bar{a} k^{-r} |e_k^1 - e_k^2| = \bar{a} d(e^1 - e^2) \leq \delta,$$

as required for continuity. We therefore conclude with the desired result:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{ni}^2 x_{ni} u_{n,i+1} = \sum_{k=0}^{\infty} a_k N_{nk} \Rightarrow \sum_{k=0}^{\infty} a_k N_k = \int_0^1 w^2 \nu \sigma \sum_{k=0}^{\infty} a_k \, dB_k = A \int_0^1 w^2 \nu \sigma \, dW_0.$$

Q.E.D.

PROOF OF THEOREM 1: Note that

$$\frac{1}{n} \sum_{i=1}^{n} w_{ni}^2 \Rightarrow \int_0^1 w^2 > 0 \quad \text{a.s.}$$

by Assumption 4 and the continuous mapping theorem. Note as well that since $h_2(x) \geq b$, then $\nu^2(s) \leq b$, and

$$\int_0^1 w^2 \nu^2 \geq b^2 \int_0^1 w^2 > 0 \quad \text{a.s.}$$

Thus

$$\sqrt{n}(\tilde{\beta}_n - \beta) = \left(\frac{1}{n} \sum_{i=1}^{n} w_{ni}^2 x_{ni}^2 + o_p(1) \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{ni}^2 x_{ni} u_{n,i+1} + o_p(1) \right)$$

$$\Rightarrow \left(\int_0^1 w^2 \nu^2 \right)^{-1} A^{-1} \int_0^1 w^2 \sigma \nu \, dW_0$$

by Lemmas A.3, A.4, A.5, and A.6. *Q.E.D.*

PROOF OF THEOREM 2: Since $\sigma_{n[ns]}^2 \Rightarrow \sigma^2(s)$, it is sufficient to show that

$$\max_{N \le i \le n} |\hat{\sigma}_{ni}^2 - \sigma_{ni}^2| \to_p 0 \tag{32}$$

and

$$\max_{1 \le i < N} |\hat{\sigma}_{ni}^2 - \sigma_{ni}^2| \to_p 0. \tag{33}$$

Let

$$\bar{k}(j/N) = \frac{k(j/N)}{\sum_{j=0}^{N} k(j/N)},$$

so that for $i \ge N$,

$$\hat{\sigma}_{ni}^2 = \sum_{j=0}^{N} \bar{k}(j/N) \hat{u}_{n,i-j}^2.$$

Note that

$$\sum_{j=0}^{N} \bar{k}(j/N) = 1 \tag{34}$$

and

$$N \sum_{j=0}^{N} \bar{k}(j/N)^2 = \frac{\frac{1}{N} \sum_{j=0}^{N} k(j/N)^2}{\left(\frac{1}{N} \sum_{j=0}^{N} k(j/N)\right)^2} \to \frac{\int_0^1 k(x)^2\, dx}{\left(\int_0^1 k(x)\, dx\right)^2} \equiv K < \infty. \tag{35}$$

We start with (32). Note that

$$\hat{\sigma}_{ni}^2 - \sigma_{ni}^2 = R_{ni}^a + \sigma_{ni}^2 R_{ni}^b + R_{ni}^c + R_{ni}^d, \tag{36}$$

where

$$R_{ni}^a = \sum_{j=0}^{N} \bar{k}(j/N)(\sigma_{n,i-j-1}^2 - \sigma_{ni}^2),$$

$$R_{ni}^b = \sum_{j=0}^{N} \bar{k}(j/N)(\varepsilon_{n,i-j}^2 - 1),$$

$$R_{ni}^c = \sum_{j=0}^{N} \bar{k}(j/N)(\sigma_{n,i-j-1}^2 - \sigma_{ni}^2)(\varepsilon_{n,i-j}^2 - 1),$$

$$R_{ni}^d = \sum_{j=0}^{N} \bar{k}(j/N)(\hat{u}_{n,i-j}^2 - u_{n,i-j}^2).$$

First, observe that

$$\max_{N \le i \le n} |R_{ni}^a| \le \max_{i \le n,\, j \le N} |\sigma_{ni}^2 - \sigma_{n,i-j}^2| \to_p 0 \tag{37}$$

by (34) and Lemma A.1. Second, since $\{\varepsilon_{ni}^2 - 1, \Im_{ni}\}$ is a martingale difference, by Lemma 2 of Hansen (1991),

$$\|R_{ni}^b\|_p \leq P\left(\sum_{j=0}^{N} \bar{k}(j/N)^2 \|\varepsilon_{n,i-j}^2 - 1\|_p^2\right)^{1/2} \leq P(Q_{\varepsilon\varepsilon} + 1)K^{1/2}N^{-1/2}$$

where $P = 18p^{3/2}/\sqrt{p-1}$. Hence since $N = Bn^{\psi}$ and $\psi p/2 > 1$,

$$E \max_{N \leq i \leq n} |R_{ni}^b|^p \leq n \max_{N \leq i \leq n} E|R_{ni}^b|^p \leq P^p(Q_{\varepsilon\varepsilon} + 1)^p K^{p/2} B^{-p/2} n^{1-\psi p/2} \to 0 \tag{38}$$

as $n \to \infty$. By Markov's inequality, $\max_i |R_{ni}^b| = o_p(1)$. Since $\max_i \sigma_{ni}^2 = O_p(1)$, it follows that

$$\max_{N \leq i \leq n} |\sigma_{ni}^2 R_{ni}^b| \to_p 0. \tag{39}$$

Third,

$$\max_{N \leq i \leq n} |R_{ni}^c| \leq \max_{i \leq n,\, j \leq N} |\sigma_{ni}^2 - \sigma_{n,i-j}^2| \left(1 + \max_{N \leq i \leq n} \sum_{j=0}^{N} \bar{k}(j/N)\varepsilon_{n,i-j}^2\right) \tag{40}$$

$$\leq \max_{i \leq n,\, j \leq N} \left[|\sigma_{ni}^2 - \sigma_{n,i-j}^2|\left(2 + \max_{N \leq i \leq n} |R_{ni}^b|\right)\right] \to_p 0$$

by (37) and (38). Fourth, consider R_{ni}^d. To simplify the notation, consider the case without an intercept. Note that

$$\left|\sum_{j=0}^{N} \bar{k}(j/N)(\hat{u}_{n,i-j}^2 - u_{n,i-j}^2)\right|$$

$$\leq 2\left|\frac{1}{\sqrt{n}} \sum_{j=0}^{N} \bar{k}(j/N) u_{n,i-j} x_{n,i-j-1}\right| \sqrt{n}\,|\hat{\beta} - \beta| + \sqrt{n}\,|\hat{\beta} - \beta|^2$$

$$\times \left|\frac{1}{n} \sum_{j=0}^{N} \bar{k}(j/N) x_{n,i-j-1}^2\right|.$$

Now $\sqrt{n}\,|\hat{\beta} - \beta| = O_p(1)$ and $\max_i |1/n \Sigma_{j=0}^N \bar{k}(j/N) x_{n,i-j-1}^2| = O_p(N/n) = o_p(1)$. Further, since $\{\bar{k}(j/N) u_{n,i-j} x_{n,i-j}\}$ is a MDA, by a derivation similar to that used to show (39), we can show that $\max_i |(1/\sqrt{n}) \Sigma_{j=0}^N \bar{k}(j/N) u_{n,i-j} x'_{n,i-j}| \to_p 0$. It follows that

$$\max_{N \leq i \leq n} |R_{ni}^d| \to_p 0. \tag{41}$$

Equation (32) follows from (36), (37), (39), (40), and (41).

Finally, note that for $i \leq N$,

$$|\hat{\sigma}_{ni}^2 - \sigma_{ni}^2| = |\hat{\sigma}_{nN}^2 - \sigma_{ni}^2| \leq |\hat{\sigma}_{nN}^2 - \sigma_{nN}^2| + |\sigma_{ni}^2 - \sigma_{nN}^2|$$

so (33) follows from (32) and Lemma A.1, completing the proof. *Q.E.D.*

Proof of Theorem 3: As discussed in the text, the adjusted variance estimator satisfies $\tilde{\sigma}_{n[ns]}^2 \Rightarrow \sigma^2(s)$. The conditions of Theorem 1 are nearly applicable, except that the array $\tilde{\sigma}_{ni}^2$ is not adapted to $\Im_{ni}$ for $i < N$. A review of the proof shows that this is only used for the convergence of the numerator $(1/\sqrt{n})\Sigma_{i=1}^n \sigma_{ni}^{-2} x_{ni} u_{ni}$. The discrepancy only involves the term

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{N} \sigma_{ni}^{-2} x_{ni} u_{n,i+1}. \tag{42}$$

Fix some $\eta > 0$. We can then find a $\delta > 0$ such that $E|\int_0^\delta \phi\, dW_0| \leq \eta$. Yet for large enough n, $N = Bn^{\psi} < \delta n$. Thus

$$E\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{N}\sigma_{ni}^{-2}x_{ni}u_{n,i+1}\right| \leq E\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\delta n}\sigma_{ni}^{-2}x_{ni}u_{n,i+1}\right| \to E\left|\int_0^\delta \phi\, dW_0\right| \leq \eta.$$

Since η is arbitrary, the discrepancy term (42) is asymptotically negligible, as required. *Q.E.D.*

REFERENCES

ANDERSEN, T. G. (1993): "Stochastic Autoregressive Volatility: A Framework for Volatility Modeling," Kellogg Graduate School of Management Working Paper 161.

ANDERSON, T. W., AND J. B. TAYLOR (1979): "Strong Consistency of Least Squares Estimates in Dynamic Models," *The Annals of Statistics*, 7, 484–489.

BASAWA, I. V., AND D. J. SCOTT (1983): *Asymptotic Optimal Inference for Non-Ergodic Models*. New York: Springer-Verlag.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

BOLLERSLEV, T., AND R. F. ENGLE (1993): "Common Persistence in Conditional Variances," *Econometrica*, 61, 167–186.

CHESNEY, M., AND L. SCOTT (1989): "Pricing European Currency Options: A Comparison of the Modified Black-Scholes Model and a Random Variance Model," *Journal of Financial and Quantitative Analysis*, 24, 267–284.

DAVIDSON, J. (1992): "A Central Limit Theorem for Globally Nonstationary Near-Epoch Dependent Functions of Mixing Processes," *Econometric Theory*, 8, 313–329.

DELONG, J. B., AND L. H. SUMMERS (1986): "The Changing Cyclical Variability of Economic Activity in the United States," in *The American Business Cycle*, ed. by R. J. Gordon. Chicago: University of Chicago Press, pp. 679–734.

EICKER, F. (1963): "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions," *Annals of Mathematical Statistics*, 34, 447–456.

HANSEN, B. E. (1991): "Strong Laws for Dependent Heterogeneous Processes," *Econometric Theory*, 7, 213–221; and (1992) "Erratum," *Econometric Theory*, 8, 421–422.

——— (1992): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500.

HARVEY, A. C., AND P. C. ROBINSON (1988): "Efficient Estimation of Nonstationary Time Series Regression," *Journal of Time Series Analysis*, 9, 201–214.

HARVEY, A. C., E. RUIZ, AND N. SHEPHARD (1994): "Multivariate Stochastic Volatility Models," *Review of Economic Studies*, 61, 247–264.

HULL, J., AND A. WHITE (1987): "The Pricing of Options on Assets with Stochastic Volatilities," *The Journal of Finance*, 42, 281–300.

KURTZ, T. G., AND P. PROTTER (1991): "Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations," *The Annals of Probability*, 19, 1035–1070.

MELINO, A., AND S. M. TURNBULL (1990): "Pricing Foreign Currency Options with Stochastic Volatility," *Journal of Econometrics*, 45, 239–265.

NELSON, D. B. (1990): "ARCH Models as Diffusion Approximations," *Journal of Econometrics*, 45, 7–39.

NELSON, D. B., AND D. P. FOSTER (1994): "Asymptotic Filtering Theory for Univariate ARCH Models," *Econometrica*, 62, 1–41.

PAGAN, A. R., AND G. W. SCHWERT (1990a): "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics*, 45, 267–290.

——— (1990b): "Testing for Covariance Stationarity in Stock Market Data," *Economics Letters*, 33, 165–170.

PARK, J. Y., AND P. C. B. PHILLIPS (1988): "Statistical Inference in Regressions with Integrated Processes: Part I," *Econometric Theory*, 4, 468–497.

PHILLIPS, P. C. B. (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.

PHILLIPS, P. C. B., AND M. LORETAN (1990): "Testing Covariance Stationarity under Moment Condition Failure with an Application to Common Stock Returns," Cowles Foundation Discussion Paper No. 947.

ROBINSON, P. (1978): "On Consistency in Time Series Analysis," *The Annals of Statistics*, 6, 214–223.

SHEPHARD, N. (1994): "Local Scale Models: State Space Alternatives to Integrated GARCH Processes," *Journal of Econometrics*, 60, 181–202.
WHITE, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.
WIGGINS, J. B. (1987): "Option Values under Stochastic Volatility: Theory and Empirical Estimates," *Journal of Financial Economics*, 19, 351–372.
WOOLDRIDGE, J. M., AND H. WHITE (1988): "Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes," *Econometric Theory*, 4, 210–230.