Discussion Paper No. 543

THE LINEAR MODEL WITH STOCHASTIC REGRESSORS AND HETEROSCEDASTIC DEPENDENT ERRORS

by

Ian Domowitz*

revised December 1982

Northwestern University, Department of Economics, 2003 Sheridan Road, Evanston, IL 60201.

* The author would like to thank Herman Bierens, Ron Gallant, Halbert White and the participants of the 1982 Yale Summer Econometrics Workshop for helpful comments. All errors are solely the author's. This work was supported in part by National Science Foundation Grant SES81-97552 and by a grant from the University of California.

ABSTRACT

General conditions are provided which ensure the consistency and asymptotic normality of the ordinary least squares estimator. These conditions apply to time-series, cross-section, panel, or experimental data for single equations as well as systems of equations. The errors of the regression may be heteroscedastic and/or serially correlated. A consistent estimator of the asymptotic parameter estimator covariance matrix is proposed. The consequences of misspecifying the regression function are discussed, and new tests of model specification are proposed.

1. INTRODUCTION

The linear regression model is the workhorse of empirical economics. Ease of computation and interpretation have contributed to the popularity of the linear model in analyzing data in every form encounted by economists.

The most common technique used to estimate the parameters of linear models is the method of least squares. The classical theory of least squares is based on the assumption that the observations on explanatory variables are fixed, not random. Such an assumption is inappropriate in a nonexperimental science such as economics in which data often appear as the realizations of a stochastic time-series, as a random cross-section, or as a panel.

General conditions sufficient to ensure classical properties such as consistency and asymptotic normality of the least squares estimator in the case of stochastic explanatory variables have been investigated by several researchers. Pierce (1972), Hannan (1973), Sims (1978), Kohn (1979), and Nicholls and Pagan (1982) provide conditions when the regressors are stochastic time-series, for example. Crowder (1980) develops such conditions by a careful analysis of the behavior of moment matrices, an approach also taken by Anderson and Taylor (1979) and Christopeit and Helmes (1980) in the analysis of consistency properties. White (1980b) provides very general conditions for the case in which the explanatory variables are independent, but not identically distributed, coming from a stratified cross-section, for example.

Careful investigation of the stochastic regressor problem has also led researchers to question the appropriateness of assuming that errors are independent of regressors and independent over time. Many of the above treatments assume only that the errors have zero mean conditional on current and lagged explanatory variables, as well as on past errors. Thus, errors may be dependent, but are serially uncorrelated with the explanatory variables. Important exceptions are provided by Hannan (1973) and Pierce (1972), who allow the errors to have an arbitrary, but stationary, dependence structure. White (1980b) retains the serial independence assumption, but the errors need not be identically distributed, permitting a rigorous treatment of the heteroscedasticity problem.

In this paper, these results are unified and extended by providing conditions which ensure the classical properties of the ordinary least squares estimator for most kinds of situations encountered by economists. The results given here are essentially special cases of the very general theorems for nonlinear models given in Domowitz and White (1982). Valuable insights are gained by examining the special case of the linear model, and the results are more easily interpretable. The conditions guaranteeing consistency and asymptotic normality allow the data to come from a time-series, a cross-section, a panel, or an experiment. Explanatory variables and errors may be serially correlated and/or heteroscedastic. This treatment rigorously justifies the use of ordinary least squares in a more realistic context than allowed in the classical framework. Additional insight into special cases of the theory is provided by phrasing the conditions directly in terms of the regressors and errors, instead of placing broad restrictions on moment matrices.

The role of various regularity conditions is discussed in the context of consistent estimation in Section 2. Restrictions on the allowable amount of serial dependence in terms of correlational properties engenders bounds on the moments of explanatory variables that may be unacceptable in some applications. Alternative mixing conditions are presented, allowing a direct characterization of the trade-off between dependence and knowledge of higher moments, while reducing moment restrictions in general. Similar conditions sufficient to estabish the asymptotic normality of the least squares estimator are given in Section 3. The White (1980b) heteroscedasticity-consistent covariance matrix estimator is extended to obtain a covariance matrix estimator which is consistent regardless of the presence of heteroscedasticity and/or serial correlation of unknown form in the errors. The consequences of misspecifying the model are discussed in Section 4, and specification testing is discussed in Section 5. Tests based on the Hausman (1978) and White (1982a) paradigms are presented. Section 6 concludes the paper.

2. REGULARITY CONDITIONS AND CONSISTENT ESTIMATION

The model considered in this paper is given by

ASSUMPTION 1. The model is known to be

$$Y_t = X_t\beta_0 + \varepsilon_t \qquad (t = 1,\ldots,n) \tag{2.1}$$

where $\{X_t\}$ is a sequence of random 1xp vectors and $\{\varepsilon_t\}$ is a sequence of zero-mean random scalars. The parameter vector β_0 is an unknown px1 vector of finite constants.

In a nonexperimental science such as economics, the data which form the basis for the estimation of β_0 are usually beyond the control of the investigator. It is therefore appropriate to view both the dependent and explanatory variables as realizations of a stochastic process. Since considerable heterogeneity may occur in nonexperimental data, the random vectors (X_t, ε_t) are not required to be identically distributed. As a large amount of economic data comes in the form of time-series, (X_t, ε_t) may also be serially dependent.

The model (2.1) may also be thought of as a sequential control model [e.g., Goodwin and Payne (1977)], in which X_t is a function of X_τ for $\tau < t$. If $\{X_t\}$ is a sequence of nonstochastic regressors, fixed in repeated samples, and $\{\varepsilon_t\}$ is serially independent, (2.1) is the classical linear regression model. A special case of particular interest is the q^{th} order univariate autoregressive scheme with s=p-q additional explanatory variables; i.e., $X_t = (Y_{t-1}, \ldots, Y_{t-q}, Z_{1t}, \ldots, Z_{st})$ and $\beta' = [\alpha', \gamma']$.

The ordinary least squares (OLS) estimator is defined as $\hat{\beta}_n = (X'X)^{-1}X'Y$, where X is the nxp matrix with rows X_t, and Y is the nx1 vector with elements Y_t. It is of practical interest to provide conditions which ensure that the OLS estimator retains its desirable classical properties under the wide variety of situations encountered by economists.

The OLS estimator exists almost surely (a.s.) for all n sufficiently large[1] provided (X'X/n) is nonsingular a.s. for n large. When this is true, Assumption 1 allows one to write

$$\hat{\beta}_n = \beta_0 + (X'X/n)^{-1}(X'\varepsilon/n), \tag{2.3}$$

where ε is the nx1 vector with elements ε_t. The consistency of $\hat{\beta}_n$ for β_o follows if $(X'X/n)^{-1}$ is appropriately dominated and if $X'\varepsilon/n$ converges to zero. Interpretation of the regression function as the conditional mean of Y_t given all information up to time t is often sufficient to guarantee the latter requirement. In this case, the errors satisfy $E(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_0, X_t, \ldots, X_0) = 0$. This implies

$E(X_t'\varepsilon_t \,|X'_{t-1}\varepsilon_{t-1},\ldots,X'_0\varepsilon_0) = 0$, so a strong law of large numbers for martingales [e.g., Stout (1974), Theorem 3.3.1] may then be applied to $X'\varepsilon/n$. This condition is an important weakening of the assumption of errors independent of regressors and across time [cf. Kohn (1979)]. It may be unsuitable for some applications, however, ruling out serial correlation in the errors.

Convergence of the cross-product matrix generally requires assumptions concerning the allowable dependence among regressors. For example, if the regressors have uniformly bounded fourth moments and $\text{corr}[X_{it}X_{jt}, X_{it-m}X_{jt-m}] \to 0$ sufficiently fast as $m \to \infty$ for $i,j \in \{1,\ldots,p\}$, a law of large numbers [e.g., Stout (1974), Theorem 3.7.2] may be applied directly to $X'X/n$, ensuring the appropriate convergence provided $E(X'X/n)$ is uniformly nonsingular. Covariance stationarity among the regressor cross-products is not required, but the "long distance" correlations between cross-product terms must vanish asymptotically.

In the case of the stable p^{th} order autoregressive model, consistency of the least squares estimator follows directly from the consistency of the estimated moment matrix [cf. Anderson and Taylor (1979)]. The asymptotic theory given in Anderson (1971, pages 188-211) relies on covariance stationarity and independent errors. When the independence assumption is relaxed to that of errors with mean zero given past information, both variables and errors are generally required to have bounded fourth moments [cf. Anderson and Taylor (1979), Lemma 2; Christopeit and Helmes (1980), Theorem 3; Nicholls and Pagan (1982), Theorem 1].

It is known that as the degree of dependence in the random variables increases, higher moments are required to be uniformly bounded to establish limiting results [cf. McLeish (1975); White and Domowitz (1981)]. The types of dependence restrictions discussed so far are weak in the sense that they are formulated in terms of correlational properties. Lack of correlation certainly does not imply an absence of dependence. Without knowledge of the behavior of higher moments, it may be difficult to ensure that a process eventually contains new information. The complete absence of serial correlation implied by the martingale difference assumption provides additional structure, but rules out some common situations in practice.

Some of these problems may be resolved by adopting so-called mixing conditions as ways of describing economic data which may exhibit both serial correlation and heterogeneity. These conditions restrict the memory of a process in a fashion analogous to the role of ergodicity for stationary stochastic processes. Mixing conditions are formulated in terms of probabilities instead of moments. The relevant dependence properties of a mixing stochastic process are thus invariant under a wide variety of transformations. A single assumption may then be made about the behavior of individual explanatory variables, rather than imposing broad restrictions on the cross-product matrix directly, for example. Although mixing processes may exhibit considerable dependence and heterogeneity, they are sufficiently well behaved to allow the establishment of laws of large numbers and central limit theorems, making possible a satisfactory and complete theory of estimation and inference.

Let $\{Z_t\}$ be a sequence of random vectors defined on a probability space $(\Omega, \ ,P)$, and let $\mathfrak{F}_a^b$ be the σ-algebra[2] of events generated by $(Z_a, Z_{a+1}, \ldots, Z_b)$. Define

$$\phi(m) = \sup_n \sup_{\{F\varepsilon \mathfrak{F}^n_{-\infty}, G\varepsilon \mathfrak{F}^{\infty}_{n+m}, P(F)>0\}} |P(G|F) - P(G)|$$

and

$$\alpha(m) = \sup_n \sup_{\{F\varepsilon \mathfrak{F}^n_{-\infty}, G\varepsilon \mathfrak{F}^{\infty}_{n+m}\}} |P(FG) - P(F)\ P(G)| \quad .$$

A sequence for which $\phi(m) \to 0$ as $m \to \infty$ is called uniform or ϕ- mixing [Ibragimov and Linnik (1971)], and a sequence for which $\alpha(m) \to 0$ as $m \to \infty$ is called strong or α- mixing [Rosenblatt (1956)]. Both coefficients measure the dependence between events separated by at least m time periods in the usual terms of statistical independence; i.e., how much the probability of a joint event differs from the product of the probability of each event occurring. The coefficient α provides an absolute measure of dependence, while ϕ measures dependence relative to P(F). The asymptotic independence implied between Z_k and Z_{k+m}, as $m \to \infty$, is analogous to the average asymptotic independence embodied in the definition of ergodicity for stationary stochastic processes [Rosenblatt (1972)]. A discussion of the applicabiity of these conditions to processes commonly encountered in economics is given in White and Domowitz (1981) and Domowitz and White (1982).

The next assumption, together with Assumption 1, is sufficient to demonstrate the strong consistency of the OLS estimator by making use of mixing conditions.

ASSUMPTION 2. (a) There exist positive constants $r_1 \geq 1$, $0<\delta \leq r_1$, and $\Delta<\infty$ such that $E|\varepsilon_t^2|^{r_1+\delta} < \Delta$ and $E|X_{it}X_{jt}|^{r_1+\delta} < \Delta$ for all t and $i,j\ \varepsilon\{1,\ldots,p.\}$. (b) The average moment matrix $M_n = n^{-1}\Sigma^n_{t=1}E(X_t'X_t)$ is such that $\det M_n > c > 0$ for n large. (c) The random sequence $\{X_t,\varepsilon_t\}$ is either (i) ϕ - mixing, with $\phi(m) = O(m^{-\lambda}), \lambda>r_1/(2r_1-1)$ or (ii) α-mixing with $\alpha(m) = O(m^{-\lambda})$, $\lambda > r_1/(r_1-1)$, $r_1>1$. (d) $E(X_t'\varepsilon_t) = 0$ for all t.

When condition 2(c) is met, $\phi(m)$ will be said to be of <u>size</u> $r_1/(2r_1-1)$, and similarly for $\alpha(m)$.

As the dependence restrictions become stronger, the moment restrictions become weaker (as indexed by r_1). Explosive nonstationarity is ruled out, however. In dynamic models, where $X_t = (Y_{t-1},\ldots,Y_{t-q},Z_{1t},\ldots,Z_{st})$, $\beta' = [\alpha', \gamma']$, nonexplosiveness generally requires that all roots of the characteristic equation $\lambda^q - \alpha_1\lambda^{q-1} - \ldots - \alpha_q = 0$ are less than one in absolute value.

If (X_t,ε_t) is serially independent, the process is ϕ - mixing, allowing $r_1 = 1$, which corresponds to the conditions in White (1980b, Lemma 1). If ϕ or α vanish exponentially fast, as would be the case with stationary Markov processes, r_1 may be set arbitrarily close to unity.

Assumption 2(b) ensures that $(X'X/n)^{-1}$ is well defined for large n and that its elements are uniformly bounded asymptotically. This is sufficient to guarantee the existence of the OLS estimator. Note that M_n, hence $(X'X/n)^{-1}$, is not required to converge to any limit, analogous to the analysis in White (1980b) of the case of independent regressors and errors. Assumption 2(d) is the required orthogonality condition, which is always satisfied if $E(\varepsilon_t|X_t) = 0$, for example. By assuming only contemporaneous zero correlation in the cross-product, $(X_t'\varepsilon_t)$ is allowed to be serially correlated. Such correlation constitutes a form of heteroscedasticity for which a test is formulated in Section 5.

THEOREM 2.1: Given Assumptions 1 and 2, $\hat{\beta}_n \to \beta_0$ a.s. as $n \to \infty$.

Previous consistency results for the linear model with stochastic regressors have usually exploited a martingale difference assumption on the error term, e.g., Nicholls and Pagan (1982). Results for generally dependent stationary errors are given by Pierce (1972) and Hannan (1973). The main difference between Theorem 2.1 and earlier results is that here regressors and errors may exhibit rather arbitrary time dependence and heterogeneity simultaneously.

3. ASYMPTOTIC NORMALITY AND A HETEROSCEDASTICITY-CONSISTENT COVARIANCE MATRIX ESTIMATOR

Given Assumptions 1 and 2, it is easy to show that if $(1/\sqrt{n})B_n^{-1/2}\,\Sigma_{t-1}^{n}\,X_t'\varepsilon_t$ is asymptotically $N(0,I_p)$ for some sequence of positive definite matrices $\{B_n\}$, then the OLS estimator is asymptotically normally distributed. Typically $B_n = \mathrm{var}[n^{-1/2}\Sigma_{t=1}^{n}X_t'\varepsilon_t]$. If $X_t\beta_o$ is the conditional expectation of Y_t given the past history of the process and the explanatory variables, a central limit for martingales [e.g. Scott (1973)] can be applied to $X_t'\varepsilon_t$. The asymptotic covariance matrix of $n^{-1/2}\Sigma_{t=1}^{n}X_t'\varepsilon_t$ is then $B_n = n^{-1}\Sigma_{t=1}^{n}E(\varepsilon_t^2X_t'X_t)$, the form given in White (1980b). The simplicity of the average covariance matrix is a result of the lack of serial correlation in $X_t'\varepsilon_t$. Allowing for serial correlation in both regressors and errors, the covariance matrix may be written as

$$B_n = n^{-1}\sum_{t=1}^{n} E(\varepsilon_t^2\, X_t'X_t) \tag{3.1}$$

$$+\, n^{-1}\sum_{\tau=1}^{n-1}\sum_{t=\tau+1}^{n} E(\varepsilon_t\varepsilon_{t-\tau}[X_t'\, X_{t-\tau}+X_{t-\tau}'\, X_t]),$$

In the case of fixed regressors, B_n is simply $X'\Sigma X/n$ where $\Sigma = [E(\varepsilon_t\varepsilon_\tau)]$.

The next assumption formally identifies the limit of (3.1) as the asymptotic covariance of $n^{-1/2}\ \Sigma_{t=1}^{n}\ X_t'\varepsilon_t$ and specifies the additional conditions under which the OLS estimator has the normal distribution asymptotically.

ASSUMPTION 3. (a) There exist positive constants $r_2 > 1$, $\Delta < \infty$, such that $E|\varepsilon_t^2 X_{it}^2|^{r_2} < \Delta$, $i = 1,\ldots p$. (b) Define $B_{a,m} = \mathrm{var}(n^{-1/2}\ \Sigma_{t=a+1}^{a+m}\ X_t'\ \varepsilon_t)$. There exists a matrix B such that det $B > 0$ and $\lambda' B_{a,m}\ \lambda - \lambda' B\ \lambda \to 0$ as $m \to \infty$ uniformly in a for any real nonzero px1 vector λ. (c) The random vector (X_t, ε_t) is either (i) ϕ- mixing, with $\phi(m)$ of size $r_2/(r_2-1)$ or (ii) α-mixing, with $\alpha(m)$ of size max $[r_1/(r_1-1),\ r_2/(r_2-1)]$, $r_1 > 1$.

Assumption 3(b) requires that the asymptotic covariance matrix B not depend on n, imposing a restriction on the allowable amount of heterogeneity in the data. Processes $\{Z_t\}$ for which $\lim_{n\to\infty}\ n^{-1}\ \Sigma_{t=1}^{n}\ E(Z_t Z_{t-\tau})$ converges to a stationary covariance $r(\tau)$, $\tau = 0,1,\ldots,$ have been studied by Kampi de Feriet and Frenkiel (1962), Rao (1978), and Parzen (1962), who termed such series "asymptotically weakly stationary." Central limit results were not obtained for such processes, however. The uniform convergence requirement of Assumption 3(b) is an additional restriction which allows a central limit theorem to hold. In the special case of m-dependent processes, however, B may depend on n.

THEOREM 3.1: Given Assumptions 1-3, $\sqrt{n}\ B_n^{-1/2} M_n(\hat{\beta}_n - \beta_0) \overset{a}{\sim} N(0, I_p)$, where $B_n \equiv B_{0,n}$.

A similar result was obtained by Eicker (1967) for linear models with fixed regressors. The present theorem incorporating random regressors is the dependent variable analogue of Lemma 2 of White (1980b). The covariance matrix $C_n \equiv M_n^{-1} B_n M_n^{-1}$ has the same form as the conditional parameter estimator convariance matrix of Hansen (1982) in the linear case. Hansen assumes jointly strictly stationary, ergodic, regressors and errors, but his covariance matrix accommodates complicated conditional covariance structures for the regressors and errors. The present result allows for both conditional and unconditional variation, but the practical implication of both results (i.e., the form of the covariance matrix) is the same in both cases.

Asymptotic approximations to the normal law are usually used to construct tests of hypotheses. Given a consistent estimator of the present result yields test statistics which are robust to the presence of heteroscedastcity and/or serial correlation of unknown or incorrectly specified form. Suppose it is desired to test the linear hypothesis[3]

$$H_0 : R\beta_o = r$$

versus

$$H_1 : R\beta_o \neq r,$$

where R is a finite qxp matrix of full row rank and r is a finite qx1 vector. The appropriate form of the Wald statistic is given by the next result.

THEOREM 3.2: Given the conditions of Theorem 3.2, H_0, and $\hat{B}_n$ such that $\hat{B}_n - B_n \xrightarrow{p} 0$,

$$n(R\hat{\beta} - r)' \, [R(X'X/n)^{-1} \hat{B}_n (X'X/n)^{-1} R']^{-1} (R\hat{\beta} - r) \overset{a}{\sim} \chi^2_p .$$

Under identical conditions, the analogous Lagrange Multiplier (LM) test statistic is asymptotically equivalent. It should be emphasized that the usual form of the Wald or LM statistic uses $\hat{\sigma}_n^2(X'X/n)^{-1}$ as the covariance matrix estimator, where $\hat{\sigma}_n^2 = n^{-1}\sum_{t=1}^{n}(Y_t - X_t\hat{\beta}_n)^2$. Heteroscedasticity and/or serial correlation of unknown form generally invalidates inferences based this standard estimator. The use of the appropriate estimate of C_n in its place is required to ensure a test of proper size. Since C_n is more complicated to compute, the information matrix testing principle of White (1982a) is applied in Section 5 to test for conditions which ensure the consistency of $\hat{\sigma}_n^2(X'X/n)^{-1}$.

The computation of C_n in the general case requires an estimator of (3.1). A natural candidate is an estimator of the form

$$\hat{B}_n = n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_t^2 X_t'X_t + n^{-1}\sum_{\tau=1}^{\ell}\sum_{t=\tau+1}^{n}\hat{\varepsilon}_t\hat{\varepsilon}_{t-\tau}[X_t'X_{t-\tau} + X_{t-\tau}'X_t] \qquad (3.2)$$

where $\hat{\varepsilon}_t = Y_t - X_t\hat{\beta}_n$. The second term of (3.1) is truncated at $\tau = \ell \leq n-1$ and the integer ℓ is called the truncation lag.

In some situations, it may be thought that $E(\varepsilon_t\varepsilon_{t-\tau}|X_t, X_{t-\tau}) = 0$ for all $\tau > \ell > 0$. This is the case when $\{\varepsilon_t\}$ is an MA(ℓ) process, for example. If $E(Y_t|X_{t-m}) = X_t\beta_0$ ($m=0,1,\ldots$; $t = 1,\ldots,n$), which is a common assumption in dynamic models containing lagged dependent variables, $E(\varepsilon_t\varepsilon_{t-\tau}|X_t, X_{t-\tau}) = 0$ for $\tau \geq 1$ In all such cases, the following additional moment condition is used to prove the consistency of $\hat{B}_n$ for B_n.[4]

ASSUMPTION 4. There exist positive constants $r_1 \geq 1$, $0 < \delta \leq r_1$ and $\Delta < \infty$ such that for all t, $E|\varepsilon_t^2 X_{it}^2|^{r_1+\delta} < \Delta$ and $E|X_{it}^4|^{r_1+\delta} < \Delta$, $i=1,\ldots p$.

The assumption that $E(\varepsilon_t \varepsilon_{t-\tau}|X_t, X_{t-\tau}) = 0$ for $\tau > \ell > 0$ may be too much to accept in some applications. In such cases, the mixing property ensures that the contribution of distant lags will be negligible, as long as ℓ grows with n. The truncation lag cannot grow too quickly relative to n, however, since not enough information will be available to estimate all covariances consistently. With the next assumption, admissible growth rates for ℓ may be specified.

ASSUMPTION 5. There exist positive constants $r_1 \geq 1$, $0 < \delta \leq r_1$ and $\Delta < \infty$ such that for all t, $E|X_{it}^6|^{r_1+\delta} < \Delta$, i=1,.., p.

The next theorem formalizes the foregoing discussion.

THEOREM: 3.3: (a) Given Assumptions 1-4, if $E(\varepsilon_t \varepsilon_{t-\tau}|X_t, X_{t-\tau}) = 0$ for all $\tau > \ell$, $\ell < \Delta < \infty$, $\hat{B}_n - B_n \xrightarrow{p} 0$.
(b) Given Assumptions 1-5, if $\ell \to \infty$ as $n \to \infty$ such that $\ell = O(n^\gamma)$, $0 < \gamma < \delta/(r_1+\delta) \leq 1/2$, either (a) $\phi(m)$ is of size 2 or (b) $\alpha(m)$ is of size $2(r_1 + \delta)/(r_1 + \delta - 1)$, $r_1 > 1$, then $\hat{B}_n - B_n \xrightarrow{p} 0$.

The result provides the conditions under which $\hat{B}_n$ is consistent for B_n, given the possible dependence assumptions on the error structure of the model (2.1). In the case of generally dependent errors, ℓ must grow with n, but more slowly than $n^{\delta/(r_1+\delta)}$. Heuristic methods of choosing ℓ in applications are discussed in White and Domowitz (1981). The marginal computational costs of including extra lags are negligible, however, suggesting that a potentially reasonable strategy would be to include <u>all</u> lags up to $n^{\delta/(r_1+\delta)}$. Further study of this issue is needed.

The covariance estimator $\hat{C}_n = (X/X/n)^{-1}\hat{B}_n(X'X/n)^{-1}$ is the time-series generalization of White's (1980b) heteroscedasticity-consistent covariance matrix estimator. Taken together, Theorems 3.2 and 3.3 extend White's (1980b) Theorem 1 to cover most situations encountered by economists. The results apply to time-series, cross-section, panel, or experimental data, and may be applied to systems of equations as well as single equation models.

These theorems are derived assuming that the conditional expectation or behavioral law, $X_t\beta_0$, is known to the investigator. This assumption is often difficult to accept in practice and is relaxed in the next section, in which the consequences of misspecifying the regression function are briefly examined.

4. THE LEAST SQUARES APPROXIMATION TO AN UNKNOWN REGRESSION FUNCTION

In this section, the assumption of a known model is replaced by

ASSUMPTION 1'. The sequence of real-valued responses Y_t is generated as

$$Y_t = g_t(Z_t) \qquad (t=1,\ldots,n) \tag{4.1}$$

where the g_t are unknown measurable functions of the real-valued random vector Z_t. The vector Z_t is finite-dimensional, may contain unobservable elements, and is jointly distributed with distribution function F_t on Ω, a Euclidean space.

Once again, it is not assumed that Y_t or Z_t is stationary. Stationarity is a particularly strong assumption in the context of potential misspecification.

The researcher chooses the linear regression function $X_t\beta$ to approximate $g_t(Z_t)$. In order to ensure that this approximation is well defined, it is assumed that X_t is a measurable function of Z_t. If suffices to consider X_t as a subvector of Z_t, where $\partial g_t/\partial z_i$ may be identically zero for some i, allowing for the inclusion of irrelevant variables.

Under the regularity conditions given below, the OLS estimator is consistent for β_n^*, the parameter vector which minimizes the average approximation (or prediction) mean squared error (MSE),

$$\bar{\sigma}_n^2 = n^{-1} \sum_{t=1}^{n} \int (g_t(Z_t) - X_t\beta)^2 \, dF_t. \tag{4.2}$$

For example, suppose $g_t(Z_t) = Z_t\gamma_0$. Then the vector of parameters which is estimated is given by the minimizers

$$\beta_n^* = [n^{-1} \sum_{t=1}^{n} E(X_t'X_t)]^{-1} [n^{-1} \sum_{t=1}^{n} E(X_t'Z_t)\gamma_0] \quad . \tag{4.3}$$

Under stationarity assumptions, $\beta^* = [E(X_t'X_t)]^{-1}[E(X_t'Z_t)\gamma_0]$ [cf. Hendry (1979)]. It is undesirable to assume that the covariance between Z_t and X_t is constant over time, however, lacking any knowledge of Z_t. An example in which β_n^* may fail to converge is given in Domowitz and White (1982).

Let $u_t = g_t(Z_t) - X_t\beta$. The following modification of Assumption 2 provides regularity conditions sufficient to ensure the consistency of $\hat{\beta}_n$ for the minimizers β_n^*.

ASSUMPTION 2'. (a) Assumptions 2(a) and 2(b) hold, replacing ε_t by u_t. (b) The random vector Z_t is either (i) ϕ-mixing with $\phi(m)$of size $r_1/(2r_1 - 1)$,

$r_1 \geq 1$ or (ii) α-mixing with $\alpha(m)$ of size $r_1/(r_1-1), r_1 > 1$.

THEOREM 4.1. Under Assumptions 1' and 2', $\hat{\beta}_n - \beta_n^* \to 0$ a.s. as $n \to \infty$.

This result is the time-series generalization of Theorem 2 of White (1980b). The theorem says that the least squares estimator is a strongly consistent estimator of the parameter vector which minimizes the average MSE of prediction. In fact, β_n^* is the parameter vector of a least squares approximation [5] $X_t\beta$ to an unknown function $g_t(Z_t)$, with weighting functions dF_t. Note that Assumption 2 restricts the moments of the approximation error, ruling out the use of a linear trend to approximate a constant, for example.

If $g_t(Z_t) \equiv X_t\beta_0 + \varepsilon_t$ for all t, $\beta_n^* = \beta_0$ for any sequence of weighting functions dF_t, as Theorem 4.2 below establishes. Otherwise, the parameter vector of the approximation will depend on the weighting functions. Define the weights $\{W_t\}$ as positive measureable functions of the Z_t taking values on a compact interval, and normalized such that $\int W_t dF_t = 1$, $t=1,\ldots,n$. Let Σ^{-1} be a diagonal matrix with nonzero elements W_t. The weighted least squares (WLS) estimator is then

$$\tilde{\beta}_n = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y. \tag{4.3}$$

If $\tilde{\beta}_n^*$ minimizes the average MSE of approximation with weighting functions $d\tilde{F}_t = W_t\, dF_t$, $\tilde{\beta}_n - \tilde{\beta}_n^* \to 0$ a.s. under Assumptions 1' and 2'. The next result provides conditions sufficient to ensure that β_n^* does not depend on the weights.

THEOREM 4.2. Suppose $g_t(Z_t) = X_t\beta_0 + \varepsilon_t$ where $E(\varepsilon_t) \equiv 0$ and $E(W_tX_t'\varepsilon_t) = 0$ for all t. If Assumptions 1' and 2' are satisfied, then $\tilde{\beta}_n \to \beta_0$ a.s. as $n \to \infty$.

The theorem extends the results of White (1980b) to the case of weighted least squares for dependent observations and errors. When the model represents a conditional expectation, the orthogonality condition of the theorem is always satisfied, and $\tilde{\beta}_n \to \beta_0$ a.s. regardless of the weights W_t (and distributions F_t). This case will play an important role in the specification analysis of the next section.

The asymptotic normality of the OLS estimator may be obtained with an extension of Assumption 3.

ASSUMPTION 3: (a) There exist positive constants $r_2 > 1$, $\Delta < \infty$, such that for all t, $E|u_t^2 X_{it}^2|^{r2} < \Delta$, $i=1,\ldots,p$. (b) Define $B^*_{a,n} = \text{var}(n^{-1/2}\Sigma_{t=a+1}^{a+n} X_t'u_t(\beta_n^*))$. There exists a matrix B^* such that $\lambda' B^*_{a,n}\lambda - \lambda' B^*\lambda \to 0$ as $n \to \infty$ uniformly in a for any real nonzero px1 vector λ. (c) The random vectors $\{Z_t\}$ are either (i) ϕ-mixing, with $\phi(m)$ of size $r_2/(r_2-1)$ or (ii) α-mixing, with $\alpha(m)$ of size $\max\,[r_1/(r_1-1),\ r_2/(r_2-1)]$, $r_1 > 1$.

THEOREM 4.3. Under Assumptions 1', 2' and 3',

$$\sqrt{n}\, B_n^{*-1/2} M_n(\hat{\beta}_n - \beta_n^*) \overset{a}{\sim} N(0, I_p)$$

where $B_n^* \equiv B^*_{o,n}$.

This asymptotic normality result would establish the basis for tests of hypotheses concerning parameters of the approximation if a consistent estimator of B_n^* could be found. Unfortunately, this is not possible in general. The estimator (3.2) is not generally consistent when the regression function is misspecified, and the covariance matrix of the least squares estimator depends on the true data-generating process, as noted by Chow (1981) and Burguete, Gallant, and Souza (1982). When the observations are independent, it can be shown that (3.2) provides an upper bound for B_n^* asymptotically. This property does not generalize for $\ell \geq 1$. Theorem 3.4 of Domowitz and White (1982) establishes a necessary and sufficient condition for $\hat{B}_n$ to consistently estimate B_n^*, namely

$$\lim_{n\to\infty} n^{-1} \sum_{\tau=\ell+1}^{n-1} \sum_{t=\tau+1}^{n} E(X_t' u_t(\beta_n^*)) \; E\,(u_{t-\tau}(\beta_n^*)X_{t-\tau}) = 0.$$

This condition is satisfied if the model is correctly specified or if g_t and F_t are time invariant.

5. SPECIFICATION TESTS

Most models used in empirical studies represent, intentionally or not, approximations to some unknown underlying data-generating process. The results of the last section are useful insofar as they provide insight into the strengths and limitations of our approximations. The least squares approximation has optimal prediction properties, but OLS estimates the parameters of the approximation, rather than parameters of interest to the economist, in general.

The understanding of empirical phenomena, rather than prediction, is often of primary interest, however. Interpretation in terms of economic theory relies both on a model correct in the sense of Theorem 4.2 and on valid hypothesis testing procedures. The latter requires a consistent estimate of the OLS covariance matrix estimator, while the former depends on $E(X_t'\varepsilon_t W_t) = 0$. Such zero correlation may be difficult to verify for arbitrary weighting functions, W_t. The tests discussed below maintain the stronger hypothesis of the following definition.

DEFINITION 5.1. If $E(Y_t|X_t) = X_t\beta_0$ a.s. for all t, the model is said to be correct to first order.

If the model is correct to first order, the investigator can be confident that OLS provides consistent estimates of parameters of interest. The next definition formalizes the condition under which valid inferences can be made based on the standard least squares covariance matrix estimator $\hat{\sigma}_n^2(X'X)^{-1}$, where $\hat{\sigma}_n^2 = n^{-1}\Sigma_{t=1}^n(Y_t-X_t\hat{\beta}_n)^2$.

DEFINITION 5.2. If the model is correct to first order and $B_n = \sigma_0^2 n^{-1}\Sigma_{t=1}^n E(X_t' X_t)$, the model is said to be correct to second order.

Serially uncorrelated homoscedastic errors independent of the regressors are sufficient to satisfy the above definitions [cf. White (1980b)]. The following conditions are also sufficient in the sense that the White (1982a) test for $B_n=\sigma_0^2 n^{-1}\Sigma_{t=1}^n E(X_t'X_t)$ may be obtained under them.

ASSUMPTION 6. (a) There exist positive constants $r_1 \geq 1$, $0 < \delta \leq r_1$ and $\Delta < \infty$ such that for all t, $E|\varepsilon_t|^{4r1+\delta} < \Delta$ and $E|X_{it}|^{8r1+\delta} < \Delta$. Define $V_{a,n} =$ var $[n^{-1/2}\Sigma_{t=a+1}^{a+n}(1, \psi_{to})'(\varepsilon_t^2-\sigma_0^2)]$, where ψ_{to} is a $1xk_o$ vector with elements $X_{it}X_{jt}$, $i,j \in \{1,\ldots p\}$. There exists a matrix V_o such that $\det V_o > c > 0$ and $\lambda' V_{a,n}\lambda - \lambda' V_o \lambda \to 0$ as $n \to \infty$ uniformly in a for any real nonzero $(k_o+1)\mathrm{x}1$ vector λ. (c) $E(\varepsilon_t^4 | X_t) = \kappa_o < \infty$ a.s. for all t, and $E(\varepsilon_t^2 | X_t, X_{t-1}, \ldots ; \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \ldots) = \sigma_o^2 < \infty$ a.s. for all t.

The conditions given in Assumption 6 are not the most general conditions under which the White test for heteroscedasticity may be obtained. Instead, the conditions ensure that the test may be computed as n times the constant-adjusted R^2 of the artificial regression

$$\hat{\varepsilon}_t^2 = \alpha_0 + \psi_{t0}\alpha \quad (t=1,\ldots,n). \tag{5.1}$$

Assumption 6(d) essentially ensures that the errors of (5.1) are homoscedastic and serially uncorrelated, so that OLS is appropriate. The dimension of ψ_{t0} is generally $1\mathrm{x}k_0$, $k_0 \leq p(p+1)/2$; i.e., ψ_{t0} is the vector containing the elements of the lower triangle of $X_t' X_t$. Redundancies may occur, violating Assumption 6(c). For example, X_t may contain a vector of ones. In such cases, the redundant elements of ψ_{t0} may be deleted, reducing k_0 and degrees of freedom for the test statistic below.

THEOREM 5.3. Given Assumptions 1-3 and 6, if the model is correct to second order, n times the constant-adjusted R^2 from the regression (5.1) is asymptotically distributed as $\chi^2_{k_0}$.

Serial correlation in ε_t is usually sufficient to cause $E(\varepsilon_t^2 | X_t, X_{t-1}, \ldots ; \varepsilon_{t-1}^2, \ldots) \neq \sigma_0^2$. The artificial regression errors are then serially correlated, resulting in the incorrect size of the nR^2 statistic.

The presence of autoregressive conditional heteroscedasticity [ARCH, See Engle (1982)] leads to the same difficulty. A test for ARCH disturbances can be constructed under similar conditions, replacing the vector ψ_{t0} in (5.1) by a q-vector of lagged squared residuals, $(\hat{\varepsilon}^2_{t-1}, \hat{\varepsilon}^2_{t-2}, \ldots, \hat{\varepsilon}^2_{t-q})$. When testing for q^{th} order ARCH, the asymptotic χ^2 statistic will have q degrees of freedom.

A White test for τ^{th} order serial correlation in $X_t'\varepsilon_t$ may be constructed by comparing the τ^{th} term of $\hat{B}_n$ to zero. The test statistic is computed by $n-\tau$ times the constant-unadjusted R^2, denoted R_0^2, from the artificial regression

$$\varepsilon_t \varepsilon_{t-\tau} = \psi_{t\tau}\,\alpha \qquad (\tau=1,2,\ldots;\ t=\tau+1,\ldots n)\ , \qquad (5.2)$$

where $\psi_{t\tau}$ is a $1 \times k_\tau$ vector with elements $X_{it}X_{jt-\tau}$, $i,j \in \{1,..,p.\}$.

The following conditions are imposed to ensure that OLS is appropriate for (5.2).

ASSUMPTION 7. (a) Define $V_{a,n,\tau} = \mathrm{var}(n^{-1/2}\sum_{t=a+1}^{a+n} \psi'_{t\tau}\, \varepsilon_t \varepsilon_{t-\tau})$. There exists a matrix V_τ such that $\det V_\tau > c > 0$ and $\lambda' V_{a,n,\tau}\lambda - \lambda' V_\tau \lambda \to 0$ as $n \to \infty$ uniformly in a for any real nonzero $k_\tau \times 1$ vector λ. (b) $E(\varepsilon_t | X_{t+\tau}, X_t, X_{t-\tau}) = 0$ a.s. for all t, and $E(\varepsilon_t^2 \varepsilon_{t-\tau}^2 | X_t, X_{t-\tau}) = \kappa_\tau < \infty$ a.s. for all t.

THEOREM 5.4. Given Assumptions 1-3, 6(a), (b), and 7, if the model is correct to second order, $n-\tau$ times R_0^2 from the regression (5.2) is asymptotically distributed as $\chi^2_{k_\tau}$.

Serial correlation in $X_t'\varepsilon_t$ may arise in models in which the parameters are erroneously believed constant, for example. In such cases, the mean of the (random) parameter vector is often consistently estimated by OLS. The covariance structure of the OLS estimator then includes B_n^* as in Section 4, since the errors are functions of omitted dynamic coefficients.

The tests of Theorems 5.2 and 5.3 are sensitive to failures in the assumption of a model correct to first order, as well as to forms of heteroscedasticity and/or serial correlation which cause $B_n \neq \sigma_0^2 n^{-1}\Sigma_{t=1}^n E(X_t'X_t)$. If the hypothesis that the model is correct to second order is rejected, it may still be that the model is correct to first order. A test of the latter hypothesis may be based on the results of Section 4. When the model is correct to first order, the WLS estimator is consistent for β_o using any set of positive weights which are measurable functions of X_t and take values on a compact interval. Whenever two such WLS estimators are available, the distance between them can be used as an indicator of model misspecification. This distance is zero asymptotically if the model is correct, but generally does not vanish otherwise.

This fact has been exploited previously in the context of maximum likelihood estimation and asymptotically efficient estimators [e.g., by Wu (1974), Hausman (1978), and White (1982a)]. The next test does not rely on asymptotic efficiency, since this would generally require a knowledge of the joint distribution of the errors and covariance stationarity. In this sense, the test is similar in form and spirit to that of White (1980b) and to those proposed for nonlinear regression models in White (1981) and Domowitz and White (1982).

Let $\{\hat{\beta}_{1n}\}$ and $\{\hat{\beta}_{2n}\}$ be two sequences of WLS estimators using weights $\{W_{1t}\}$ and $\{W_{2t}\}$ as defined in Section 4. The sequence $\{W_{1t}\}$ may simply be a sequence of ones, making $\hat{\beta}_{1n}$ the OLS estimator. Define

$$R_n = n^{-1} \sum_{t=1}^{n} E(W_{1t} W_{2t} \varepsilon_t^2 X_t' X_t) \tag{5.3}$$

$$+ n^{-1} \sum_{\tau=1}^{n-1} \sum_{t=\tau+1}^{n} E(\varepsilon_t \varepsilon_{t-\tau} (X_t' X_{t-\tau} W_{1t} W_{2t-\tau} + X_{t-\tau}' X_t W_{1t-\tau} W_{2t}]).$$

The average covariance matrix of the difference between the two WLS estimators given by

$$S_n = M_{1n}^{-1} B_{1n} M_{1n}^{-1} + M_{2n}^{-1} B_{2n} M_{2n}^{-1} - M_{1n}^{-1} R_n M_{2n}^{-1} - M_{2n}^{-1} R_n M_{1n}^{-1} \tag{5.4}$$

where $M_{in} = X' \Sigma_i^{-1} X$, $\Sigma_i^{-1} = \mathrm{diag}(W_{it})$ as defined in Section 4, and B_{in} is computed with weights W_{it}, $t=1,\ldots,n$.

Let $\hat{\varepsilon}_t = Y_t - X_t \hat{\beta}_{1n}$. Under the null hypothesis of a model correct to first order, the covariance estimator required for the test statistic is given by

$$\hat{S}_n = (X' \Sigma_1^{-1} X/n)^{-1} \hat{B}_{1n} (X' \Sigma_1^{-1} X/n)^{-1} \tag{5.5}$$

$$+ (X' \Sigma_2^{-1} X/n)^{-1} \hat{B}_{2n} (X' \Sigma_2^{-1} X/n)^{-1}$$

$$- (X' \Sigma_1^{-1} X/n)^{-1} \hat{R}_n (X' \Sigma_2^{-1} X/n)^{-1}$$

$$- (X' \Sigma_2^{-1} X/n)^{-1} \hat{R}_n (X' \Sigma_1^{-1} X/n)^{-1} ,$$

where

$$\hat{R}_n = n^{-1} \sum_{t=1}^{n} W_{1t} W_{2t} \hat{\varepsilon}_t^2 X_t' X_t \tag{5.6}$$

$$+ n^{-1} \sum_{\tau=1}^{\ell} \sum_{t=\tau+1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-\tau} (X_t' X_{t-\tau} W_{1t} W_{2t-\tau} + X_{t-\tau}' X_t W_{1t-\tau} W_{2t})$$

and $\hat{B}_{in}$ is evaluated using $\hat{\varepsilon}_{it} = Y_t - X_t \hat{\beta}_{in}$.

The next assumption formally identifies S_n as the covariance matrix of $\hat{\beta}_{1n} - \hat{\beta}_{2n}$ and provides the additional moment restrictions required to guarantee consistent estimation of S_n.

ASSUMPTION 8. Define $S_{a,n} = \mathrm{var}[n^{-1/2}\Sigma_{t=a+1}^{a+n} (M_{1n}^{-1} X_t' W_{1t}\varepsilon_t - M_{2n}^{-1} X_t' W_{2t}\varepsilon_t)]$. There exists a matrix S such that det $S>c>0$ and $\lambda' S_{a,n}\lambda - \lambda' S \lambda \to 0$ as $n\to\infty$ uniformly in a, for any real nonzero px1 vector λ.

The memory condition is also modified to ensure that $\hat{S}_n$ consistently estimates S_n.

ASSUMPTION 9. Either (a) $\phi(m)$ is of size 2 or (b) $\alpha(m)$ is of size $2(r_1 + \delta)/(r_1 + \delta - 1)$, $r_1 > 1$.

THEOREM 5.5. Let $\{\hat{\beta}_{1n}\}$ and $\{\hat{\beta}_{2n}\}$ be sequences of WLS estimators using weights $\{W_{1t}\}$ and $\{W_{2t}\}$. If Assumptions 1 - 3, 8, and 9 hold, and if $\ell \to \infty$ as $n \to \infty$, $\ell = O(n^{\gamma})$, $0 < \gamma < \delta/(r_1 + \delta) \leq 1/2$,

$$n(\hat{\beta}_{1n}-\hat{\beta}_{2n})\hat{S}_n^{-1}(\hat{\beta}_{1n}-\hat{\beta}_{2n}) \overset{a}{\sim} \chi_p^2 \tag{5.7}$$

when the model is correct to first order.

The statistic (5.7) is the time-series generalization of the specification test given in White (1980b). The statistic requires the computation of two WLS estimates. Although (5.7) appears complicated, all matrices and their inverses are available as byproducts of the WLS estimates. Note that only one truncation lag for $\hat{B}_{1n}$, $\hat{B}_{2n}$ and $\hat{R}_n$ has been defined. This need not be the case in practice, so long as each truncation lag satisfies the conditions of the theorem.

The test given by (5.7) is limited by the choice of weights, which is apparently arbitrary. Heuristically, it seems desirable to weight portions of the parameter space most heavily where the approximation is the poorest, in hope of improving the power of the test. A set of weights which satisfies this requirement is given by the fitted values of the regression (5.1). A discussion of this case may be found White (1980b,1981). A treatment of the weighting problem is beyond the scope of the present study. It is, however, an important topic for future research, since issues of test power and relative estimator efficiency also depend on the merits of alternative weighting schemes.

6. CONCLUSION

General conditions sufficient to ensure the consistency and asymptotic normality of the OLS estimator have been provided in this paper. Regression errors may be heteroscedastic and/or serially correlated, while the data may come from a cross-section, a panel, time-series, or an experiment. A heteroscedasticity-consistent covariance matrix estimator is introduced for the time-series framework considered here. The estimator allows the development of hypothesis testing procedures robust to serial correlation and/or heteroscedasticity of unknown or misspecified form in the errors.

The consequences of misspecifying the mean regression function are then explored. The OLS estimator is found to be consistent for the parameters of a well-defined least squares approximation to an unknown regression function. The parameters of the approximation need not be stable over time. Additional conditions are provided which ensure that the OLS estimator is asymptotically normal. In the presence of general model misspecification, the estimator of the asymptotic parameter covariance matrix need not be consistent, however. General tests of model misspecification are derived under the null hypothesis of a correctly specified model, based on principles advocated by Hausman (1978) and White (1980a, 1982).

It is well known that unknown forms of serial correlation in the errors of dynamic models lead to inconsistent parameter estimates. In such cases, the method of instrumental variables is an appropriate estimation technique. Future work on the linear model will include an extension of the analysis here to that case, extending the framework introduced by White (1982b).

MATHEMATICAL APPENDIX

All notation, definitions, and assumptions correspond to those given in the text. Limits are taken as $n \to \infty$, unless stated otherwise. Footnote 2 applies in the Appendix as well as in the main text.

Proof of Theorem 2.1: The argument follows the basic line given by White (1980b, Lemma 1). Provided that $X'X/n$ is nonsingular for n large,

$$\hat{\beta}_n = \beta_0 + (X'X/n)^{-1} (X'\varepsilon/n) \tag{A.1}$$

under Assumption 1. By Lemma 2.1 of White and Domowitz (1981), $\{X_{it}X_{jt}\}$, $i,j \in \{1,\ldots,p\}$ are mixing sequences with the size properties of Assumption 2(c). Since $E|X_{it}X_{jt}|^{r_1+\delta} < \Delta$ for all t, it follows by Theorem 2.10 of McLeish (1975) that $|X'X/n - M_n| \to 0$, a.s., where $|\cdot|$ indicates convergence element by element.

By Assumption 2(b), M_n is nonsingular for n large, and thus so is $X'X/n$. Assumption 2(a) bounds the elements of M_n^{-1} for n large, so that $|(X'X/n)^{-1} - M_n^{-1}| \to 0$, a.s.

Now, $E|X_{it}\varepsilon_t|^{r_1+\delta} \leq E^{1/2}|X_{it}^2|^{r_1+\delta} \; E^{1/2}|\varepsilon_t^2|^{r_1+\delta} < \Delta$, for all t, and $i = 1,\ldots,p$, by Assumption 2(a) and the Holder inequality. $\{X_{it}\varepsilon_t\}$ is a mixing sequence meeting the size requirements of Theorem 2.10 of McLeish (1975), so that $|X'\varepsilon/n - n^{-1} \Sigma_{t=1}^{n} E(X_t'\varepsilon_t)| \to 0$, a.s. Since $X_t'\varepsilon_t$ has uniformly bounded elements for n large, as does $(X'X/n)^{-1}$, and since $E(X_t'\varepsilon_t) = 0$, $|(X'X/n)^{-1}(X'\varepsilon/n)| \to 0$, a.s. The result follows from (A.1).

Proof of Theorem 3.1: Consider the quantity $n^{-1/2}\Sigma_{t=1}^{n} X_t'\varepsilon_t$. Under Assumption 3, the sequence $\{X_t'\varepsilon_t\}$ is mixing, and the average covariance matrix is given by

$$B_n = n^{-1}\sum_{t=1}^{n} E(\varepsilon_t^2 X_t'X_t)$$
$$+ n^{-1}\sum_{\tau=1}^{n-1}\sum_{t=\tau+1}^{n} E(\varepsilon_t\varepsilon_{t-\tau}[X_t'X_{t-\tau} + X_{t-\tau}'X_t]). \qquad \text{(A.2)}$$

B_n is positive definite for n large by Assumption 3(b), and so the matrix square root $B_n^{1/2}$ and its inverse, $B_n^{-1/2}$, are well defined. Applying Theorem 2.6 of Domowitz and White (1982), a central limit result for mixing random variables, it follows that $\sqrt{n}\, B_n^{-1/2}\, \Sigma_{t=1}^{n} X_t'\varepsilon_t \overset{a}{\sim} N(0,I_p)$. By (A.1),

$$\sqrt{n}\, B_n^{-1/2} M_n (X'X/n)^{-1} \sum_{t=1}^{n} X_t'\varepsilon_t = \sqrt{n}\, B_n^{-1/2} M_n(\hat{\beta}_n - \beta_0) \qquad \text{(A.3)}$$

a.s. for n large. Since $|X'X/n)^{-1} - M_n^{-1}| \to 0$ a.s., as argued in Theorem 2.1, $|M_n(X'X/N)^{-1} - I_p| \xrightarrow{p} 0$ by Lemma 3.2 of White (1980c) and it follows that $\sqrt{n}\, B_n^{-1/2} M_n(\hat{\beta}_n - \beta_o) \overset{a}{\sim} N(0,I_p)$.

Proof of Theorem 3.2: Under H_0,

$$R\hat{\beta}_n - r = R\beta_0 - r + R(\hat{\beta}_n - \beta_0) = R(\hat{\beta}_n - \beta_0). \qquad \text{(A.4)}$$

Let $S_n = RM_n^{-1}B_nM_n^{-1}R'$, which has uniformly bounded elements and is positive definite large given Assumptions 1-3 and H_0. The result follows from Lemma 3.3 of White (1980c), provided

$$\sqrt{n}\, S_n^{-1/2} R(\hat{\beta}_n - \beta_0) - n^{-1/2} S_n^{-1/2} R M_n^{-1} \sum_{t=1}^{n} X_t' \varepsilon_t \xrightarrow{p} 0, \tag{A.5}$$

$$|\hat{S}_n - S_n| \xrightarrow{p} 0, \tag{A.6}$$

where $\hat{S}_n = R(X'X/n)^{-1} \hat{B}_n (X'X/n)^{-1} R'$, and

$$n^{-1/2} \sum_{t=1}^{n} S_n^{-1/2} R M_n^{-1} X_t' \varepsilon_t \overset{a}{\sim} N(0, I_p). \tag{A.7}$$

(A.7) is easily verified by noting that the random variables $S_n^{-1/2} R M_n^{-1} X_t' \varepsilon_t$ are mixing, with size properties given by Assumption 3, and with the identity matrix as covariance matrix. Assumption 3 allows the application of Theorem 2.6 of Domowitz and White (1982), establishing (A.7).

(A.6) follows since $|(X'X/n)^{-1} - M_n^{-1}| \to 0$ a.s., and $|\hat{B}_n - B_n| \xrightarrow{p} 0$ by hypothesis, applying Lemma 3.2 of White (1980c).

(A.5) follows from (A.2), noting that the elements of $S_n^{-1/2} R$ are uniformly bounded.

The result then follows by applying Lemma 3.3 of White (1980c).

Proof of Theorem 3.3:[6] (a) From (3.1) and (3.2) in the text

$$\hat{B}_n - B_n = n^{-1} \sum_{t=1}^{n} \hat{\varepsilon}_t^2 X_t' X_t - E(\varepsilon_t^2 X_t' X_t) \tag{A.8}$$

$$+ n^{-1} \sum_{\tau=1}^{\ell} \sum_{t=\tau+1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-\tau} (X_t' X_{t-\tau} + X_{t-\tau}' X_t)$$

$$- E(\varepsilon_t \varepsilon_{t-\tau} [X_t' X_{t-\tau} + X_{t-\tau}' X_t])$$

when $E(\varepsilon_t \varepsilon_{t-\tau} | X_t, X_{t-\tau}) = 0$ for all $\tau > \ell$, ℓ finite. Now, $\varepsilon_t \varepsilon_{t-\tau}[X_t'X_{t-\tau}+X_{t-\tau}'X_t] = (Y_t - X_t\beta)(Y_{t-\tau} - X_{t-\tau})\beta)[X_t'X_{t-\tau}+X_{t-\tau}'X_t]$. It suffices to consider only

$$(Y_t - X_t\beta)(Y_{t-\tau} - X_{t-\tau}\beta)(X_t'X_{t-\tau}). \tag{A.9}$$

Under Assumption 1, (A.9) is equal to

$$[\varepsilon_t - X_t(\beta - \beta_0)][\varepsilon_{t-\tau} - X_{t-\tau}(\beta - \beta_0)]X_{it}X_{jt-\tau} \tag{A.10}$$

for $i,j \in \{1,\ldots,p\}$. Taking absolute values,

$$\begin{aligned} &|[\varepsilon_t - X_t(\beta-\beta_0)][\varepsilon_{t-\tau} - X_{t-\tau}(\beta-\beta_0)]\; X_{it}X_{jt-\tau}| \\ &\leq |[\varepsilon_t - X_t[\beta-\beta_0)]X_{it}|^2 \; |[\varepsilon_t - X_t(\beta-\beta_0)]\; X_{jt-\tau}|^2 \end{aligned} \tag{A.11}$$

Let β_k and β_{0k} be the k^{th} elements of β and β_0, respectively. An argument identical to that of White (1980b, Theorem 1(i)) establishes that there exists a $\overline{\beta} < \infty$ such that $(\beta_j - \beta_{0j})^2 < \overline{\beta}$, $j = 1,\ldots,p$. Using this fact, and by repeated application of the inequality $|a + b|^r \leq 2^{r-1}|a|^r + 2^{r-1}|b|^r$, $r \geq 1$, the expression in (A.11) is bounded above by

$$\begin{aligned} &[\gamma_0|\varepsilon_t X_{it}|^2 + \sum_{k=1}^{p} \gamma_k |X_{kt}X_{it}|^2 \overline{\beta}\,] \\ &\times[\mu_0|\varepsilon_{t-\tau}X_{jt-\tau}| + \sum_{k=1}^{p} \mu_k |X_{kt-\tau}X_{jt-\tau}|^2 \overline{\beta}] \end{aligned} \tag{A.12}$$

where γ_k, μ_k, $k=0,1,\ldots,p$, are finite constants. The expectation of each term is uniformly bounded under Assumptions 2-4. Applying the same argument to $(Y_t - X_t\beta)(Y_{t-\tau} - X_{t-\tau}\beta)(X'_{t-\tau}X_t)$ establishes that $\varepsilon_t\varepsilon_{t-\tau}[X'_tX_{t-\tau} + X'_{t-\tau}X_t]$ is dominated by functions that are uniformly $r_1+\delta$ -integrable. Application of Theorem 2.5 of Domowitz and White (1982) then ensures that

$$|(n-\tau)^{-1} \sum_{t=\tau+1}^{n} (\varepsilon_t\varepsilon_{t-\tau}[X'_tX_{t-\tau} + X'_{t-\tau}X_t] \tag{A.13}$$

$$-E(\varepsilon_t\varepsilon_{t-\tau}[X'_tX_{t-\tau} + X'_{t-\tau}X_t]))| \to 0 \text{ a.s.}$$

uniformly in β for $\tau = 0, 1, \ldots, \ell$. Since $(n-\tau)/n \to 1$ as $n \to \infty$, and $\hat{\beta}_n \to \beta_0$ a.s., it follows from Theorem 2.3 of Domowitz and White (1982) that

$$|n^{-1} \sum_{t=\tau+1}^{n} \hat{\varepsilon}_t\hat{\varepsilon}_{t-\tau} [X'_tX_{t-\tau} + X'_{t-\tau}X_t]$$

$$-E(\varepsilon_t\varepsilon_{t-\tau}[X'_tX_{t-\tau} + X'_{t-\tau}X_t])| \longrightarrow 0 \text{ a.s.}$$

for $\tau = 0, \ldots, \ell$. By Lemma 3.2 of White (1980c), it follows that $\hat{B}_n - B_n \xrightarrow{p} 0$.

(b) Define

$$B_n(\beta) = n^{-1} \sum_{t=1}^{n} (Y_t - X_t\beta)^2 X'_tX_t \tag{A.14}$$

$$+n^{-1} \sum_{\tau=1}^{\ell} \sum_{t=\tau+1}^{n} (Y_t - X_t\beta)(Y_{t-\tau} - X_{t-\tau}\beta)[X'_tX_{t-\tau} + X'_{t-\tau}X_t] \quad ,$$

and

$$\tilde{B}_n(\beta) = n^{-1} \sum_{t=1}^{n} E[(Y_t - X_t\beta)^2 X'_tX_t] \tag{A.15}$$

$$+ n^{-1} \sum_{\tau=1}^{\ell} \sum_{t=\tau+1}^{n} E[(Y_t - X_t\beta)(Y_{t-\tau} - X_{t-\tau}\beta)(X'_tX_{t-\tau} + X'_{t-\tau}X_t)].$$

It will first be shown that $\tilde{B}_n(\beta_0) - B_n \longrightarrow 0$. It is then demonstrated that $B_n(\beta) - \tilde{B}_n(\beta) \xrightarrow{a.s.} 0$, uniformly in β. Finally, it is shown that $\tilde{B}_n(\hat{\beta}_n) - \tilde{B}_n(\beta_0) \xrightarrow{p} 0$, yielding the desired result.

First,

$$\tilde{B}_n(\beta_0) - B_n = n^{-1} \sum_{\tau=\ell+1}^{n-1} \sum_{t=\tau+1}^{n} E(\varepsilon_t \varepsilon_{t-\tau}[X_t'X_{t-\tau} + X_{t-\tau}'X_t]) \quad .$$

It suffices to show that

$$n^{-1} \sum_{\tau=\ell+1}^{n-1} \sum_{t=\tau+1}^{n} |E(\varepsilon_t \varepsilon_{t-\tau} X_t'X_{t-\tau})| \to 0 \quad . \tag{A.16}$$

Assumption 2 ensures that $E(X_t'\varepsilon_t) = 0$. Since Assumption 5 ensures that $E|X_t\varepsilon_t|^{2+2\eta} < \infty$, where $\eta \equiv r_1 + \delta - 1$, it follows from Lemma 2.2 of White and Domowitz (1981) that either

$$|E(\varepsilon_t \varepsilon_{t-\tau} X_t'X_{t-\tau})| \leq c_1 \phi(\tau)^{1/2} \quad ,$$

or

$$|E(\varepsilon_t \varepsilon_{t-\tau} X_t'X_{t-\tau})| \leq c_2\, \alpha(\tau)^{\eta/(2+2\eta)}$$

where the c_i are henceforth finite constants. Therefore, either

$$n^{-1} \sum_{\tau=\ell+1}^{n-1} \sum_{t=\tau+1}^{n} |E(\varepsilon_t \varepsilon_{t-\tau} X_t'X_{t-\tau})| \leq c_1 \sum_{\tau=\ell+1}^{n-1} \phi(\tau)^{1/2} \quad , \tag{A.17}$$

or

$$n^{-1} \sum_{\tau=\ell+1}^{n-1} \sum_{t=\tau+1}^{n} |E(\varepsilon_t \varepsilon_{t-\tau} X_t'X_{t-\tau})| \leq c_2 \sum_{\tau=\ell+1}^{n-1} a(\tau)^{\eta/(2+2\eta)} . \tag{A.18}$$

To see that the right-hand sides of (A.17) and (A.18) converge to zero appropriately, write

$$\lim \sum_{\tau=\ell+1}^{n-1} \phi(\tau)^{1/2} = \lim \sum_{\tau=0}^{n-1} \phi(\tau)^{1/2} - \lim \sum_{\tau=0}^{\ell} \phi(\tau)^{1/2} \qquad (A.19)$$

$$\lim \sum_{\tau=\ell+1}^{n-1} \alpha(\tau)^{\eta/(2+2\eta)} = \lim \sum_{\tau=0}^{n-1} \alpha(\tau)^{\eta/(2+2\eta)} \qquad (A.20)$$

$$- \lim \sum_{\tau=0}^{\ell} \alpha(\tau)^{\eta/(2+2\eta)} .$$

The size requirements of the theorem ensure that $\Sigma_{\tau=0}^{\infty} \phi(\tau)^{1/2} < \infty$ and $\Sigma_{\tau=0}^{\infty} \alpha(\tau)^{\eta/(2+2\eta)} < \infty$. Since $\ell \to \infty$, (A.16) follows from (A.17) and (A.19) or (A.18) and (A.20). Thus, $\tilde{B}_n(\beta_0)-B_n \to 0$.

It is now shown that $B_n(\beta) = \tilde{B}_n(\beta) \xrightarrow{a.s.} 0$ uniformly in β. Let $d_{t\tau} = \varepsilon_t \varepsilon_{t-\tau} X_{it} X_{jt-\tau} - E(\varepsilon_t \varepsilon_{t-\tau} X_{it} X_{jt-\tau})$, noting that $d_{t\tau}$ is a function of β, but suppressing the extra notation as well as the i,j subscripts. By the triangle inequality, $B_n(\beta) - \tilde{B}_n(\beta) \xrightarrow{p} 0$ uniformly in β provided that for all i and j

$$|n^{-1} \sum_{t=1}^{n} d_{t0}| \xrightarrow{a.s.} 0, \text{ and} \qquad (A.21)$$

$$\sum_{=1}^{\ell} |n^{-1} \sum_{t=\tau+1}^{n} d_{t\tau}| \xrightarrow{a.s.} 0 \quad , \qquad (A.22)$$

uniformly in β, where $\ell = O(n^{\gamma})$, $0 < \gamma < \delta/(r_1+\delta)$.

To show (A.22) first note that

$$\sup_{\theta\varepsilon\Theta} \sum_{\tau=1}^{\ell} | n^{-1} \sum_{t=\tau+1}^{n} d_{t\tau}| \leqslant \sum_{\tau=1}^{\ell} \sup_{\theta\varepsilon\Theta} | n^{-1} \sum_{t=\tau+1}^{n} d_{t\tau}|.$$

The $d_{t\tau}$ are continuous uniformly in both t and τ. Assumption 4 and the Cauchy-

Schwartz inequality ensure that the $d_{t\tau}$ are dominated by $r_1 + \delta$ integrable functions for all τ. By Lemma A.3 of White and Domowitz (1981), the mixing requirements of Theorem 2.5 of Domowitz and White (1982) are satisfied for all $\tau=1,\ldots,=\ell$, such that for n large, $\sup_\theta|(n-\tau)^{\gamma-1} \sum_{t=\tau+1}^{n} d_{t\tau}| < \xi$, uniformly for $0 < \gamma < \delta/(r_1 + \delta)$, and it follows that

$$\sup_{\theta\varepsilon\Theta} \sum_{\tau=1}^{\ell} | n^{-1} \sum_{t=\tau+1}^{n} d_{t\tau}| < \ell\, n^{-\gamma}\, \xi$$

for n large and almost every sequence $\{X_t,\varepsilon_t\}$. The sets F_ℓ of sequences $\{X_t,\varepsilon_t\}$ such that $\sup_\theta |n^{\gamma-1} \sum_{t=1}^{n} d_{t\tau}| \not\longrightarrow 0$ for $\tau=1,\ldots,\ell$ given $\ell = 0(n^\gamma)$, $0 < \gamma < \delta/(r_1 + \delta)$, constitute an increasing sequence of sets of measure zero, such that $P(U_{\ell=1}^{\infty} F_\ell) = 0$, establishing (A.22). (A.21) holds as a consequence of (A.13).

From the mean value theorem of Jennrich (1969), and the triangle inequality, for i, j = 1,..., p, $|\tilde{B}_n^{ij}(\hat{\beta}_n) - \tilde{B}_n^{ij}(\beta_o)| \leq |\nabla \tilde{B}_n^{ij}(\hat{\beta}_n - \beta_o)| \leq \sum_{i=1}^{p} |\partial\tilde{B}_n^{ij}/\partial\beta_i| \; |\hat{\beta}_{ni} - \beta_{oi}|$, where β_n lies on the segment connecting $\hat{\beta}_n$ and β_o, and $\tilde{B}_n(\beta)$ is the matrix with elements $\tilde{B}_n^{ij}(\beta)$. Assumptions 4 and 5 ensure that $|\partial\tilde{B}_n^{ij}(\bar{\beta}_n)/\partial\beta_i| \leq \Delta\, n^\gamma + \Delta$, $\Delta < \infty$, for all i, so that
$|\tilde{B}_n^{ij}(\hat{\beta}_n) - \tilde{B}_n^{ij}(\beta_o)| \leq \Delta n^{\gamma-1/2} \sum_{i=1}^{p} |n^{1/2}(\hat{\beta}_{ni} - \beta_{oi})| + \Delta \sum_{i=1}^{p} |\hat{\beta}_{ni} - \beta_{oi}|$.
Since $\hat{\beta}_n \xrightarrow{a.s.} \beta_o$ and $n^{1/2}(\hat{\beta}_n - \beta_o)$ is $O_p(1)$ as a consequence of the asymptotic normality result, and since $\gamma < 1/2$, the left-hand side of the last inequality is $o_p(1)$ by 2c.4 (x.a) and 2c.4 (xiii) of Rao (1973). Therefore,
$\tilde{B}_n(\tilde{\beta}_n) - \tilde{B}_n(\beta_o) \xrightarrow{p} 0$.

Since $\hat{B}_n - \tilde{B}_n(\hat{\beta}_n) \xrightarrow{a.s.} 0$, $\tilde{B}_n(\hat{\beta}_n) - \tilde{B}_n(\beta_o) \xrightarrow{p} 0$, and $\tilde{B}_n(\beta_o) - B_n \longrightarrow 0$, it follows that $\hat{B}_n - B_n \xrightarrow{p} 0$.

Proof of Theorem 4.1: Define $\beta_n^* = M_n^{-1}[n^{-1}\Sigma_{t=1}^n E(X_t' \, g_t(Z_t))]$. A slight modification of the argument in White (1980a, Theorem 2) establishes that β_n^* is identifiably unique (Domowitz and White (1982), Definition 2.1] relative to $\bar{\sigma}_n^2$. Given Assumption 1',

$$\hat{\beta}_n = (X'X/n)^{-1}[n^{-1}\sum_{t=1}^{n} X_t' g_t(Z_t)].$$

The X_t are measurable functions of the Z_t, and by Lemma 2.1 of White and Domowitz (1981), the X_t are mixing with size properties given in Assumption 2'. As in the proof of Theorem 2.1, $(X'X/n)^{-1} - M_n^{-1} \to 0$ a.s.

Since the $g_t(Z_t)$, $t = 1,2,\ldots,$ are measurable functions of Z_t, it follows similarly that $|n^{-1}\sum_{t=1}^n [X_t' g_t(Z_t) - E(X_t' g_t(Z_t))]| \to 0$, a.s. The result follows directly.

Proof of Theorem 4.2: The proof proceeds completely analogously to that of Theorem 2.1, noting that the weights are positive functions taking values on a compact interval, and are mixing with the size properties of the Z_t by Lemma 2.1 of White and Domowitz (1981).

Proof of Theorem 4.3: Consider the quantity $n^{-1/2}\Sigma_{t=1}^n X_t' u_t$, where $u_t \equiv g_t(Z_t) - X_t\beta_n^*$. Assumption 3' allows the application of Theorem 2.6 of Domowitz and White (1981), yielding $\sqrt{n}\, B_n^{*\,-1/2}\Sigma_{t=1}^n X_t' u_t \overset{a}{\sim} N(0, I_p)$. An argument similar to that in the proof of Theorem 3.2 completes the proof.

Proof of Theorem 5.3: The proof is identical to that of Theorem 2, Corollary 1, of White (1980b), except that Theorem 2.6 of Domowitz and White (1982) is applied instead of the Liapounov central limit theorem, Theorem 2.5 of Domowitz and White (1982) is applied instead of Lemma 2.3 of White (1980c), and the Markov strong law of large numbers is replaced by Theorem 2.10 of McLeish (1975).

Proof of Theorem 5.4: For $\tau \geqslant 1$, define

$$D_{n\tau}=(n-\tau)^{-1}\sum_{t=\tau+1}^{n} \psi'_{t\tau}(X_t-X_t\beta)(Y_{t-\tau}-X_{t-\tau}\beta) \tag{A.29}$$

where $\psi_{t\tau}$ is the $1xk_\tau$ vector with elements $X_{it}X_{jt-\tau}$, $i,j \in \{1,\ldots p\}$, and where it is recognized that $D_{n\tau}$ is a function of β, suppressing the extra notation. Let $\hat{D}_{n\tau} = \hat{D}_{n\tau}(\hat{\beta}_n)$ and $D^o_{n\tau} = D_{n\tau}(\beta_0)$. If

$$(n-\tau)^{1/2}V_\tau^{-1/2}D^o_{n\tau} \overset{a}{\sim} N(0,I_{k_\tau}), \tag{A.30}$$

$$|\hat{V}_{n\tau}-V_\tau| \xrightarrow{P} 0 \text{ for some } \hat{V}_{n\tau}, \text{ and} \tag{A.31}$$

$$|(n-\tau)^{1/2}V_\tau^{-1/2}[\hat{D}_{n\tau}-D^o_{n\tau}]| \xrightarrow{P} 0 , \tag{A.32}$$

then White's (1980c) Lemma 3.3 may be applied to establish

$$(n-\tau)\hat{D}_{n\tau}\hat{V}_{n\tau}^{-1}\hat{D}_{n\tau} \overset{a}{\sim} \chi^2_{k_\tau} \tag{A.33}$$

and the desired result follows by showing that (A.33) is $(n-\tau)R_0^2$ from the artificial regression (5.2).

Consider (A.30). By definition, $(n-\tau)^{1/2}D^o_{n\tau} = (n-\tau)^{-1/2}\Sigma_{t=\tau+1}^{n}\Psi'_{t\tau}\varepsilon_t\varepsilon_{t-\tau}$.

An argument as in the proof of Theorem 3.4 establishes that the elements of the summand are appropriately dominated, under Assumptions 3 and 6. Given Assumption 7, the average covariance matrix is $V_\tau = (n-\tau)^{-1}\kappa_\tau \cdot \Sigma_{t=\tau+1}^{n} E(\psi'_{t\tau}\psi_{t\tau})$. Lemma 2.1 of White and Domowitz (1981), together with the mixing conditions of Assumption 3, then allow the application of Theorem 2.6 of Domowitz and White (1982), establishing $(n-\tau)^{1/2} V_\tau^{-1/2} D^o_{n\tau} \overset{a}{\sim} N(0, I_{k_\tau})$.

Now consider (A.31). An estimator for V_τ is given by $\hat{V}_{n\tau} = (n-\tau)^{-1}\hat{\kappa}_\tau \Sigma_{t=\tau+1}^{n} \psi_{t\tau}\psi_{t\tau}$, where $\hat{\kappa}_\tau = (n-\tau)^{-1} \Sigma_{t=\tau+1}^{n} \ (Y_t - X_t\hat{\beta}_n)^2 (Y_{t-\tau} - X_{t-\tau}\hat{\beta}_n)^2$. It is straightforward to verify that appropriate dominating functions exist under Assumption 6 such that Theorem 2.5 of Domowitz and White (1982) may be applied, ensuring that

$$\Big|(n-\tau)^{-1} \sum_{t=\tau+1}^{n} [(Y_t - X_t\beta)^2 (Y_{t-\tau} - X_{t-\tau}\beta)^2$$

$$-E[(Y_t - X_t\beta)^2 (Y_{t-\tau} - X_{t-\tau}\beta)^2]]\Big| \longrightarrow 0 \text{ a.s.}$$

uniformly in β. Under the assumptions, Theorem 2.10 of McLeish (1975) may be applied to show that

$$(n-\tau)^{-1} \sum_{t=\tau+1}^{n} [\psi'_{t\tau}\psi_{t\tau} - E(\psi'_{t\tau}\psi_{t\tau})] \longrightarrow 0 \text{ a.s.}$$

Since $\hat{\beta}_n \longrightarrow \beta_0$ a.s., Theorem 2.3 of Domowitz and White (1982) may be applied, and it follows that $|\hat{V}_{n\tau} - V_\tau| \longrightarrow 0$ a.s.

In order to establish (A.32), a mean value expansion is applied to $D_{n\tau}$. The assumptions allow the application of Lemma 3 of Jennrich (1969), a mean value theorem for random functions. The expansion applies to a sequence tail-equivalent to $\hat{\beta}_n$, but to simplify things, the notation will not be changed. With this proviso,

$$(n-\tau)^{1/2}\hat{D}_{n\tau} = (n-\tau)^{1/2} D^o_{n\tau} + (n-\tau)^{1/2} \nabla\overline{D}_{n\tau}(\hat{\beta}_n - \beta_0)$$

where $\nabla\bar{D}_{n\tau}$ is the Jacobian of $D_{n\tau}$ with columns i=1,...,p, evaluated at suitable mean values $\bar{\beta}_n^{(i)}$ lying on the segment connecting $\hat{\beta}_n$ and β_0. Rearranging,

$$|(n-\tau)^{1/2}[\hat{D}_{n\tau}-D^o_{n\tau}]| = |((n-\tau)/n)^{1/2}\sqrt{n}\ \nabla\bar{D}_{n\tau}(\hat{\beta}_n - \beta_0)| \ . \quad \text{(A.34)}$$

Now, $|\sqrt{n}\ (\hat{\beta}_n-\beta_0)|$ is bounded in probability by Theorem 3.1, which holds under the assumptions given. Also, $(n-\tau)/n \to 1$ as $n \to \infty$. Taking the indicated derivatives,

$$\partial\bar{D}^{(j)}_{n\tau}\partial\beta_j = -(n-\tau)^{-1}\sum_{t=\tau+1}^{n} \psi'_{t\tau}[(Y_t-X_t\bar{\beta}^{(j)}_n)X_{jt-\tau} + (Y_{t-\tau} - X_{t-\tau}\bar{\beta}^{(j)}_n)X_{jt}]$$

where $\bar{\beta}_n^{(j)}$ is the appropriate mean value. Assumption 6 ensures that $|\psi'_{t\tau}(X_t-X_t\beta)X_{jt}|$ is appropriately dominated, so that Theorem 2.5 of Domowitz and White (1982) may be applied to establish that

$$|(n-\tau)^{-1}\sum_{t=\tau+1}^{n}[\psi'_{t\tau}[(Y_t-X_t\beta)X_{jt-\tau}+(Y_{t-\tau}-X_{t-\tau}\beta)X_{jt}]$$

$$-E(\psi'_{t\tau}[(Y_t-X_t\beta)X_{jt-\tau} + (Y_{t-\tau}-X_{t-\tau}\beta)X_{jt}])] \to 0 \text{ a.s.},$$

uniformly in β. Since $\hat{\beta}_n \to \beta_0$ a.s., Theorem 2. 3 of Domowitz and White (1982) may be applied to establish that

$$\partial\bar{D}^{(j)}_{n\tau}/\partial\beta_j-E[\partial D^{(j)}_{n\tau}/\partial\beta_j] \to 0 \text{ a.s., } j=1,\dots,p.$$

Finally, given Assumption 7, the expectations above vanish, so that $\partial\bar{D}^{(j)}_{n\tau}/\partial\beta_j \to 0$, a.s., j=1,...,p. The symmetric positive definite matrix square root $V_\tau^{-1/2}$ is well defined for n large, and has uniformly bounded elements under Assumptions 6 and 7. (A.32) therefore follows from (A.35).

Lemma 3.3 of White (1980c) establishes (A.33), given (A.30), (A.31), and (A.32). From equation (5.2) of the text, the OLS estimates of α from the artifical regression are

$$\hat{\alpha} = \left[(n-\tau)^{-1} \sum_{t=\tau+1}^{n} \psi'_{t\tau}\, \psi_{t\tau}\right]^{-1}\left[(n-\tau)^{-1} \sum_{t=\tau+1}^{n} \psi'_{t\tau}\, \hat{\varepsilon}_t \hat{\varepsilon}_{t=\tau}\right]$$

where $\hat{\varepsilon}_t = Y_t - X_t\hat{\beta}_n$. Thus,

$$\hat{D}'_{n\tau}\hat{V}^{-1}_{n\tau}\, \hat{D}_{n\tau} = \hat{\alpha}' \left[(n-\tau)^{-1} \sum_{t=\tau+1}^{n} \psi'_{t\tau}\psi_{t\tau}\right] \hat{\alpha}/\kappa_\tau$$

which is R_0^2 (Theil (1971, p. 164)] for the regression (5.2). The result follows.

Proof of Theorem 5.5: The argument is identical to that of Theorem 4 of White (1980a), except that Komolgorov's strong law of large numbers is replaced by Theorem 2.10 of McLeish (1975), provided that $|\hat{S}^{-1} - S^{-1}| \xrightarrow{p} 0$. The latter follows under the assumptions given, if $|\hat{S}_n - S| \xrightarrow{p} 0$. This is demonstrated by applying the argument of Theorem 3.4 to $\hat{B}_{1n}$, $\hat{B}_{2n}$, and $\hat{R}_n$, noting that the weights in all cases are positive, measurable, mixing random variables taking values on a compact interval.

FOOTNOTES

[1]In what follows, the qualifiers "all" and "sufficiently" in such statements will be implicitly understood.

[2]Economists usually think of σ- algebras as information sets. We write $E(X_t|\mathfrak{F}_{t-1})=E(X_t|X_{t-1},X_{t-2},\ldots)$ to mean the expectation of X_t given information in the form of lagged X's available at time t-1, for example. Formally, a σ- algebra is a collection $\mathcal{B}$ of subsets (events) of a set (sample space) Ω such that: (i) ϕ and Ω belong to $\mathcal{B}$; (ii) if B belong to $\mathcal{B}$, then B^c belong to $\mathcal{B}$; (iii) if $\{B_n\}$ is a sequence of sets in, then $\cup_{n=1}^{\infty} B_n$ belongs to $\mathcal{B}$. The reader may consult Billingsley (1979) for further details.

[3]General nonlinear hypotheses $h(\beta_0) = 0$, where h is a continuously differentiable function in β with bounded Jacobian, ∇h, having full row rank (at β_0) may be handled completely analogously. Asymptotically, ∇h is like R.

[4]In this case, limiting results also follow by assuming the errors to be ℓ- dependent (cf. Billingsley (1968, page 167)), an assumption also used extensively by Anderson (1971).

[5]The mathematical aspects of the linear least squares approximation are rigorously discussed in Rice (1969), chapters 2 and 12, and in Sard (1963).

[6]I am indebted to Halbert White for suggesting this line of proof.

REFERENCES

Anderson, T. W., 1971, The Statistical Analysis of Time Series (Wiley, New York).

Anderson, T. W. and John B. Taylor, 1979, Strong Consistency of Least Squares Estimates in Dynamic Models, Annals of Statistics 3, 484-489.

Bierens, Herman J., 1970, Robust Methods and Asymptotic Theory in Nonlinear Econometrics (Springer-Verlag, Berlin).

Billingsley, P., 1979, Probability and Measure (Wiley, New York).

Burguete, J., A. R. Gallant, and G. Souza, 1982, On Unification of the Asymptotic Theory of Nonlinear Econometric Models, Econometric Reviews, forthcoming.

Chow, G. C. 1981, Selection of Econometric Models by the Information Criterion, in: E. G. Charatsis, ed., Proceedings of the Econometric Society European Meeting 1979 (North Holland, Amsterdam).

Christopeit, N., and K. Helmes, 1980, Strong Consistency of Least Squares Estimators in Linear Regression Models, Annals of Statistics 4, 778-788.

Crowder, Martin J., 1980, On the Asymptotic Properties of Least-Squares Estimators in Autoregression, Annals of Statistics 8, 132-146.

Domowitz, Ian, and Halbert White, 1982, Misspecified Models with Dependent Observations, Annals of Econometrics, forthcoming.

Eicker, F., 1967, Limit Theorems for Regressions with Unequal and Dependent Errors, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1 (University of California Press, Berkeley).

Goodwin, G. C. and R. L. Payne, 1977, Dynamic System Identification (Academic Press, New York).

Hannan, E. J., 1973, The Asymptotic Theory of Linear Time-Series Models, Journal of Applied Probability 1, 130-145.

Hansen, L. P., 1982, Large Sample Properties of Generalized Method of Moments Estimators, Econometrica 50, 1029-1054.

Hausman, J. A., 1978, Specification Tests in Econometrics, Econometrica 46, 1251-1272.

Hendry, D. F., 1979, The Behavior of Inconsistent Instrumental Variables Estimators in Dynamic Systems with Autocorrelated Errors, Journal of Econometrics 9, 295-314.

Ibragimov, I. A., and Y. V. Linnik, 1971, Independent and Stationary Sequences of Random Variables (Wolters-Noordhoff, The Netherlands.)

Jennrich, Robert I., 1969, Asymptotic Properties of Non-linear Least Squares Estimators, Annals of Mathematical Statistics 40, 633-643.

Kohn, R., 1979, Asymptotic Estimation and Hypothesis Testing Results for Vector Linear Time Series Models, Econometrica 47, 1005-1030.

Kampe de Feriet, J. and F. N. Frenkiel, 1962, Correlation and Spectra for Nonstationary Random Functions, Mathematical Computation 16, 1-21.

Lukacs, Eugene, 1975, Stochastic Convergence (Academic Press, New York).

McLeish, D. L., 1975, A Maximal Inequality and Dependent Strong Laws, Annals of Probability 3, 826-836.

Nicholls, D. F., and A. R. Pagan, 1982, Heteroscedasticity in Models with Lagged Dependent Variables, mimeo, Australian National University.

Parzen, E., 1962, Spectral Analysis of Asymptotically Stationary Time Series, Bulletin de L'Institut International de Statistique 39, 87-103.

Pierce, D. A., 1972, Least Squares Estimation in Dynamic Disturbance Time Series Models, Biometrika 52, 73-78.

Rao, C. R., 1973, Linear Statistical Inference and its Applications (Wiley, New York).

Rao, M. M., 1978, Covariance Analysis of Nonstationary Time Series, in: P. R. Krishnaiah, ed., Developments in Statistics Volume 1 (Academic Press, New York).

Rice, John, 1979, The Approximation of Functions (Addison-Wesley, Massachusetts).

Rosenblatt, M., 1956, A Central Limit Theorem and a Strong Mixing Condition, Proceedings of the National Academy of Science, U.S.A. 42, 43-47.

Rosenblatt, M., 1972, Uniform Ergodicity and Strong Mixing, Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandete Gebiete 24, 79-84.

Sard, Arthur, 1963, Linear Approximation (American Mathematical Society, Providence).

Scott, D. J., 1973, Central Limit Theorems for Martingales and for Processes with Stationary Increments Using a Skorokhod Representation Approach, Advances in Applied Probability 5, 118-136.

Sims, C. A., 1978, Least-Squares Estimation of Autoregressions with Some Unit Roots, University of Minnesota, Department of Economics, Discussion Paper 78-95.

Stout, William F., 1974, Almost Sure Convergence (Academic Press, New York).

Theil, H., 1971, Principles of Econometrics (John Wiley and Sons, Inc., New York).

White, Halbert, 1980a, Using Least Squares to Approximate Unknown Regression Functions, International Economic Review 21, 149-170.

White, Halbert, 19980b, A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity, Econometrica 48, 817-838.

White, Halbert, 1980c, Nonlinear Regression on Cross-Section Data, Econometrica 48, 721-746.

White, Halbert, 1981, Consequences and Detection of Misspecified Nonlinear Regression Models, Journal of the American Statistical Association 76, 419-433.

White, Halbert, 1982a, Maximum Likelihood Estimation of Misspecified Models, Econometrica 50, 1-25.

White, Halbert, 1982b, Instrumental Variables Regression with Independent Observations, Econometrica 50, 483-499.

White, Halbert, and I. Domowitz, 1981, "Nonlinear Regression with Department Observations, UCSD Economics Department Working Paper 81-32.

Wu, De-Min, 1973, Alternative Tests of Independence Between Stochastic Regressors and Disturbances, Econometrica 41, 733-750.