Rochester Center for
Economic Research

Regression with Non-Stationary Variances

Hansen, Bruce E.

Working Paper No. 331
September 1992

University of
Rochester

Regression With Non-Stationary Variances

Bruce E. Hansen

Rochester Center for Economic Research
Working Paper No. 331

Regression With Non-Stationary Variances

Bruce E. Hansen*

Department of Economics

University of Rochester

September 1992

Abstract

A completely new asymptotic theory of regression is introduced for possibly non-stationary time series. The variables are assumed to be generated by a vector linear process with martingale difference innovations. The conditional variances of these martingale differences are allowed to be non-stationary processes. The primary requirement imposed on these processes is that they converge weakly in the Skorohod metric to cadlag stochastic processes. The types of non-stationary variances thereby permitted include deterministic variances, multiple structural breaks with random shift points, and positive functions of integrated or near-integrated processes.

*A very preliminary version of this research was circulated as "Regression Theory When Variances Are Non-Stationary," RCER Working Paper No. 226. This research was supported by the NSF.

It contrast to much of the existing literature which attempts to generalize the assumptions of the classic regression model, we find that, in general, the asymptotic distributions of the ordinary least squares (OLS) and generalized least squares (GLS) estimators are not normal, nor even mixtures of normals, unless the asymptotic variance processes are exogeneous in a specific sense. Under the latter assumption, however, GLS estimation is efficient within the class of weighted least squares. An adaptive estimator is proposed, based on local averaging of the squared OLS residuals, which is asymptotically equivalent to GLS.

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 auto covariances. Casual examination of plots and recursive of rolling estimates of variances for many series, however, suggests non-constancy. 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 non-stationarity 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 non-stationarity in the variance.

To handle the difficult concept of non-stationary variances, we work with linear processes with martingale difference innovations. We allow for non-stationarity by assuming that the conditional covariance matrix of the martingale differences can be approximated as a *cadlag* stochastic process. This allows (as special cases) for constant variances, single or multiple structural breaks, polynomial or sinusiodal deterministic trends, and positive functions of Brownian motions or Gaussian diffusion processes.

Under this assumption, a large sample theory of inference can be derived. Consider the ordinary least squares (OLS) estimator. We can show that the coefficient vector is consistently estimated. The asymtotic distribution, however, is only multivariate normal if the limiting variance processes are deterministic which is the assumption commonly found in the current literature, such as Wooldridge and White (1988), Harvey and Robinson (1988) and Davidson (1992). In general, when the limiting variance processes are stochastic, the asymptotic distribution can be represented by a ratio of random variables. The distributions bear a striking resemblance to those obtained in the unit root literature, as the numerator of the distribution has a stochastic integral representation, and the denominator is an integral over a stochastic process. In the special case in which the asymptotic variance processes are independent of the stationary part of the variables, the asymptotic distribution is a random mixture of normals, so inference can proceed conventionally. Without this assumption, however, the asymptotic distribution of test statistics is non-standard.

Since we are allowing for the variance of the regression error to be time-varying, it makes sense to consider generalized least squares (GLS) estimation as well. In particular, we propose an adaptive GLS estimator which uses conditional variance estimates constructed by local averaging of the squared OLS residuals. This estimator process is shown to be uniformly consistent for the limiting variance process, and the adaptive GLS estimator is asymptotically equivalent to the GLS estimator using the true conditional variance sequence.

Section 2 introduces the model and assumptions. Section 3 examines OLS estimation. Section 4 examines GLS estimation. Section 5 contains the proofs of all lemmas and theorems.

Throughout the paper $|\cdot|$ refers to the Euclidean norm $|A| = (tr(A'A))^{1/2}$, $\|\cdot\|_p$ to the L_p- norm $\|A\|_p = (E|A|^p)^{1/p}$, $[\cdot]$ refers to integer part, vec(A) is the vector operator which stacks the columns of the matrix A, and $\Rightarrow$ denotes weak convergence with respect to the Skorohod metric. All limits are taken as the sample size, n, diverges to positive infinity.

2 Model and Assumptions

2.1 Regression Model

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

$$y_{ni} = x_{ni}'\beta + u_{ni}.$$

For some array of sigma-fields $\{\Im_{ni} : 1 \leq i \leq n\}$ to which x_{ni} and u_{ni+1} are adapted, we assume that u_{ni} is a martingale difference array:

$$E\left(u_{ni} \mid \Im_{ni}\right) = 0$$

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

$$x_{ni} = \sum_{k=0}^{\infty} A_j v_{ni-j}, \quad E\left(v_{ni} \mid \Im_{ni-1}\right) = 0$$

where A_0 is normalized to be the $m \times m$ identity matrix I_m, and the coefficients satisfy the summability condition

$$\sum_{j=0}^{\infty} j \left|A_j\right| < \infty. \tag{1}$$

The regressor is well defined under the following moment condition.

Assumption 1 *For some* $p > 2$, $\sup_{n\geq 1} \sup_{i\leq n} \|v_{ni}\|_p^2 \leq C < \infty$.

These conditions describe a fairly typical set of assumptions in linear time series analysis.

2.2 Conditional Variances

Define the conditional variances:

$$E\left(u_{ni}^2 \mid \Im_{ni}\right) = \sigma_{ni}^2$$

and

$$E\left(v_{ni}v_{ni}' \mid \Im_{ni-1}\right) = \Omega_{ni}.$$

We will frequently desire to separate out the variance part of these arrays. This can be accomplished by defining the standardized arrays

$$z_{ni} = u_{ni}/\sigma_{ni} \tag{2}$$

and

$$e_{ni} = D_{ni}^{-1}v_{ni} \tag{3}$$

where D_{ni} is the upper triangular square root matrix of Ω_{ni}. By construction, z_{ni} is a martingale difference array with unit conditional variance, and e_{ni} is a vector martingale difference array with conditional covariance matrix I_m. Equations (2) and (3) yield the equivalent expressions $u_{ni} = \sigma_{ni}z_{ni}$ and $v_{ni} = D_{ni}e_{ni}$ which give the desired separation.

It turns out that the asymptotic limit theory depends upon the following $m^2 \times m$ random array. Set

$$\Phi_{ni} = \sum_{k=0}^{\infty} e_{ni-k}z_{ni} \otimes A_k',$$

and its associated partial sum process

$$B_{ni} = \frac{1}{\sqrt{n}}\sum_{j=1}^{i} \Phi_{nj},$$

and set

$$\phi_{ni} = \operatorname{vec}\Phi_{ni} = \sum_{k=0}^{\infty} e_{ni-k} z_{ni} \otimes a_k,$$

where $a_k = \operatorname{vec}(A_k')$. Note that $\{\phi_{ni}, \Im_{ni}\}$ is a martingale difference array, and for each $k \geq 0$, $E\left(e_{ni-k}e_{ni-k}' z_{ni}^2\right) = E\left(e_{ni-k}e_{ni-k}' E\left(z_{ni}^2 \mid \Im_{ni}\right)\right) = E\left(e_{ni-k}e_{ni-k}'\right) = I_m$. Thus ϕ_{ni} has the degenerate covariance matrix

$$V_\phi = E\sum_{k=0}^{\infty} e_{ni-k} z_{ni}^2 e_{ni-k}' \otimes a_k a_k' = I_m \otimes \sum_{k=0}^{\infty} a_k a_k'.$$

If $|\phi_{ni}|^2$ is uniformly integrable, then it is well known that $B_{n[nr]} \Rightarrow B(r)$ where $\operatorname{vec} B(r)$ is a Brownian motion with covariance matrix V_ϕ. We assume the latter directly in assumption 2.

The purpose of the present paper is to study distributional theory when the conditional variance is "non-stationary." This is a rather vague concept which needs to be made more precise. We will use the following condition.

Assumption 2

$$\left(B_{n[nr]},\ \sigma_{n[nr]}^2,\ \Omega_{n[nr]}\right) \Rightarrow \left(B(r), \sigma^2(r),\ \Omega(r)\right)$$

where $B(r)$ is a vector Brownian motion with covariance matrix V_ϕ, $\sigma^2(\cdot)$ is an element of $D[0,1]$, and $\Omega(\cdot)$ is an element of $D[0,1]^{m\times m}$.

This condition directly assumes that the conditional variances can be approximated (for large n) by a random element on the space of *cadlag* matrix functions. This doesn't seem like a very strong assumption, and captures fairly neatly the basic idea of non-stationarity. The fact that Ω_{ni} doesn't need to be scaled by a function of n is with no loss of generality, since we are directly working with random arrays. (This turns out to be particularly convenient for most examples we find of interest).

Note that assumption 2 includes constant conditional variances as a special case.

Note as well the parellels with the standard theory of non-stationarity. If a variable $x_i = i$ is a deterministic trend, when normalized its conditional expectation (trivially) is approximated by the linear function r. If x_i is I(1), its conditional expectation is cointegrated with it, and both can be approximated by a vector Brownian motion. If x_{ni} is a "near-integrated" array, its conditional expectation is approximated by a Gaussian diffusion process. Assumption 2 allows for the conditional variance (as opposed to the conditional mean) to be approximated by a cadlag process.

Assumption 2 allows for either the conditional variance of the regression error or the conditional variance of the regressors to be non-stationary, or both. This allows for a range of possible applications. Some researchers may believe with high confidence that their regression errors will not display non-stationarity in the conditional variance, because they have scrutinized the time-series properties of their residuals to eliminate such possibilities. In this case, they would likely assume that the limiting process $\sigma^2(r)$ is a constant σ^2. The same researcher, though, may not be confident that the conditional variance of the regressors is stationary. This mix is allowed under assumption 2 as an interesting special case.

The assumption requires that the three arrays converge jointly. This is frequently easy to verify when the processes are asymptotically continuous, so long as the variables are defined on the same probability space. When the variance processes are asymptotically discontinuous, then the joint convergence requirement is more restrictive. Note that no assumption has been made regarding the relationship of $B(r), \sigma^2(r)$, and $\Omega(r)$. Thus the processes may be independent or interdependent. As we discuss later, this has important implications for inference.

Our final three conditions are required for idenfication and other technicalities:

Assumption 3 $M = \int_0^1 \Omega(s)ds > 0$.

Assumption 4 *For some* $q > 2$, $\sup_{n\geq 1} \sup_{i\leq n} \left\|\sigma^2_{ni}\right\|_q < \infty$.

Assumption 5 $\sup_{1\leq i\leq n} \sqrt{n} \left\|D_{ni} - D_{ni-1}\right\|_4 = O(1)$.

We discuss particular examples which satisfy assumptions 2-5 in the following section.

2.3 Examples

The conditional variances σ^2_{ni} and Ω_{ni} can be constructed as non-negative definite functions of random arrays with weak cadlag limits. That is, if the $s \times 1$ vector S_{ni} is an array such that $S_{n[nr]} \Rightarrow S(r)$ for some process $S(r)$, and $\Omega_{ni} = \Omega(S_{ni})$ where $\Omega(\cdot)$ is a mapping from R^s to $M^{m\times m}$, the space of positive semi-definate $m \times m$ matrices, then Ω_{ni} is a valid conditional covariance matrix and $\Omega_{n[nr]} \Rightarrow \Omega(S(r))$. Similarly for the scalar σ^2_{ni}. The array S_{ni} can be any standard example used in the literature on empirical processes. Possible candidates include deterministic functions such as polynomials and sinusoids. (For a linear trend, set $S_{ni} = i/n$.) Structural change is permitted by functions such as

$$S_{ni} = \begin{cases} S_1, & i < n\tau \\ S_2, & i \geq n\tau \end{cases},$$

where the timing of structural change τ is a random variable in $[0,1]$.

Stochastic nonstationarity can be allowed by using functions of partial sum processes. Take a sequence of iid random variable ξ_i and set $S_{ni} = \frac{1}{\sqrt{n}}\sum_{j=1}^{i}\xi_j$, so that $S_{n[nr]} \Rightarrow S(r)$, a Brownian motion with variance $E(\xi_i)^2$. A useful functional for the mapping $\sigma^2(S)$ in this context is $\sigma^2(S) = (a+S)^2$. Thus $\sigma_{ni} = (a+S_{ni})^2$ and the limit variance process is $\sigma^2(r) = (a+S(r))^2$ which is bounded above zero, a useful property for a variance process.

The random walk model assumes that shocks to the variance process are "persistent."A process with less persistence than a random walk can be achieved by setting

$$S_{ni} = (1 - c/n)S_{ni-1} + \frac{1}{\sqrt{n}}\xi_i. \tag{4}$$

Here, $S_{n[nr]} \Rightarrow S(r)$, a diffusion process which satisfies the stochastic differential equation $dS = -cS + dW$, where W is a Brownian motion. This process is a continuous-time approximation to a first-order autoregressive process, and is used in the literature on near-integration. Again, setting $\sigma_{ni}^2 = (a+S_{ni})^2$ bounds the conditional variance above zero.

To construct a martingale difference with the given conditional variance simply requires the construction of a mixture random variable. For example, let z_{ni} and e_{ni} be iid random variables, and set $v_{ni} = D_{ni}e_{ni}$ and $u_{ni} = \sigma_{ni}z_{ni}$, where D_{ni} is the upper triangular square root of Ω_{ni}. Then set $\Im_{ni}$ to be the smallest sigma-field containing the past history of $(z_{ni}, e_{ni}, D_{ni}, \sigma_{ni})$. Then the arrays $\{v_{ni}, \Im_{ni}\}$ and $\{u_{ni}, \Im_{ni}\}$ are martingale differences with conditional covariances Ω_{ni} and σ_{ni}^2, respectively. The latter satisfy the convergence of assumption 2, and the moment requirements can be satisfied directly by assumption on the iid innovations.

Recently, Nelson and Foster (1992) have argued that ARCD models may be viewed as useful filters to estimate conditional variances when the data are discrete measurements from an underlying continuous-time diffusion process. In fact, their model is that the observations and conditional variance are generated by a *joint* diffusion process, which is the continuous-time analog of the specification in (4). The asymptotic theory in their paper is derived as the time interval goes to zero, which is analogous to our near-integration specification. While not strictly nested, their models is quite close to ours, although their purpose is to investigate the properties of ARCH models as optimal filters, not to investigate the properties of classical regression methods as in this paper.

Before we turn to the anaysis of regression methods, we need to examine the unusual condition given in assumption 5. We can show that this condition is valid for several examples of interest.

First, suppose that D_{ni} is a function of a polynomial in time, so that $D_{ni} = f\left(\frac{i}{n}\right)$ for some uniformly differentiable f. Then

$$\sqrt{n}\max_{i\leq n}|D_{ni} - D_{ni-1}| = \sqrt{n}\max_{i\leq n} f\left|\left(\frac{i}{n}\right) - f\left(\frac{i-1}{n}\right)\right| \leq \frac{1}{\sqrt{n}}\sup_{0\leq r\leq 1}\frac{df(x)}{dx} \to 0.$$

Second, suppose that D_{ni} arises from a structural shift, so $D_{ni} = D_1 +$

$D_2\{i \geq n\tau\}$ where $\{\}$ denotes the indicator function, and $\tau \in [0,1]$ is a random variable which denotes the timing of the structural shift. In this cas,

$$\sqrt{n}\max_{i\leq n}\|D_{ni} - D_{ni-1}\|_4 = \Delta D_2\sqrt{n}\max_{i\leq n} P\{i = [n\tau]\} \to 0$$

if the random variable τ has a density function on $[0,1]$.

Third, take that case of D_{ni} being a near-integrated partial sum process. Then

$$\sqrt{n}\max_{i\leq n}\|D_{ni} - D_{ni-1}\|_4 \leq \frac{c}{\sqrt{n}}\max_{i\leq n}\|S_{ni-1}\|_4 + \max_{i\leq n}\|\xi_i\|_4 = O(1)$$

if $\max_{i\leq n}\|\xi_i\|_4 = O(1)$.

3 Ordinary Least Squares

The ordinary least squares (OLS) estimate of the regression parameter β is given by

$$\hat{\beta}_n = \left(\sum_{i=1}^{n} x_{ni}x'_{ni}\right)^{-1} \left(\sum_{i=1}^{n} x_{ni}y_{ni}\right).$$

Centered and standardized, this equals

$$\sqrt{n}\left(\hat{\beta}_n - \beta\right) = \left(\frac{1}{n}\sum_{i=1}^{n} x_{ni}x'_{ni}\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_{ni}u_{ni}\right).$$

3.1 Numerator

Our first result is for the numerator of the OLS estimator. Define $D(s)$ as the upper triangular square root matrix of $\Omega(s)$, and $\sigma(s)$ as the positive square root of $\sigma^2(s)$. Here and elsewhere, let $X^{-}(r)$ denote the left limit of X at r.

Theorem 1 $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_{ni}u_{ni} \Rightarrow G = \int_0^1 dB(r)' \operatorname{vec} D(r)^{-} \sigma(r)^{-}.$

The limit random variable G has a distribution which can be represented as a stochastic integral with respect to a matrix Brownian motion. When $D(s)$ and $\sigma(s)$ are deterministic functions, G specializes to a normal random vector. This is essentially the situation studied by the previous literature allowing for non-stationarity in the variance, such as Wooldridge and White (1988), Harvey and Robinson (1988), and Davidson (1992). When either $D(\cdot)$ or $\sigma^2(\cdot)$ is a non-degenerate random process, however, G has a non-normal distribution. This is a new result and will be discussed in more detail in the section 3.3.

3.2 Denominator

The derivation of the limiting representation for $\frac{1}{n}\sum_{i=1}^{n} x_{ni}x_{ni}'$ is based on the decomposition

$$x_{ni}x_{ni}' = M_{ni}^{a} + M_{ni}^{b} + M_{ni}^{b\prime} \tag{5}$$

where

$$M_{ni}^{a} = \sum_{j=0}^{\infty} A_j v_{ni-j}v_{ni-j}' A_j' \tag{6}$$

and

$$M_{ni}^{b} = \sum_{k=1}^{\infty}\sum_{j=0}^{\infty} A_j v_{ni-j}v_{ni-j-k}' A_{j+k}'. \tag{7}$$

This decomposition was introduced by Phillips and Solo (1992) for the analysis of scalar linear processes (with asymptotically constant variances).

Lemma 1 $\frac{1}{n}\sum_{i=1}^{n} M_{ni}^{a} \Rightarrow J = \sum_{j=0}^{\infty} A_j M A_j'$, *where* $M = \int_0^1 \Omega(s)ds$.

Lemma 2 $\frac{1}{n}\sum_{i=1}^{n} M_{ni}^{b} \to_p 0$.

These two lemmas immediately yield our desired result.

Theorem 2 $\frac{1}{n}\sum_{i=1}^{n} x_{ni}x'_{ni} \Rightarrow J(r)$.

Theorem 2 gives the asymptotic distribution of the regression design matrix. Note that the limit variate J is function only of the random matrix M and the constant matrix $\sum_{s=0}^{\infty}(A_s \otimes A_s)$, since

$$\text{vec}(J) = \sum_{j=0}^{\infty} \text{vec}\left(A_j M A_j'\right) = \sum_{s=0}^{\infty}(A_s \otimes A_s)\,\text{vec}(M).$$

3.3 Asymptotic Distribution

We can now establish a general theorem for the OLS estimator.

Theorem 3 $\sqrt{n}\left(\hat{\beta}_n - \beta\right) \Rightarrow J^{-1}G$.

The least squares estimator converges to the ratio of random varaibles discussed in the previous sections. The distribution given in Theorem 3 resembles those found in the literature on unit roots and cointegration, as the numerator is a stochastic integral with respect to a Brownian motion, and the denominator is an integral of a stochastic process.

The nature of the distribution in Theorem 3 will depend upon the relations between the limiting variance processes $(\sigma^2(\cdot), \Omega(\cdot))$ and the limiting partial sum process $B(\cdot)$. We start our analysis with the relatively simple case in which these two processes are independent. Set $\Im_{\Omega\sigma} = \sigma(\Omega(s), \sigma^2(s) : 0 \leq s \leq 1)$, the sigma-field generated by the limiting variance processes.

Theorem 4 *If $B(\cdot)$ is independent of $\Im_{\Omega\sigma}$, then*

$$\sqrt{n}\left(\hat{\beta}_n - \beta\right) \Rightarrow J^{-1}G = \int N\left(0, J^{-1}SJ^{-1}\right) dP\left(J, S\right),$$

where $P(.,.)$ *is the probability measure over the joint distribution of* J *and*

$$S = \sum_{k=0}^{\infty} A_k \int_0^r \Omega(s)\sigma^2(s)dsA_k'.$$

Corollary 1 *If* $\sigma^2(r)$ *and* $\Omega(r)$ *have degenerate probability distributions (that is, are deterministic functions), then* J *and* S *are constant matrices and*

$$\sqrt{n}\left(\hat{\beta}_n - \beta\right) \Rightarrow N\left(0, J^{-1}SJ^{-1}\right).$$

Corollary 2 *If* $\sigma^2(r) = \sigma^2$ *and* $\Omega(r) = \Omega$ *are constants, then* $J = \sum_{k=0}^{\infty} A_k\Omega A_k$ *and* $S = \sigma^2 J$ *so*

$$\sqrt{n}\left(\hat{\beta}_n - \beta\right) \Rightarrow N\left(0, J^{-1}\sigma^2\right).$$

Corollary 2 gives the asymptotic distribution theory for the conventional regression model. The parameters estimates are asymptotically normal with a covariance matrix of conventional form. Corollary 2 gives the asymptotic distribution theory when the conditional variances are allowed to asymptotically non-stationary, but deterministic. The asymptotic distribution of the OLS estimator is again multivariate normal. This case is analogous to the models studied by other authors, such as Wooldridge and White (1988) and Davidson (1992). The implicit message in results such as Corollary 2 seems to have been that non-stationarity in the variance does not matter for inference in large samples. This conclusion is based, however, on the important restriction to asymptotically deterministic variance processes.

Corollary 2 is a specialization of Theorem 4. In the latter, the asymptotic variance processes are allowed to be stochastic, but are required to be independent of the asymptotic partial sum process B. In this special case, the asymptotic distribution $\int N\left(0, J^{-1}SJ^{-1}\right) dP\left(J, S\right)$ is a variance mixtures of normals, similar to the distributional results obtained in the cointegration literature. Mixtures of normals are essentially normal distributions with a random covariance matrix. If the random covariance matrix is known, then confidence intervals with the correct asymptotic coverage probabilities can be constructed. The fact that the covariance matrix is random means that the amount of information in a sample is random, and so will be different in repeated samples.

The assumption of independence in Theorem 4 is quite strong. We do not expect this condition to hold, in general. Without a complete probability model for the data, it is hard to know whether or not it is reasonable. It is not a vacuous assumption since it is easy to construct processes which satisfy this condition, such as the examples of section 2.3.

In the most general case, where the asymptotic variance processes are stochastic and not necessarily independent of the process B, the limiting distribution of the OLS estimator given in Theorem 3 is non-standard. The divergence from the normal distribution will depend critically on the extent to which the processes $(\sigma^2(\cdot), \Omega(\cdot))$ are correlated with $B(\cdot)$.

3.4 Conditionally Homoskedastic Regression Error

It is common for researchers to pay close attention to the time-series properties of their regression residuals. When there is evidence of serial correlation or heteroskedasticity, most will attempt to correct the problem through an appropriate data transformation. As a result, it might be expected that the likelihood of the variance of the regression error, σ_{ni}^2, being non-stationary is quite low, even if the regressors x_{ni} have time-varying variances. In this case, $\sigma_{n[nr]}^2 \Rightarrow \sigma^2$, a constant, and the limiting distribution of Theorem 4 specializes to $\int N\left(0, \sigma^2 J^{-1}\right) dP\left(J\right)$. This is still a mixtures-of-normals distribution, but the conditional covariance matrix is of a simpler form, as expected. In this case, the natural estimate of the covariance matrix for calculation of test statistics and confidence intervals is given by

$$\hat{J}^{-1}\hat{\sigma}^2$$

where

$$\hat{J} = \frac{1}{n}\sum_{i=1}^{n} x_{ni}x_{ni}'$$

and

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_{ni}^2$$

with $\hat{u}_{ni}^2$ being the least squares residuals, $\hat{u}_{ni} = y_{ni} - x_{ni}'\hat{\beta}_n$.

3.5 Covariance Matrix Estimation

In the general case in which the regression error is conditionally heteroskedastic, the conventional covariance matrix estimate will not be appropriate for calculation of test statistics and construction of confidence intervals. An appropriate estimator will be of the Eicker-White form, that is

$$\hat{J}^{-1}\hat{S}\hat{J}^{-1}$$

where $\hat{J}$ is defined above and

$$\hat{S} = \frac{1}{n} \sum_{i=1}^{n} x_{ni} x'_{ni} \hat{u}_{ni}^2.$$

We know (Theorem 2) that $\hat{J} \Rightarrow J$, and our hope is that $\hat{S} \Rightarrow S$.

As a first step it is natural to analyze the idealized estimator

$$\tilde{S} = \frac{1}{n} \sum_{i=1}^{n} x_{ni} x'_{ni} u_{ni}^2$$

and demonstrate that $\tilde{S} \Rightarrow S$. The second step of course would be to demonstrate that $\hat{S} - \tilde{S} \rightarrow_p 0$. Unfortunately, the proof for just the first step appears to be quite complicated and lengthy. Rather than burden the present paper with this derivation, we now turn to the more interesting problem of efficient estimation.

4 Adaptive Estimation

4.1 Weighted 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. Since the conditional variance is not observed, however, an estimate must be used in its place.

We first consider the general problem of weighted least squares regression, for a fairly arbitrary array of real-valued weights, $\{h_{ni}\}$ with asymptotically continuous trajectories.

Theorem 5 *Define the weighted least squares estimator*

$$\hat{\beta}_n^h = \left(\sum_{i=1}^{n} h_{ni}^{-1} x_{ni} x_{ni}' \right)^{-1} \left(\sum_{i=1}^{n} h_{ni}^{-1} x_{ni} y_{ni} \right).$$

If h_{ni} is adapted to $\Im_{ni}$, $h_{ni} \geq c > 0$. and $h_{n[nr]} \Rightarrow h(r) \in C[0,1]$, where the convergence is joint with the processes $\Omega_{ni}, \sigma_{ni}^2$, and B_{ni}, then

$$\sqrt{n}\left(\hat{\beta}_n^h - \beta\right) \Rightarrow J_h^{-1} G_h$$

where

$$J_h = \sum_{k=0}^{\infty} A_k \int_0^1 h(r)^{-1} \Omega(r) dr A_k'$$

and

$$G_h = \int_0^1 h(r)^{-1} dB(r)' \operatorname{vec}\left(D(r)^-\right) \sigma(r)^-.$$

4.2 Generalized Least Squares

If it were possible to directly observe the array $\{\sigma_{ni}^2\}$, then one could obtain the GLS estimator by setting $h_{ni} = \sigma_{ni}^2$ in the definition of the weighted least squares estimator. Even though it is unlikely that σ_{ni}^2 could be directly observed, we study this case as a precursor to adaptive estimation. Theorem 5 gives the asymptotic distribution of the GLS estimator when σ_{ni}^2 is asymptotically continuous and bounded above zero. We have

$$\hat{\beta}_n^\sigma = \left(\sum_{i=1}^n \sigma_{ni}^{-2} x_{ni} x_{ni}'\right)^{-1} \left(\sum_{i=1}^n \sigma_{ni}^{-2} x_{ni} y_{ni}\right)$$

and

$$\sqrt{n}\left(\hat{\beta}_n^\sigma - \beta\right) \Rightarrow J_\sigma^{-1} G_\sigma$$

where

$$J_\sigma = \sum_{k=0}^\infty A_k \int_0^1 \sigma(r)^{-2} \Omega(r) dr A_k'$$

and

$$G_\sigma = \int_0^1 \sigma(r)^{-1} dB(r)' \operatorname{vec}\left(D(r)^-\right).$$

When $B(\cdot)$ is independent of $\Im_{\Omega\sigma}$, this distribution is the mixture of normals given by

$$\int N(0, J_\sigma^{-1}) dP(J_\sigma)$$

where $P(\cdot)$ is the probability measure for the random design matrix J_σ. Note that the inverse of the design matrix is the asymptotic conditional precision matrix, hence the estimator $\hat{\beta}_n^\sigma$ is asymptotically efficient within the class of weighted least squares regression. Asymptotic standard errors can be calculated using the matrix

$$\hat{J}_\sigma^{-1} = \left(\frac{1}{n} \sum_{i=1}^{n} \sigma_{ni}^{-2} x_{ni} x_{ni}' \right)^{-1} .$$

It is important to note, however, that the asymptotic efficiency of the GLS estimator relies on the assumption that the process $B(\cdot)$ is independent of $\Im_{\Omega\sigma}$. This is the same requirement needed for asymptotic mixture normality of the OLS estimator. The endogeneity of the conditional variance of either the regressors or the regression error is sufficient to invalidate asymptotic mixture normality. Reweighting cannot solve this problem.

4.3 Non-Parametric Variance Estimation

If the variance of the regression error is asymptotically continuous in $C[0,1]$, it should be possible to consistently estimate it using a non-parametric estimator. While it may be possible to consistently estimate variance processes which are not necessarily asymptotically continuous, we will exclude this case from consideration as this would require a more complex estimation technique and asymptotic theory.

Assumption 6 *$\sigma^2(\cdot) \in C[0,1]$, $\sigma^2(r) \geq c$.*

Our idea is to estimate σ^2_{ni} by averaging the regression residuals $\hat{u}^2_{ni-j}$ for small j. Specifically, we use a non-parametric kernel of the form

$$\hat{\sigma}^2_{ni} = \sum_{j=0}^{N} w_{nj} \hat{u}^2_{ni-j} \ , \ i \geq N$$

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

where the weights $w_{nj} > 0$ satisfy

$$\sum_{j=0}^{N} w_{nj} = 1 \text{ and } N \sum_{j=0}^{N} w^2_{nj} = O(1).$$

The weights w_{nj} can be any typical kernel weights (normalized to sum to unity), such as from the Bartlett or Parzen kernels. One simple choice would be a rectangular kernel, where $w_{nj} = 1/N$. The integer N is a bandwidth number, and controls the degree of local smoothing. We require that the bandwidth number grow like a power of sample size:

Assumption 7 *$N = Bn^a$ for some $0 < B < \infty$ and $\frac{2}{p} < \alpha < 1$, where p is defined in assumption 1.*

We also require the following moment bound:

Assumption 8

$$\max_{n \geq 1} \max_{i \leq n} \ E\left(|u_{ni}|^{2p} \right) < \infty.$$

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 α needs to be close to 1. If u_{ni} has more finite moments, then α can be smaller. Intuitively, a smaller α implies less smoothing, and an attempt to estimate the variance array σ^2_{ni} at a higer resolution. This is more difficult without the presence of higher moments in which case more smoothing is required (and hence a larger α) to estimate the variance array uniformly.

We have:

Theorem 6 $\hat{\sigma}^2_{n[nr]} \Rightarrow \sigma^2(r)$.

Theorem 6 shows that consistent estimation in $C[0,1]$ (in the sense of weak convergence) of the asymptotic variance process is possible by a simple non-parametric kernel technique. This result is a consequence of the assumption that $\sigma^2(r)$ is continuous, so local averaging can reveal the underlying variance process. As a practical matter, it should be obvious upon reflection that the theorems are a bit more optimistic than should perhaps be warranted. The estimated process $\hat{\sigma}^2_{ni}$ will tend to be more smooth than the true conditional variance. The estimates will reveal the long-run trends in the conditional variance, but cannot hope to uncover high-frequency movements, in the style of ARCH or GARCH estimation.

4.4 Adaptive Least Squares

Our definition of $\hat{\sigma}^2_{ni}$ uses a one-sided backward-looking estimator. Consistent estimates could also be obtained by using a two-sided estimator, or a one-sided forward-looking estimator. The use of a backward-looking estimator, however, is particularly convenient for use in a weighted least squares context. The estimate array $\hat{\sigma}^2_{ni}$ is adapted to $\Im_{ni}$, so the conditions for Theorem 5 are nearly applicable. We need to be concerned, however, about the possibility that a sample realization of $\hat{\sigma}^2_{ni}$ might be too close to zero. We therefore suggest using the following trimmed version:

$$\tilde{\sigma}^2_{ni} = \max(\hat{\sigma}^2_{ni}, c).$$

Since $\hat{\sigma}^2_{n[nr]}$ converges weakly to $\sigma^2(r)$, and we have assumed that $\sigma^2(r) \geq c$, $\tilde{\sigma}^2_{n[nr]}$ will converge weakly to $\sigma^2(r)$ as well.

We can now define the adaptive least squares estimator.

$$\hat{\beta}^a_n = \left(\sum_{i=1}^{n} \tilde{\sigma}^{-2}_{ni} x_{ni} x'_{ni} \right)^{-1} \left(\sum_{i=1}^{n} \tilde{\sigma}^{-2}_{ni} x_{ni} y_{ni} \right).$$

Theorem 7 $\sqrt{n} \left(\hat{\beta}^a_n - \beta \right) \Rightarrow J^{-1}_\sigma G_\sigma$.

Theorem 7 shows that the adaptive estimator achieves the same asymptotic distribution as the generalized least squares estimator. We give special cases in the following results.

Corollary 3 *If $B(\cdot)$ is independent of $\Im_{\Omega\sigma}$, then*

$$J^{-1}_\sigma G_\sigma = \int N\left(0, J^{-1}_\sigma\right) dP\left(J_\sigma\right),$$

where $P(.)$ is the probability measure over the distribution of J_σ.

Corollary 4 *If $\sigma^2(r)$ and $\Omega(r)$ have degenerate probability distributions (that is, are deterministic functions), then J_σ is a constant matrix and*

$$J^{-1}_\sigma G_\sigma = N\left(0, J^{-1}_\sigma\right).$$

The general properties of the asymptotic distributions are quite similar to those obtained for the least squares estimator in section 3.3. When, and only when, the asymptotic variance processes are deterministic, the limiting distribution of the adaptive GLS estimator is multivariate normal. Mixture normality is obtained under the strong assumption of independence of B from $\Im_{\Omega\sigma}$, which we cannot expect to generally hold. In the general case, the limiting distribution is not a mixture of normals, even though the estimator is asymptotically equivalent to the true GLS estimator. This lack of normality occurs if either the variance of the regression error or the covariance matrix of the regressor innovations is endogeneously related to the partial sum process B_{ni}.

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, however, the long-run properties of the second moment properties of both the regression error and the regressors matter for the large sample distribution of estimators. If the asymptotic variance processes are not exogeneous in the particular sense made precise in section 3.3, the asymptotic distributions of the OLS and GLS estimators are not mixtures of normals, so conventional inference procedures are not justified.

The major lesson of this paper, therefore, is that empirical researchers should pay more attention to these properties of their data. It is not exactly clear how this should be done, but visual examination of time series plots of the data might be a useful first step.

When the regression error has a conditional variance which displays long-run non-stationarity, OLS estimation is not efficient. Feasible GLS techinques are available, one of which is outlined here, which allow for the conditional variance to be approximated by a weighted average of the squared residuals. Such techniques do not require explicit modeling of the variance process but can lead to potentially major gains in estimation efficiency.

Unfortunately, conventional inference procedures rely upon asymptotic mixture normality, and this only arises in the restrictive case in which the asymptotic variance processes are exogeneous to the stationary part of the variables. Generalized least squares techniques cannot eliminate this problem. Methods to deal with this situation are currently unknown, and would be an interesting subject for future research.

6 Mathematical Proofs

Proof of Theorem 1: We first show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\sum_{k=0}^{\infty}A_k\left(D_{ni}-D_{ni-k}\right)e_{ni-k}\right)z_{ni}\sigma_{ni}\rightarrow_p 0, \tag{8}$$

which holds if $\left(\sum_{k=0}^{\infty}A_k\sqrt{n}\left(D_{ni}-D_{ni-k}\right)e_{ni-k}\right)z_{ni}\sigma_{ni}$ is uniformly integrable. (See, for instance, Theorem 2.2 of Hall and Heyde (1980)). Indeed, by Minkowski's inequality, double application of Holder's inequality, the fact that $\left\|e_{ni-k}z_{ni}\right\|_2=1$, and assumptions 4 and 5,

$$\begin{aligned}
&\left\|\left(\sum_{k=0}^{\infty}A_k\sqrt{n}\left(D_{ni}-D_{ni-k}\right)e_{ni-k}\right)z_{ni}\sigma_{ni}\right\|_{4q/(3q+2)} \\
&\leq\sum_{k=0}^{\infty}\left|A_k\right|\sum_{j=1}^{k}\left\|\sqrt{n}\Delta D_{ni-j}e_{ni-k}z_{ni}\sigma_{ni}\right\|_{4q/(3q+2)} \\
&\leq\sum_{k=0}^{\infty}\left|A_k\right|\sum_{j=1}^{k}\sqrt{n}\left\|\Delta D_{ni-j}\sigma_{ni}\right\|_{4q/(q+2)}\left\|e_{ni-k}z_{ni}\right\|_2 \\
&\leq\sum_{k=0}^{\infty}k\left|A_k\right|\sqrt{n}\max_{i\leq n}\left\|\Delta D_{ni}\right\|_4\max_{i\leq n}\left\|\sigma_{ni}\right\|_{2q}<\infty,
\end{aligned} \tag{9}$$

uniformly in i. Since $4q/(3q+2)>1$ under assumption 4, the array is uniformly integrable and (8) is established.

Second, by Theorem 2.1 of Hansen (1992), which is a special case of Theorem 4.6 of Kurtz and Protter (1991),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi_i'\operatorname{vec}(D_{ni})\sigma_{ni}\Rightarrow\int_0^1 dB(s)'\operatorname{vec}(D(s))^{-}\sigma(s)^{-}. \tag{10}$$

Combining (8) and (10), we find

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_{ni}u_{ni} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=0}^{\infty} A_k D_{ni-k} e_{ni-k} z_{ni}\sigma_{ni}$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=0}^{\infty} A_k D_{ni} e_{ni-k} z_{ni}\sigma_{ni} + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=0}^{\infty} \left(z_{ni}e'_{ni-k} \otimes A_k\right) \operatorname{vec}(D_{ni})\sigma_{ni} + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \Phi'_{ni} \operatorname{vec}(D_{ni})\sigma_{ni} + o_p(1) \Rightarrow \int_0^1 dB(s)' \operatorname{vec}(D(s))^- \sigma(s)^-. \ \square$$

Proof of Lemma 1: Expression (6) for M^a_{ni} is not particularly amenable to linear analysis, but the work becomes easier by vectorizing the expression:

$$m^a_{ni} = \operatorname{vec}(M^a_{ni}) = \sum_{j=0}^{\infty}(A_j \otimes A_j) \operatorname{vec}(v_{ni-j}v'_{ni-j}) = A^0(L) \operatorname{vec}(v_{ni}v'_{ni}) \quad (11)$$

where

$$A^0(L) = \sum_{j=0}^{\infty}(A_j \otimes A_j)L^j.$$

$A^0(L)$ is a standard linear operator, to which we can apply the Beveridge-Nelson decomposition:

$$A^0(L) = A^0(1) - (1-L)\tilde{A}^0(L)$$

where

$$\tilde{A}^0(L) = \sum_{j=0}^{\infty} \tilde{A}_j^0 L^j$$

and

$$\tilde{A}_j^0 = \sum_{s=j+1}^{\infty} (A_s \otimes A_s).$$

For a recent exposition of the Beveridge-Nelson decomposition, see Phillips and Solo (1992). We can rewrite (11) as

$$m_{ni}^a = A^0(1)\,\mathrm{vec}(v_{ni}v_{ni}') - (1-L)w_{ni}, \tag{12}$$

where

$$w_{ni} = \tilde{A}^0(L)\,\mathrm{vec}(v_{ni}v_{ni}'). \tag{13}$$

Summing over (12) we obtain

$$\frac{1}{n}\sum_{i=1}^{n} m_{ni}^a = \frac{1}{n}\sum_{i=1}^{n} A^0(1)\,\mathrm{vec}(v_{ni}v_{ni}') + \frac{1}{n}w_{n0} - \frac{1}{n}w_{nn}. \tag{14}$$

By the triangle inequality, uniformly in i,

$$E\,|w_{ni}| = E\left|\sum_{j=0}^{\infty} \tilde{A}_j^0\,\mathrm{vec}(v_{ni-j}v_{ni-j}')\right| \leq \sum_{j=0}^{\infty} \left|\tilde{A}_j^0\right| E\left|v_{ni-j}v_{ni-j}'\right|$$

$$\leq C\sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty} |A_s \otimes A_s| \leq C\sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty} |A_s|^2 \leq C\sum_{j=0}^{\infty} j\,|A_j|^2 < \infty, \tag{15}$$

where C is defined in assumption 1. Thus by Markov's inequality

$$\frac{1}{n}\,|w_{n0} - w_{nn}| \to_p 0. \tag{16}$$

Further,

$$\frac{1}{n}\sum_{i=1}^{n} v_{ni}v_{ni}' = \frac{1}{n}\sum_{i=1}^{n}(v_{ni}v_{ni}' - \Omega_{ni}) + \frac{1}{n}\sum_{i=1}^{n}\Omega_{ni} \Rightarrow \int_0^1 \Omega(s)ds = M$$

since $\{v_{ni}v_{ni}' - \Omega_{ni}, \Im_{ni}\}$ is a uniformly integrable martingale difference array. Together with (14) and (16) this yields

$$\frac{1}{n}\sum_{i=1}^{n} m_{ni}^{a} = \frac{1}{n}\sum_{i=1}^{n} A^0(1)\,\mathrm{vec}(v_{ni}v_{ni}') + o_p(1) \Rightarrow A^0(1)\,\mathrm{vec}\,(M) = \mathrm{vec}\left(\sum_{j=0}^{\infty} A_j M A_j'\right)$$

which completes the proof. □

Proof of Lemma 2: Vectorizing M_{ni}^{b}, we obtain

$$m_{ni}^{b} = \mathrm{vec}\,M_{ni}^{b} = \sum_{k=1}^{\infty}\sum_{j=0}^{\infty}(A_{j+k}\otimes A_j)(v_{ni-j-k}\otimes v_{ni-j}) = \sum_{k=1}^{\infty} A^k(L)(v_{i-j-k}\otimes v_{i-j})$$

where

$$A^k(L) = \sum_{j=0}^{\infty}(A_{j+k}\otimes A_j)L^j.$$

A Beveridge-Nelson decomposition on $A^k(L)$ for each k yields

$$A^k(L) = A^k(1) - (1-L)\tilde{A}^k(L) \tag{17}$$

where

$$\tilde{A}^k(L) = \sum_{j=0}^{\infty}\tilde{A}_j^k L^j,$$

and

$$\tilde{A}_j^k = \sum_{s=j+1}^{\infty} (A_{k+s} \otimes A_s).$$

Applying the decomposed filter (17) to m_{ni}^b yields

$$m_{ni}^b = \eta_{ni} - (1-L)\lambda_{ni} \tag{18}$$

where

$$\eta_{ni} = \sum_{k=1}^{\infty} A^k(1)(v_{ni-k} \otimes v_{ni})$$

and

$$\lambda_{ni} = \sum_{k=1}^{\infty} \tilde{A}^k(L)(v_{i-k} \otimes v_i).$$

Summing over (18) we obtain

$$\frac{1}{n}\sum_{i=1}^{n} m_{ni}^b = \frac{1}{n}\sum_{i=1}^{n} \eta_{ni} + \frac{1}{n}\lambda_{n0} - \frac{1}{n}\lambda_{nn}. \tag{19}$$

First, note by the triangle inequality, for all i,

$$\begin{aligned} E\left|\lambda_{ni}\right| &= E\left|\sum_{k=1}^{\infty} \tilde{A}^k(L)(v_{ni-k} \otimes v_{ni})\right| \\ &\leq \sum_{k=1}^{\infty}\sum_{j=0}^{\infty} \left|\tilde{A}_j^k\right| E\left|v_{ni-k-j} \otimes v_{ni-j}\right| \\ &\leq \sum_{k=1}^{\infty}\sum_{j=0}^{\infty}\sum_{s=j+1}^{\infty} \left|A_{k+s} \otimes A_s\right| \left\|v_{ni-k-j}\right\|_2 \left\|v_{ni-j}\right\|_2, \\ &\leq C\sum_{k=1}^{\infty}\sum_{j=0}^{\infty} j\left|A_j\right|\left|A_{j+k}\right| \leq C\left(\sum_{j=0}^{\infty} j\left|A_j\right|\right)\left(\sum_{k=1}^{\infty}\left|A_k\right|\right) < \infty. \end{aligned}$$

Thus by Markov's inequality,

$$\frac{1}{n}\left|\lambda_{n0}+\lambda_{nn}\right| \rightarrow_p 0. \tag{20}$$

Second, by Minkowski's inequality,

$$\left\|\eta_{ni}\right\|_{p/2} \leq \sum_{k=1}^{\infty}\left|A^k(1)\right|\left\|v_{ni-k}\otimes v_{ni}\right\|_{p/2} \leq C\sum_{k=1}^{\infty}\sum_{j=0}^{\infty}\left|A_{j+k}\right|\left|A_j\right|, \tag{21}$$

$$\leq C\left(\sum_{j=0}^{\infty}\left|A_j\right|\right)^2 < \infty$$

and thus $|\eta_{ni}|$ is uniformly integrable (since $p/2>1$). Since $\{\eta_{ni}, \Im_{ni}\}$ is also a martingale difference array, this implies

$$\frac{1}{n}\sum_{i=1}^{n}\eta_{ni} \rightarrow_p 0. \tag{22}$$

(19), (20) and (22) combine to yield the desired result. □

Proof of Theorem 2: Summing over (5), we obtain

$$\frac{1}{n}\sum_{i=1}^{n}x_{ni}x_{ni}' = \frac{1}{n}\sum_{i=1}^{n}M_{ni}^{a} + \frac{1}{n}\sum_{i=1}^{n}M_{ni}^{b} + \frac{1}{n}\sum_{i=1}^{n}M_{ni}^{b\prime}.$$

The result is immediate from Lemmas 1 and 2. □

Proof of Theorem 3. Theorems 1, 2, assumption 3 and the continuous mapping theorem complete the proof. □

Proof of Theorem 4. We can rewrite the matrix process B as

$$B(r) = \sum_{k=0}^{\infty} B_k(r) \otimes A_k'$$

where the $B_k(\cdot)$ are independent Brownian motions each with covariance matrix I_m. We then have

$$\int_0^1 dB(r)' \operatorname{vec}(D(r)^-)\sigma(r)^- = \int_0^1 \sum_{k=0}^{\infty} (dB_k(r)' \otimes A_k) \operatorname{vec}(D(r)^-)\sigma(r)^-$$

$$= \sum_{k=0}^{\infty} A_k \int_0^1 D(r)^- \sigma(r)^- dB_k(r). \tag{23}$$

For each k, conditional upon $\Im_{\Omega\sigma}$, the vectors $\int_0^1 D(r)^- \sigma^-(r) dB_k(r)$ are independently normally distributed with mean zero and covariance matrix $\int_o^1 D^-(s)D^-(s)(\sigma^-(s))^2 ds = \int_0^1 \Omega(s)\sigma^2(s)ds$. Thus the sum in (23) is conditionally normal with mean zero and conditional covariance matrix $\sum_{k=0}^{\infty} A_k \int_0^1 \Omega\sigma^2 A_k' = S$. Hence the limit variate $J^{-1}G$ is conditionally normal with covariance matrix $J^{-1}SJ^{-1}$. Unconditionally, the distribution is therefore a mixtures of normals over this random covariance matrix. □

The following results will be useful in the proofs of Theorems 5 and 6.

Lemma 3 . *If* $X_{n[nr]} \Rightarrow X(r) \in C[0,1]$, *and* $N = Bn^{\alpha}$ *where* $B < \infty$ *and* $0 < \alpha < 1$, *then*

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

Proof of Lemma 3. Since $X(r)$ 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 $\epsilon > 0$ and $\eta > 0$, there exists a $\delta > 0$ and an integer n_0 such that for all $n \geq n_0$,

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

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

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

and thus

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

Since ϵ and η are arbitrary, the proof is complete. □

Lemma 4 *If $X_{n[nr]} \Rightarrow X(r) \in C[0,1]$, and $X_{ni} \geq c > 0$, then*

$$\max_{i\leq n}\left|\frac{1}{X_{ni}} - \frac{1}{X_{ni-1}}\right| \to_p 0.$$

Proof of Lemma 4:

$$\max_{i\leq n}\left|\frac{1}{X_{ni}} - \frac{1}{X_{ni-1}}\right| = \max_{i\leq n}\left|\frac{X_{ni} - X_{ni-1}}{X_{ni-1}X_{ni}}\right| \leq c^{-2}\max_{i\leq n}|X_{ni} - X_{ni-1}| \to_p 0$$

by Lemma 3. □

Proof of Theorem 5.

Step 1: $\frac{1}{n}\sum_{i=1}^{n} h_{ni}^{-1} x_{ni} x_{ni}' \Rightarrow J_h$. The proof follows those of Lemmas 1 and 2. Using the notation defined in their proofs,

$$\frac{1}{n}\sum_{i=1}^{n} h_{ni}^{-1} x_{ni} x_{ni}' = \frac{1}{n}\sum_{i=1}^{n} h_{ni}^{-1} M_{ni}^{a} + \frac{1}{n}\sum_{i=1}^{n} h_{ni}^{-1} M_{ni}^{b} + \frac{1}{n}\sum_{i=1}^{n} h_{ni}^{-1} M_{ni}^{b\prime}.$$

Using equation (12),

$$\begin{aligned} \operatorname{vec}\left(\frac{1}{n}\sum_{i=1}^{n} h_{ni}^{-1} M_{ni}^{a}\right) &= \frac{1}{n}\sum_{i=1}^{n} h_{ni}^{-1} m_{ni}^{a} \\ = \frac{1}{n}\sum_{i=1}^{n} h_{ni}^{-1} A^{0}(1) \operatorname{vec}(v_{ni} v_{ni}') &- \frac{1}{n}\sum_{i=1}^{n} h_{ni}^{-1}(1-L) w_{ni} \\ = A^{0}(1)\operatorname{vec}\left(\frac{1}{n}\sum_{i=1}^{n} h_{ni}^{-1} v_{ni} v_{ni}'\right) + \frac{1}{n}\sum_{i=1}^{n} \Delta\left(h_{ni}^{-1}\right) w_{ni} &+ \frac{1}{n} h_{n1}^{-1} w_{n0} - \frac{1}{n} h_{nn}^{-1} w_{nn}, \end{aligned} \tag{24}$$

where $\Delta\left(h_{ni}^{-1}\right) = \frac{1}{h_{ni}} - \frac{1}{h_{ni-1}}$ and w_{ni} is defined in (13). We now examine each term on the right-hand-side of (24).

First,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} h_{ni}^{-1} v_{ni} v_{ni}' &= \frac{1}{n}\sum_{i=1}^{n} h_{ni}^{-1}\Omega_{ni} - \frac{1}{n}\sum_{i=1}^{n} h_{ni}^{-1}\left(\Omega_{ni} - v_{ni} v_{ni}'\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} h_{ni}^{-1}\Omega_{ni} + o_p(1) \Rightarrow \int_0^1 h(r)^{-1}\Omega(r)dr \end{aligned} \tag{25}$$

where the second equality uses the fact that $\{h_{ni}^{-1}(\Omega_{ni} - v_{ni} v_{ni}'), \Im_{ni}\}$ is a uniformly integrable MDA.

Second,

$$\left|\frac{1}{n}\sum_{i=1}^{n} \Delta\left(h_{ni}^{-1}\right) w_{ni}\right| \le \max_{i \le n}\left|\frac{1}{h_{ni}} - \frac{1}{h_{ni-1}}\right| \frac{1}{n}\sum_{i=1}^{n} |w_{ni}| \to_p 0 \tag{26}$$

by Lemma 4 and the fact that

$$E\frac{1}{n}\sum_{i=1}^{n}|w_{ni}| \leq \max_{i\leq n} E\,|w_{ni}| < \infty$$

by (15).

Third,

$$\left|\frac{1}{n}h_{n1}^{-1}w_{n0} - \frac{1}{n}h_{nn}^{-1}w_{nn}\right| \leq \frac{1}{c}\,\frac{1}{n}\left(|w_{n0}| + |w_{nn}|\right) \to_p 0 \tag{27}$$

since $E\left(|w_{n0}| + |w_{nn}|\right)$ is uniformly bounded by (15). Equations (24), (25), (26), and (27) together yield

$$\begin{aligned}
\operatorname{vec}\left(\frac{1}{n}\sum_{i=1}^{n}h_{ni}^{-1}M_{ni}^{a}\right) &\Rightarrow A^0(1)\operatorname{vec}\left(\int_0^1 h(r)^{-1}\Omega(r)dr\right) \\
&= \sum_{k=0}^{\infty}(A_k\otimes A_k)\operatorname{vec}\left(\int_0^1 h(r)^{-1}\Omega(r)dr\right) \\
&= \operatorname{vec}\left(\sum_{k=0}^{\infty}A_k\left(\int_0^1 h^{-1}\Omega\right)A_k'\right) = \operatorname{vec}(J_h)\,.
\end{aligned}$$

It remains to show that $\frac{1}{n}\sum_{i=1}^{n}h_{ni}^{-1}M_{ni}^{b}$ is asymptotically negligible. Using (18) we find

$$\begin{aligned}
\operatorname{vec}\left(\frac{1}{n}\sum_{i=1}^{n}h_{ni}^{-1}M_{ni}^{b}\right) &= \frac{1}{n}\sum_{i=1}^{n}h_{ni}^{-1}m_{ni}^{b} = \frac{1}{n}\sum_{i=1}^{n}h_{ni}^{-1}\left(\eta_{ni} - \Delta\lambda_{ni}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}h_{ni}^{-1}\eta_{ni} + \frac{1}{n}\sum_{i=1}^{n}\Delta h_{ni}^{-1}\lambda_{ni} + \frac{1}{n}\left(h_{n1}^{-1}\lambda_{n0} - h_{nn+1}^{-1}\lambda_{nn}\right).
\end{aligned} \tag{28}$$

First, since $\{\eta_{ni},\Im_{ni}\}$ is a uniformly integrable MDA (as shown in (21)) so is $\{h_{ni}^{-1}\eta_{ni},\Im_{ni}\}$, and thus

$$\frac{1}{n}\sum_{i=1}^{n}h_{ni}^{-1}\eta_{ni} \to_p 0. \tag{29}$$

Second,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\Delta h_{ni}^{-1}\lambda_{ni}\right| \leq \max_{i\leq n}\left|\frac{1}{h_{ni}}-\frac{1}{h_{ni-1}}\right| \frac{1}{n}\sum_{i=1}^{n}|\lambda_{ni}|$$

$$\leq c^{-2}\max_{i\leq n}|h_{ni}-h_{ni-1}| \cdot O(1)\rightarrow_p 0 \qquad (30)$$

by Lemma 4 and the fact that

$$E\frac{1}{n}\sum_{i=1}^{n}|\lambda_{ni}| \leq \max_{i\leq n}E|\lambda_{ni}|<\infty$$

as shown by (21).

Third,

$$\left|\frac{1}{n}\left(h_{n1}^{-1}\lambda_{n0}-h_{nn+1}^{-1}\lambda_{nn}\right)\right| \leq \frac{1}{n}c^{-1}\left(|\lambda_{n0}|+|\lambda_{nn}|\right)\rightarrow_p 0 \qquad (31)$$

by (21). Equations (28), (29), (30), and (31) together demonstrate that $\frac{1}{n}\sum_{i=1}^{n}h_{ni}^{-1}M_{ni}^{b}\rightarrow_p 0$. This completes the first step.

Step 2: $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}h_{ni}^{-1}x_{ni}u_{ni}\Rightarrow G_h$. The proof follows that of Theorem 1. First, note that

$$\left\|h_{ni}^{-1}\left(\sum_{k=0}^{\infty}A_k\sqrt{n}\left(D_{ni}-D_{ni-k}\right)e_{ni-k}\right)z_{ni}\sigma_{ni}\right\|_{4q/(3q+2)}$$

$$\leq c^{-1}\left\|\left(\sum_{k=0}^{\infty}A_k\sqrt{n}\left(D_{ni}-D_{ni-k}\right)e_{ni-k}\right)z_{ni}\sigma_{ni}\right\|_{4q/(3q+2)}<\infty$$

by (9). Thus this array is uniformly integrable and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}h_{ni}^{-1}\left(\sum_{k=0}^{\infty}A_k\left(D_{ni}-D_{ni-k}\right)e_{ni-k}\right)z_{ni}\sigma_{ni}\rightarrow_p 0, \qquad (32)$$

Therefore

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h_{ni}^{-1} x_{ni} u_{ni} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=0}^{\infty} A_k h_{ni}^{-1} D_{ni} e_{ni-k} z_{ni} \sigma_{ni} + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h_{ni}^{-1} \Phi'_{ni} \operatorname{vec}(D_{ni})\sigma_{ni} + o_p(1) \Rightarrow \int_0^1 h(s)^{-1} dB(s)' \operatorname{vec}(D(s))^{-}\sigma(s)^{-}.$$

The Theorem follows from steps 1 and 2 and the continuous mapping theorem. □

Proof of Theorem 6. We start by showing that

$$\max_{N\le i\le n}\left|\sum_{j=0}^{N} w_{nj}\left(u_{ni-j}^2 - \sigma_{ni-j}^2\right)\right| \to_p 0. \tag{33}$$

By Burkholder's inequality for martingale differences (see Hall and Heyde (1980), p. 23),

$$\max_{N\le i\le n} E\left|\sum_{j=0}^{N} w_{nj}\left(u_{ni-j}^2 - \sigma_{ni-j}^2\right)\right|^p \tag{34}$$

$$\le n \max_{N\le i\le n} E\left|\sum_{j=0}^{N} w_{nj}\left(u_{ni-j}^2 - \sigma_{ni-j}^2\right)\right|^p$$

$$\le Kn \max_{N\le i\le n} E\left|\sum_{j=0}^{N} w_{nj}^2\left(u_{ni-j}^2 - \sigma_{ni-j}^2\right)^2\right|^{p/2}$$

where $K = 18^p\left[p^3/(p-1)\right]^{p/2}$. The last term can be bounded using Minkowski's inequality by

$$Kn \max_{N\le i\le n}\left(\sum_{j=0}^{N} w_{nj}^2\left(E\left|u_{ni-j}^2 - \sigma_{ni-j}^2\right|^p\right)^{2/p}\right)^{p/2}$$

$$\le Kn\left(\sum_{j=0}^{N} w_{nj}^2\right)^{p/2} \max_{N\le i\le n} E\left|u_{ni}^2 - \sigma_{ni}^2\right|^p.$$

Now by Minkowski's inequality, the Rao-Blackwell theorem, and assumption 8,

$$E\left|u_{ni}^2-\sigma_{ni}^2\right|^p \le \left(\left\|u_{ni}^2\right\|_p+\left\|\sigma_{ni}^2\right\|_p\right)^p \le 2^p\left\|u_{ni}\right\|_{2p}^{2p}<\infty,$$

uniformly in i and n. Further, by assumption 7, $\alpha p/2>1$, so

$$n\left(\sum_{j=0}^{N} w_{nj}^2\right)^{p/2}=n\left(O(1/N)\right)^{p/2}=O(n^{1-\alpha p/2})=o(1).$$

This establishes that (34) converges to zero as $n\to\infty$, which implies (33) by Markov's inequality.

Next, note that by the triangle inequality and Lemma 3

$$\max_{N\le i\le n}\left|\sigma_{ni}^2-\sum_{j=0}^{N} w_{nj}\sigma_{ni-j}^2\right| = \max_{N\le i\le n}\left|\sum_{j=0}^{N} w_{nj}\left(\sigma_{ni}^2-\sigma_{ni-j}^2\right)\right|$$

$$\le \sum_{j=1}^{N} w_{nj}\max_{N\le i\le n}\left|\sigma_{ni}^2-\sigma_{ni-j}^2\right| \le \max_{i\le n,\ j\le N}\left|\sigma_{ni}^2-\sigma_{ni-j}^2\right| \to_p 0. \qquad (35)$$

Next, we show that

$$\max_{N\le i\le n}\left|\hat\sigma_{ni}^2-\sum_{j=0}^{N} w_{nj}u_{ni-j}^2\right| \to_p 0. \qquad (36)$$

Since $\hat u_{ni}^2-u_{ni}^2=-2u_{ni}x_{ni}'(\hat\beta-\beta)+(\hat\beta-\beta)'x_{ni}x_{ni}'(\hat\beta-\beta)$, we have

$$\left|\sum_{j=0}^{N} w_{nj}(\hat u_{ni-j}^2-u_{ni-j}^2)\right|$$

$$\le 2\left|\frac{1}{\sqrt{n}}\sum_{j=0}^{N} w_{nj}u_{ni-j}x_{ni-j}'\right|\left|\sqrt{n}(\hat\beta-\beta)\right|+\left|\sqrt{n}(\hat\beta-\beta)\right|^2\left|\frac{1}{n}\sum_{j=0}^{N} w_{nj}x_{ni-j}x_{ni-j}'\right|.$$

Now $\left|\sqrt{n}(\hat\beta-\beta)\right| = O_p(1)$ and $\max_i\left|\frac{1}{n}\sum_{j=0}^{N} w_{nj}x_{ni-j}x_{ni-j}'\right| = O_p(N/n) = o_p(1)$. Further, since $\{w_{nj}u_{ni-j}x_{ni-j}\}$ is a MDA, by a derivation similar to that to show (33), we can show that $\max_i \left| \frac{1}{\sqrt{n}} \sum_{j=0}^{N} w_{nj} u_{ni-j} x'_{ni-j} \right| \to_p 0$. (36) follows, and together with (33) and (35) establishes that

$$\max_{N \le i \le n} \left| \hat{\sigma}_{ni}^2 - \sigma_{ni}^2 \right| \to_p 0. \tag{37}$$

Finally, note that

$$\max_{i \le N} \left| \hat{\sigma}_{ni}^2 - \sigma_{ni}^2 \right| = \max_{i \le N} \left| \hat{\sigma}_{nN}^2 - \sigma_{ni}^2 \right|$$

$$\le \left| \hat{\sigma}_{nN}^2 - \sigma_{nN}^2 \right| + \max_{i \le N} \left| \sigma_{ni}^2 - \sigma_{nN}^2 \right| \to_p 0 \tag{38}$$

by (37) and Lemma 3. (37) and (38) imply that

$$\hat{\sigma}_{n[nr]}^2 = \sigma_{n[nr]}^2 + o_p(1) \Rightarrow \sigma^2(r),$$

completing the proof. □

Proof of Theorem 7: As discussed in the text, the adjusted variance estimator satisfies $\tilde{\sigma}_{n[nr]}^2 \Rightarrow \sigma^2(r)$. The conditions of Theorem 5 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: $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_{ni}^{-2} x_{ni} u_{ni} \Rightarrow G_\sigma$. The discrepancy only involves the term

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{N} \sigma_{ni}^{-2} x_{ni} u_{ni}. \tag{39}$$

Fix some $\eta > 0$. We can then find some $\delta > 0$ such that $E \left| \int_0^\delta \sigma^{-1} dB' \operatorname{vec} D^- \right| \le \eta$. Yet for large enough n, $N = Bn^\alpha < \delta n$. Thus

$$E \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{N} \sigma_{ni}^{-2} x_{ni} u_{ni} \right| \le E \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\delta n} \sigma_{ni}^{-2} x_{ni} u_{ni} \right| \to E \left| \int_0^\delta \sigma^{-1} dB' \operatorname{vec} D^- \right| \le \eta.$$

Since η is arbitrary, the discrepancy term (39) is asymptotically negligible. The conditions of Theorem 5 are satisfied and the proof is complete. □

References

[1] Beveridge, S. and C.R. Nelson (1981): "A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'," *Journal of Monetary Economics*, 7, 151-174.

[2] Billingsley, P. (1968): *Convergence of Probability Measures*, New York: Wiley.

[3] Davidson, J. (1992): "A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes," *Econometric Theory*, 8, forthcoming.

[4] 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. R.J. Gordon.

[5] Hansen, B.E. (1992): "Convergence to stochastic integrals for dependent heterogeneous processes," forthcoming in *Econometric Theory*, 8, 489-500.

[6] Harvey, A.C. and P.C. Robinson (1988): "Efficient estimation of nonstationary time series regression," *Journal of Time Series Analysis*, 201-214.

[7] Kurtz, T.G. and P. Protter (1991): "Weak limit theorems for stochastic integrals and stochastic differential equations," *The Annals of Probability*, 19, 1035-1070.

[nf] Nelson, D.B. and D.P. Foster (1992): "Asymptotic filterting theory for univariate ARCH models," University of Chicago Graduate School of Business.

[8] Pagan, A.R. and G.W. Schwert (1990a): "Alternative models for conditional stock volatility," *Journal of Econometrics*, 45, 267-290.

[9] Pagan, A.R. and G.W. Schwert (1990b): "Testing for covariance stationarity in stock market data," *Economics Letters*, 33, 165-170.

[10] 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.

[11] Phillips, P.C.B. and V. Solo (1992): "Asymptotics for linear processes," *The Annals of Statistics*, 20, 971-1001.

[12] Wooldridge, J.M. and H. White (1988): "Some invariance principles and central limit theorems for dependent heterogeneous processes, *Econometric Theory*, 4, 210-230.