INTERNATIONAL ECONOMIC REVIEW
Vol. 23, No. 3, October, 1982

DIFFERENCING AS A TEST OF SPECIFICATION*

BY CHARLES I. PLOSSER, G. WILLIAM SCHWERT AND HALBERT WHITE[1]

1. INTRODUCTION

One of the most difficult problems in applied econometrics is to determine when an econometric model is correctly specified. Although there are many types of misspecification, this paper is concerned with specification errors that lead to biased or inconsistent estimators of the regression coefficients. Specification errors of this type include omitted variables, measurement error, simultaneous equation problems or any type of misspecification that leads to a dependence between the error term and the regressors.

In the context of the linear time series regression model, Plosser and Schwert [1977, 1978] suggest that a comparison of the least squares estimators from a presumably correctly specified regression equation with the estimators from the differenced form of the same model can provide important insights into the validity of the linear regression specification. If the model is correctly specified, the estimators from the differenced and undifferenced models have the same probability limit, so the results should corroborate one another. On the other hand, if there are measurement errors, omitted variables, or other types of misspecification, the differences regression should lead to different results, indicating that corrective measures are in order.

The purpose of this paper is to formalize the arguments presented in Plosser and Schwert by constructing a test of the hypothesis that a time series regression equation is well-specified. The test is based on a comparison of the least squares estimators obtained from the differenced and undifferenced regressions. The resulting specification test is similar to those proposed by Hausman [1978].

Section 2 develops a statistic to test whether the undifferenced and differenced regressions are equal. A rejection of this test would indicate misspecification. In Section 3, we investigate the asymptotic local power of the test. Section 4 summarizes Monte Carlo experiments that investigate the size and power of our test in finite samples under some plausible types of misspecification. In addition, we compare the power of our test with the power of an instrumental variables procedure and test proposed by Wu [1973], the frequently used Durbin-Watson

* Manuscript received February 4, 1981; revised March 12, 1982.

[1] We have benefited from the comments of Eugene Fama, Martin Geisel, Michael Gibbons, Clive Granger, Charles Nelson, Adrian Pagan, Christopher Sims, and from anonymous referees. We also received valuable comments from workshop participants at the University of Chicago, Northwestern University, and the N.B.E.R./N.S.F. Seminar on Time Series. Schwert's participation is supported by the National Science Foundation.

statistic, and a specification test due to Ramsey [1969] and Ramsey and Schmidt [1976]. The concluding section summarizes our test procedures and the findings of the power investigation.

2. A DIFFERENCING TEST FOR MISSPECIFICATION

The stochastic regression model we consider is discussed by Goldberger [1964, ch. 6]. Assume

A1. The model can be written in mean-deviation form as

$$y_t = X_t \beta_0 + \varepsilon_t, \qquad t = 1,\ldots,T, \tag{1}$$

where $\{X_t\}$ is a weakly stationary sequence of random $1 \times k$ row vectors, such that plim $T^{-1} \sum_{t=1}^{T} X_t' X_t = M_{XX} \equiv E(X_t' X_t)$ is finite and nonsingular; $\{\varepsilon_t\}$ is a pure white noise sequence, such that $E(\varepsilon_t)=0$, $E(\varepsilon_t^2)=\sigma_0^2$, $0<\sigma_0^2<\infty$; ε_t is independent of X_τ for all τ; and β_0 is an unknown finite $k \times 1$ vector of constants.

When the model is correctly specified, the least squares estimator $\hat{\beta}_T = (X'X)^{-1}X'y$, is a consistent estimator for β_0. Given A1 and assumption A2 of the Appendix (which imposes mild conditions on the moments of the regressor cross-products and on the convergence of the regressor covariances given the distant past to the unconditional covariance), it is straightforward to show that $\sqrt{T}(\hat{\beta}_T - \beta_0) \overset{A}{\sim} N(0, \sigma_0^2 M_{XX}^{-1})$. However, when the regression equation (1) is misspecified, in the sense that there are omitted variables, simultaneous equation problems, etc., there is a dependence between ε_t and X_t, and the least squares estimator is generally inconsistent.

In order to test for misspecification in (1), consider the differenced regression

$$\dot{y}_t = \dot{X}_t \beta_0 + \dot{\varepsilon}_t, \qquad t = 1,\ldots, T, \tag{2}$$

where the dot notation indicates first differencing (e.g., $\dot{y}_t = y_t - y_{t-1}$). For convenience, set $y_0 \equiv y_1$, $X_0 \equiv X_1$, and $\varepsilon_0 \equiv \varepsilon_1$. Defining $\dot{X}$ as the $T \times k$ matrix with rows X_t and y as the $T \times 1$ vector with elements y_t, the least squares estimator for (2) is $\tilde{\beta}_T = (\dot{X}'\dot{X})^{-1}\dot{X}'\dot{y}$. When the model is correctly specified, $\tilde{\beta}_T$ is also a consistent (although inefficient) estimator for β_0, provided that X_t contains no lagged dependent variables. Otherwise, $\tilde{\beta}_T$ is inconsistent. The inefficiency of the estimator arises from the fact that differencing induces a first-order moving average process in the disturbances of (2).[2]

Thus, the difference $\tilde{\beta}_T - \hat{\beta}_T$ provides a misspecification indicator similar to those proposed by Hausman [1978]. When the model is correctly specified, $\tilde{\beta}_T - \hat{\beta}_T$ will be close to zero; however, $\tilde{\beta}_T - \hat{\beta}_T$ will generally diverge if the model is misspecified. Based on this fact, we derive a statistic that uses the difference between $\tilde{\beta}_T$ and $\hat{\beta}_T$ to test for the presence of misspecification.

The statistic is based on the joint asymptotic normality of $\tilde{\beta}_T$ and $\hat{\beta}_T$.[3] If

[2] See Plosser and Schwert [1977] for a discussion of this moving average process and its effects on estimates of β_0.

$\sqrt{T}(\tilde{\beta}_T - \hat{\beta}_T)$ is asymptotically normal with mean zero and nonsingular covariance matrix V (which occurs under general conditions when the model is correctly specified), it follows that

$$\Delta \equiv T(\tilde{\beta}_T - \hat{\beta}_T)' \hat{V}^{-1} (\tilde{\beta}_T - \hat{\beta}_T) \overset{A}{\sim} \chi^2_k, \tag{3}$$

where $\hat{V}$ is a consistent estimator of V. Formally, we have

THEOREM 1. *If X_t contains no lagged endogenous variables and $A1$, $A2$, and $A6$ hold, then $\Delta \overset{A}{\sim} \chi^2_k$, where*

$$\hat{V} = \hat{\sigma}^2_T[(\dot{X}'\dot{X}/T)^{-1}(\ddot{X}'\ddot{X}/T)(\dot{X}'\dot{X}/T)^{-1} - (X'X/T)^{-1}]$$

and $\hat{\sigma}^2_T = (y - X\hat{\beta}_T)'(y - X\hat{\beta}_T)/T$, provided plim $\hat{V}$ *is nonsingular.*

The double dot notation indicates second differencing ($\ddot{X}_t = \dot{X}_t - \dot{X}_{t-1}$). Assumptions A2 and A6 are also given in the Appendix, along with all mathematical proofs. The particularly simple form for $\hat{V}$ (which is just the difference of the estimated covariance matrices for $\tilde{\beta}_T$ and $\hat{\beta}_T$)[4] arises from the fact that $\hat{\beta}_T$ is asymptotically efficient relative to $\tilde{\beta}_T$.[5]

Although Theorem 1 is sufficient for many applied problems, often econometricians are faced with more complex models that involve lagged dependent variables, or autocorrelated errors. Hence, it is important to modify the differencing test to allow for such complications.

The only difficulty presented by lagged dependent variables in X_t is that $\tilde{\beta}_T$ becomes inconsistent even when (1) is correctly specified.[6] To remedy this problem, consider the instrumental variables estimator $\tilde{\tilde{\beta}}_T = (Z'\dot{X})^{-1}Z'\dot{y}$, where Z is a $T \times k$ matrix of instruments with properties fully specified by assumptions A3–A5 of the Appendix.[7] A natural choice for Z is a linear combination of

[3] It should be noted that Hausman's [1978] Lemma 2.1 implicitly requires the joint asymptotic normality of $\tilde{\beta}_T$ and $\tilde{\tilde{\beta}}_T$. Hence, it will not generally suffice to consider the asymptotic normality of $\tilde{\beta}_T$ and $\tilde{\tilde{\beta}}_T$ separately, and then apply Hausman's Lemma. In the proof of our Theorem 2, it becomes evident that although asymptotic normality can be proven separately for $\tilde{\beta}_T$ and $\tilde{\tilde{\beta}}_T$ using assumptions A1–A4, an additional assumption, A5, is required for joint asymptotic normality.

[4] Note that $\hat{\sigma}^2_T(\dot{X}'\dot{X}/T)^{-1}(\ddot{X}'\ddot{X}/T)(\dot{X}'\dot{X}/T)^{-1}$ is just the estimated covariance matrix for $\tilde{\beta}_T$ taking into account that the disturbances follow a first-order moving average process with a unit root.

[5] It is also worth noting that there is an exact F-test that can be used instead of Δ when the errors are normal and the regressors are non-stochastic or independent of ε.

[6] In the lagged dependent variable case some of the regressors are correlated with lagged errors so assumption A1 is not satisfied. A similar problem occurs in rational expectations predictive equations when the regressors (measures of expectations) will generally be correlated with lagged errors.

[7] Alternately, Adrian Pagan has pointed out to us that the inconsistency in $\tilde{\beta}_T$ can be estimated and removed. A consistent estimator is $\tilde{\beta}_T + (\dot{X}'\dot{X})^{-1}\hat{S}$, where $\hat{S}$ is a $k \times 1$ vector with $T\hat{\sigma}^2_T$ in the position corresponding to the lagged dependent variable and zeros elsewhere. This is

(*Continued on next page*)

current and lagged exogenous variables. For this more general case we have the following:

THEOREM 2. *Given A1–A5,* $\Delta \overset{A}{\sim} \chi^2_k$, *where*

$$\hat{V} = \hat{\sigma}^2_T[(Z'\dot{X}/T)^{-1}(\dot{Z}'\dot{Z}/T)(\dot{X}'Z/T)^{-1} - (X'X/T)^{-1}]$$

and $\hat{\sigma}^2_T = (y - X\hat{\beta}_T)'(y - X\hat{\beta}_T)/T$, *provided* plim $\hat{V}$ *is nonsingular.*
Note that Theorem 1 can be viewed as a simple corollary of Theorem 2 where $Z = \dot{X}$.

Next we allow for autocorrelated errors by considering more general models of the form

$$y^*_t = X^*_t\beta_0 + \varepsilon^*_t, \tag{4}$$

where ε^*_t is an *ARMA* (p, q) process $\phi(B)\varepsilon^*_t = \theta(B)\varepsilon_t$, and B is the backshift operator. If $\phi(B)$ and $\theta(B)$ have roots outside the unit circle, the model can always be written $y_t = X_t\beta_0 + \varepsilon_t$, where $y_t \equiv \phi(B)\theta(B)^{-1}y^*_t$, and $X_t = \phi(B)\theta(B)^{-1}X^*_t$. Obviously, no difficulty arises in satisfying A1 if $\phi(B)$ and $\theta(B)$ are known, so no generality is lost in this case by supposing ε_t to be white noise. In practice, however, one has at most a knowledge of p and q, so that the parameters of $\phi(B)$ and $\theta(B)$ must be estimated. Nevertheless, results of Pierce [1972] for a similar model strongly suggest that this estimation may be carried out without affecting the covariance matrices embedded in the statistics proposed here. This arises from the fact that the least squares estimators of the parameters of $\phi(B)$ and $\theta(B)$ are independent of the estimators of β_0 in large samples.

Thus, for a fairly general class of models, the differencing test can be implemented as follows:

(1) Estimate β_0 along with any ARMA parameters for the error process by least squares. If necessary, transform the model so that the errors are white noise, and retain the covariance matrix estimator for β_0, $\hat{\sigma}^2_T(X'X)^{-1}$.

(2) Difference the transformed model and compute $\tilde{\beta}_T$ and the remaining terms of the appropriate form for $\hat{V}$ based on the cross-products matrix of the instrumental variables and the first differences of the regressors, $(Z'\dot{X}/T)$, and the cross-products matrix of the first difference of the instrumental variables, $(\dot{Z}'\dot{Z}/T)$. When there are no lagged dependent variables, this only requires the computation of the cross-products matrices for the first differences $(\dot{X}'\dot{X}/T)$ and the second differences $(\ddot{X}'\ddot{X}/T)$ of the regressors.

(Continued)

asymptotically equivalent to an instrumental variables procedure with $Z \equiv \dot{X} + \hat{U}$, where $\hat{U}$ has the estimated residual from the undifferenced regression in the column corresponding to the lagged dependent variable and zeros elsewhere. The presence of estimated rather than true residuals may affect the form of $\hat{V}$ in Theorem 2 below.

(3) Finally, compute Δ and compare it to the critical value for a χ_k^2 statistic at the desired significance level. If Δ exceeds this value, reject the null hypothesis that the model is correctly specified.

3. ASYMPTOTIC LOCAL POWER OF THE DIFFERENCING TEST

For large samples, the local power of the differencing test follows immediately from Hausman's [1978] Theorem 2.2. Applied to the present case, Hausman's result states for a sequence of local alternatives such that plim $\sqrt{T}(\tilde{\beta}_T - \hat{\beta}_T) = \delta$, where δ is a finite non-zero $k \times 1$ vector, the Δ statistic is distributed asymptotically as noncentral $\chi_{k,v}^2$ with noncentrality parameter

$$v = \delta'(\text{plim } \hat{V}^{-1})\delta. \tag{5}$$

It is trivial to show that the differencing test is consistent (i.e., has power one asymptotically) against any fixed alternative such that plim $(\tilde{\beta}_T - \hat{\beta}_T) \neq 0$ and plim $\hat{V}$ is nonsingular. The local asymptotic power is

$$\pi(v) = P[\Delta > {}_\alpha\chi_k^2 \mid \Delta \sim \chi_{k,v}^2],$$

where ${}_\alpha\chi_k^2$ is the critical value for a test of size α under the χ_k^2 distribution. The power function $\pi(v)$ is increasing in v, so we shall use v to index the local power of a test. To gain more understanding, we consider several special cases.

A common source of misspecification occurs when elements of the regressor matrix contain measurement error. For example, suppose the true model is

$$y_t = \beta_0 z_t + u_t, \tag{6}$$

where for simplicity z_t and β_0 are now scalars. The model estimated is

$$y_t = \beta_0 x_t + \varepsilon_t, \tag{7}$$

where the regressor contains measurement error, $x_t = z_t + v_t$ so that $\varepsilon_t = u_t - \beta_0 v_t$. This is the textbook errors-in-variables model.

If we assume that z_t is a first-order autoregressive process, AR(1), and v_t is uncorrelated with z_t, u_t and its own past values, and we consider the sequence of local alternatives such that $\sigma_v^2 = \sigma_v^{*2}/\sqrt{T}$, the expression (5) becomes

$$v = 2\beta_0^2\rho_z^2\sigma_v^{*4}/(\sigma_u^2\sigma_z^2(1-\rho_z^2)). \tag{8}$$

where $\sigma_u^2 = E(u_t^2)$, $E(u_t) = 0$, $\sigma_z^2 = E(z_t^2)$, and $\rho_z = E(z_t z_{t-1})/\sigma_z^2$. The local power of the test depends directly on β_0^2 and ρ_z^2 and inversely on $\sigma_u^2\sigma_z^2$.

Another common source of misspecification occurs when relevant variables, which are correlated with the included variables, are omitted from the regression. For example, let the true model be

$$y_t = \beta_0 x_t + \gamma_0 z_t + u_t, \tag{9}$$

where u is independent of x and z. The misspecified model is

(10) $$y_t = \beta_0 x_t + \varepsilon_t,$$

where $\varepsilon_t = u_t + \gamma_0 z_t$.

If we consider the sequence of local alternatives $\gamma_0 = \gamma_0^*/\sqrt{T}$, the expression (5) becomes

(11) $$\nu = 2\gamma_0^{*2}\sigma_z^2(\rho_{xz}\rho_x - [\rho_{xz}(-1) + \rho_{xz}(+1)]/2)^2/\sigma_u^2(1-\rho_x^2)$$

where x_t is AR(1), $\rho_x = E(x_t x_{t-1})/\sigma_x^2$, $\rho_{xz}(k) = \text{corr}\,(x_t, z_{t+k})$, and the other notation is as previously defined. The local power of the test is more complicated than for the errors-in-variables model. It depends directly on γ_0^{*2}, ρ_{xz} and σ_z^2, inversely on σ_u^2, and in a complicated way on the cross-correlation structure of x_t and z_t.

There are two other forms of misspecification that are of frequent concern in practice. The first is usually referred to as simultaneous equation bias, which occurs when one or more of the elements of X_t is endogenous. Explicit representation of the noncentrality parameter in this case is extremely complex and requires assumptions about the complete structural model. However, it can be shown that the local power depends on the covariance structure among the disturbances in the system, the relation among the elements of X_t that are endogenous, and the relevant excluded exogenous variables. Hence, the simultaneous equation problem can be thought of as a combination of the measurement error problem and the omitted variable problem.

A second type of misspecification occurs when an incorrect functional form is used. For example, one may estimate the linear model, $y_t = \beta_0 x_t + \varepsilon_t$, when, in fact, the correct functional form is semilog, $y_t = \gamma_0 \ln x_t + u_t$. The linear specification is a misspecified regression with an omitted variable, $\ln x_t$. Consequently, many types of functional form misspecification can be considered as a special case of the omitted variable problem.

In this paper we focus on the differencing transformation or filter as a means of investigating model misspecification. The power of the test arises because the probability limit of the least squares estimators from the differenced and undifferenced regressions generally differ. It is useful to recognize, however, that there is nothing unique in the differencing filter in this context. Indeed, one could choose any linear filter for the transformation, such as quasi-differencing $(1-\lambda B)$, or higher order polynomials like $(1-\lambda_1 B - \lambda_2 B^2)$, and construct a similar test statistic. Moreover, if one has a specific alternative hypothesis in mind, it would be possible to design the "optimal" filter which would have the greatest power in that situation.[8] For example, consider filters of the class $(1-\lambda B)$ in the measurement error model of equations (6) and (7). The least squares estimator of the quasi-differenced model is $\tilde{\beta}_T(\lambda) = (\dot{x}'\dot{x})^{-1}\dot{x}'\dot{y}$, where $\dot{x}_t = x_t - \lambda x_{t-1}$, and $\dot{y}_t = y_t - \lambda y_{t-1}$. The noncentrality parameter can be shown to be

$$\nu = 2\lambda^2\beta_0^2\sigma_v^{*4}\rho_z^2/(\sigma_u^2\sigma_z^2[.5 + \lambda^2 - \lambda^4/2 + (\lambda^3-\lambda)\rho_z - \lambda^2\rho_z^2]),$$

[8] We are grateful to Clive Granger for suggesting this point.

which is maximized for $\lambda = 1$ for all $-1 < \rho_z < 1$. Hence, the differencing filter is the optimal filter of this class for local alternatives for this errors-in-variables model.

More generally, we have chosen to focus on the differencing filter for two reasons: (1) changes of variables are convenient and easily implemented in practice; and (2) in most applications the researcher has only a vague idea of the type of model misspecification, so that the design of an "optimal" filter is out of the question.

4. SIMULATION EXPERIMENTS

Although the results in Sections 2 and 3 characterize the size and asymptotic local power of the differencing test for large samples, further insight into the operating characteristics of the test in situations likely to be encountered in practice can be gained by means of Monte Carlo simulation. To provide some perspective for these experiments, the instrumental variables specification test suggested by Wu [1973], the RESET specification test suggested by Ramsey [1969] and Ramsey and Schmidt [1976], and the Durbin-Watson test are also examined.

The Wu test compares the least squares estimators of the presumably correctly specified time series regression in (1), $\hat{\beta}_T$, with the instrumental variable estimator

(12) $$\beta_T^{IV} = (Z'X)^{-1}Z'y$$

where the instrumental variables are lagged values of the regressors, $Z_t = X_{t-1}$. Wu's test statistic can be expressed as

(13) $$\text{Wu} \equiv T\cdot(\beta_T^{IV} - \hat{\beta}_T)'\hat{V}^{-1}(\beta_T^{IV} - \hat{\beta}_T) \overset{A}{\sim} \chi_k^2,$$

where

$$\hat{V}^{-1} = \hat{\sigma}_T^2[(Z'X/T)^{-1}(Z'Z/T)(X'Z/T)^{-1} - (X'X/T)^{-1}]$$

and $\hat{\sigma}_T^2 = (y - X\hat{\beta}_T)'(y - X\hat{\beta}_T)/T$, provided that $\hat{V}$ is nonsingular. Thus, the Wu test and the differencing test are closely related.

The RESET test is based on the assumption that under the null hypothesis the distribution of ε, conditional on X, is normal $(0, \sigma^2 I)$ whereas under the alternative the distribution is normal (μ, Ω). The variant of the test used here is discussed by Ramsey and Schmidt [1976]. The idea is that if μ can be approximated by a linear combination of the columns of some observable matrix D, then a test of specification can be based on the regression of y on X and D using the usual F-test to test the hypothesis that the coefficients associated with D are zero. Ramsey [1969] assumes that μ can be expressed as a function of $\chi\beta_0$ and suggests that powers of $\hat{y} = X\hat{\beta}_T$ are logical candidates for the columns of D. Following this suggestion we choose D as a $T\times 3$ matrix whose columns are $\hat{y}^2$, $\hat{y}^3$, and $\hat{y}^4$ respectively. The RESET test statistic is then distributed as $F(3, T-k-3)$.

The Durbin-Watson test statistic is sometimes used as a specification test. For example, the presence of significant serial correlation in the residuals can be viewed as an indication of an omitted variable. The test is based on the statistic

$$d = \frac{\sum_{t=2}^{T} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{T} \hat{\varepsilon}_t^{\,2}} \tag{14}$$

where $\hat{\varepsilon}_t$ is an element of $\hat{\varepsilon} = y - X\hat{\beta}_T$. Of course, it should be noted that the Durbin-Watson test is not designed to detect specification errors that cause a dependence between the error term and the regressors. Nevertheless, it is commonly used as a crude test of specification.

4.1. *Small Sample Distribution of Test Statistics.* The RESET test is an exact F-test and the finite sample (lower bound) critical value of the Durbin-Watson test can be taken from existing tables. The differencing test and the Wu test are only asymptotically distributed as χ^2. To investigate the finite sample properties of these two large sample test statistics we simulate the differencing and Wu tests under the null hypothesis that there is no specification error in the simple regression model $y_t = \beta_0 x_t + u_t$. In all of these simulations $\beta_0 = 1$, $\sigma_x^2 = \sigma_u^2$, so that the coefficient of determination is 0.50 and x_t is generated as an AR(1). The sample size is fixed at 30 observations and the autocorrelation coefficient for x_t, ρ_x, is varied between 0 and 1. Each simulation is replicated 1000 times.[9] Table 1 reports the summary statistics for these simulations.

Several things are noteworthy in Table 1. First, the frequency distributions of the differencing and Wu tests are quite close to the large sample χ_1^2 distribution, even for this relatively small sample size ($T=30$). In particular, the estimates of the actual size of a 5 percent level test are quite close to .05. This seems to hold approximately even in the case of $\rho_x = 1$ (i.e., the regressor follows a random walk) which is not covered by Theorems 1 and 2. Since each experiment is an independent drawing from a binomial distribution with a probability p of rejecting the null hypothesis, the sample proportion of rejections is unbiased, $E(\hat{p}) = p$, and its variance is $p(1-p)/1000$. For $p = .05$, the large sample standard error for $\hat{p}$ is .007, so almost all of the rejection frequencies are within two standard errors of .05. Second, the Wu test seems to have a χ_1^2 distribution when $\rho_x = 0$, even though the instrumental variable estimator β_T^{IV} is inconsistent when $\rho_x = 0$. Based on the results in Table 1, it seems reasonable to use the large sample χ^2 distribution of the differencing and Wu tests for samples as small as 30 observations.

4.2. *Errors-in-Variables.* To estimate the power of the differencing test against the alternative hypothesis of errors-in-variables, consider the model of equations (6) and (7). The noncentrality parameter for the Wu test can be

[9] The experiments were performed on the Hewlett-Packard 3000 computer at the University of Rochester. All of the random variables were generated to have normal distributions using the Box-Müller [1958] transformation of the uniform pseudo-random deviates which were generated by HP3000. Tests on the independence and the normality of the generated data indicated no problems.

TABLE 1

SMALL SAMPLE (T = 30) DISTRIBUTIONS OF DIFFERENCING AND WU TESTS UNDER THE NULL HYPOTHESIS

	Differencing Test, J				Wu Test			
	Mean	Variance	Percent .95 fractile of χ_1^2	Goodness-of-Fit Test	Mean	Variance	Percent .95 fractile of χ_1^2	Goodness-of-Fit Test
Theoretical χ_1^2 distribution	1.0	2.0	.050	—	1.0	2.0	.050	—
Simulation results Autocorrelation of x_t, ρ_x								
0	1.078	2.100	.059	11.6	.979	1.753	.045	7.3
.3	1.080	2.211	.059	14.9	1.049	2.211	.050	7.5
.5	1.031	1.977	.055	10.3	1.060	1.999	.055	10.4
.7	1.022	1.943	.059	4.8	1.112	2.341	.064	16.7
.9	1.027	1.915	.046	11.3	.986	2.028	.045	5.9
1.0	1.165	2.353	.072**	23.3*	1.076	1.960	.057	19.3*

NOTE: The Goodness-of-Fit test is based on the deciles of the theoretical χ_1^2 distribution, so the Goodness-of-Fit test should have a χ_9^2 distribution. Each simulation is replicated 1000 times for a sample size of 30. The model is $y_t = \beta_0 x_t + u_t$, where $\beta_0 = 1$ and $\sigma_x^2 = \sigma_u^2$.

* Greater than the .95 fractile of the χ_9^2 distribution.

** More than two standard errors from .050.

shown to be half that for the differencing test, i.e., $\nu = \beta_0^2 \rho_z^2 \sigma_v^{*4}/(\sigma_u^2 \sigma_z^2 (1-\rho_z^2))$. Thus, we should generally expect the differencing test to outperform the Wu test.

The RESET test should also indicate misspecification in this case since errors-in-variables can be viewed like the misspecification due to simultaneous equation complications discussed by Ramsey [1969]. The Durbin-Watson test also has the potential to detect misspecification due to measurement error since, as discussed in Grether and Maddala [1973] and Plosser [1981], measurement error typically produces autocorrelation in the estimated residuals even when the true disturbances are serially uncorrelated.

The Monte Carlo experiments based on the measurement error model set $\beta_0 = 1$, $\rho_v = 0$, and $\sigma_z^2 = 1/(1-\rho_z^2)$.[10] The variance of the measurement error, σ_v^2, is set in two different ways: (a) the measurement error variance, σ_v^2, is set equal to 10 or 50 percent of the variance of the true regressor z, which determines the bias in the estimator of β_0 from the undifferenced regression, and (b) the measurement error variance is set equal to 10 or 50 percent of the variance of the true disturbance u, which determines the precision with which β_0 is estimated. The noncentrality parameters for the differencing test and the Wu test indicate

[10] The true unobserved regressor z_t is created to follow a first order autoregressive process, $z_t = \rho_z z_{t-1} + a_t$, where a_t is normal and independent.

TABLE 2
ESTIMATED POWER FOR THE ERRORS-IN-VARIABLES MODEL
5 PERCENT LEVEL TESTS

			$T=30$				$T=100$				$T=200$			
ρ_z	σ_v^2/σ_z^2	σ_v^2/σ_u^2	Δ	*Wu*	*d*	*RESET*	Δ	*Wu*	*d*	*RESET*	Δ	*Wu*	*d*	*RESET*
	.1	.1	.060	.048	.072	.053	.068	.063	.079	.054	.085	.056	.070	.045
.3	.5	.1	.058	.061	.064	.048	.064	.072	.084	.061	.102	.064	.046	.056
	.1	.5	.081	.057	.074	.044	.112	.082	.082	.052	.164	.084	.052	.057
	.5	.5	.086	.070	.059	.042	.166	.101	.099	.027	.287	.147	.074	.055
	.1	.1	.076	.046	.069	.046	.115	.091	.079	.042	.165	.117	.050	.048
.5	.5	.1	.084	.074	.070	.042	.132	.093	.072	.052	.224	.130	.065	.053
	.1	.5	.099	.077	.069	.046	.217	.133	.063	.047	.433	.230	.044	.045
	.5	.5	.137	.095	.054	.052	.367	.217	.090	.064	.675	.395	.099	.054
	.1	.1	.098	.076	.066	.053	.176	.126	.094	.055	.282	.190	.057	.056
.7	.5	.1	.112	.068	.065	.065	.197	.142	.089	.051	.408	.253	.056	.060
	.1	.5	.169	.103	.052	.055	.497	.288	.088	.045	.816	.534	.056	.055
	.5	.5	.225	.126	.081	.042	.672	.415	.108	.054	.937	.734	.114	.048
	.1	.1	.124	.093	.077	.049	.374	.243	.065	.056	.644	.126	.047	.070
.9	.5	.1	.107	.082	.074	.049	.312	.206	.072	.063	.609	.424	.053	.068
	.1	.5	.318	.187	.061	.051	.879	.659	.079	.088	.996	.948	.059	.080
	.5	.5	.271	.163	.090	.047	.870	.672	.171	.065	.995	.958	.200	.095

NOTE: The true model is: $y_t=\beta_0 z_t+u_t$, where $\beta_0=1$, $z_t=\rho_z z_{t-1}+e_t$ and $\sigma_e^2=1$. There is independent measurement error in z_t such that the observed value is $x_t=z_t+v_t$. The relative amount of measurement error in x_t is determined by σ_v^2/σ_z^2. The model which is estimated is: $y_t=\beta_0 x_t+\varepsilon_t$, where $\varepsilon_t=u_t-\beta_0 v_t$, hence the amount of measurement error relative to the true regression disturbance is σ_v^2/σ_u^2. Δ is the test statistic based on the differences regression, and *Wu* is the test statistic based on the instrumental variables regression using x_{t-1} as the instrument. The Durbin-Watson statistic, *d*, tests for residual autocorrelation. The rejection frequencies are based on the lower bound value of the tabled distribution. *RESET* is the specification test due to Ramsey and is based on adding $\hat{y}^2$, $\hat{y}^3$, and $\hat{y}^4$ to the estimated model. *T* is the number of observations used to estimate the undifferenced regression. The power estimates are based on 1000 replications.

that the local power of both tests should increase with σ_v^2/σ_z^2 and σ_v^2/σ_u^2.

Table 2 contains estimates of power for the differencing test, the Wu test, the RESET test, and the Durbin-Watson test for different values of ρ_z, σ_v^2/σ_z^2, σ_v^2/σ_u^2, and the sample size *T*. The differencing test and the Wu test statistics are assumed to be distributed as χ_1^2. The RESET test statistic is distributed as $F(3, T-5)$. For the Durbin-Watson test statistic, *d*, the lower bound of the statistic is used to compute the rejection frequency.[11]

[11] The lower bound is a conservative procedure but is only used for $T=30$ and $T=100$. For $T=200$, we use a normal approximation for *d* and the standard deviation of the sampling distribution based on 500 replications under the null hypothesis.

In virtually all cases, the differencing test rejects more frequently than any of the other tests, and often the difference is quite substantial. The estimates of power appear to increase with the autocorrelation of z, with the variance of the measurement error, and with the sample size. It is interesting that for the range of parameter values considered, neither the Durbin-Watson nor RESET tests appear to have much power. The largest estimated rejection frequency for the Durbin-Watson test is .20 and for the RESET test is .095. Based on these experiments, it seems that the differencing test dominates the other tests for nonlocal alternatives when the alternative hypothesis is measurement error in the regressor.

4.3. *Omitted Variables* To examine the power of the differencing test against the hypothesis of an omitted variable, consider the model of equations (9) and (10). For this case, the noncentrality parameter for the Wu test can be shown to be

$$\nu = \gamma_0^{*2}\sigma_z^2(\rho_{xz}\rho_x - \rho_{xz}(+1))^2/\sigma_u^2(1-\rho_x^2). \tag{15}$$

The relation between this expression and that for the differencing test is complicated and there is no simple expression that indicates when one test is expected to outperform the other.

The simulation experiments use equations (9) and (10) and set $\beta_0=\gamma_0=1$, $\sigma_x^2=\sigma_z^2$, and the coefficient of determination for the true model is equal to .50. Hence, the parameters that vary across experiments are the first-order autocorrelation coefficient for x, ρ_x, the cross-correlation coefficients between x_t and z_{t+k}, $\rho_{xz}(k)$, for $k=-1, 0, +1$, and the sample size, T.[12] Table 3 presents the proportion of rejections of the various tests using a 5 percent level of significance based on 1000 replications of each experiment.

The estimates of power for the differencing and Wu tests vary substantially for different values of ρ_x, and $\rho_{xz}(k)$, $k=+1, 0, -1$ in Table 3. Neither the differencing test nor the Wu test seems to be very powerful for the small samples, such as $T=30$, but both tests have very high proportions of rejections for some alternative hypotheses when the sample size is moderate ($T=100$) to large ($T=200$).

As suggested by the noncentrality parameter, the differencing test is most powerful when $[\rho_x\rho_{xz}(0)-(\rho_{xz}(-1)+\rho_{xz}(+1))/2]$ is furthest from zero. Several things should be noticed about this fact. First, the cross-correlation coefficients between x and z at lead and lag 1 enter symmetrically in this test; that is, it is only the sum, and not the pattern of these coefficients that matters. This explains why the results for the differencing test are essentially the same for experiments II and III, and also for experiments IV, V, and VI. Second, the differencing test has

[12] The regressor sequences are constructed as follows. First, the included regressor x_t is created to follow a first order-autoregressive process, $x_t=\rho_x x_{t-1}+a_t$, where a_t is normal and independent. Next, the excluded regressor z_t is created by the following model, $z_t=w_1x_{t-1}+w_2x_t+w_3x_{t+1}+b_t$, where b_t is normal and independent. The coefficients w_1, w_2, and w_3 are functions of the cross-correlation coefficients $\rho_{xz}(k)$, the autocorrelation of x_t, ρ_x, and the relative standard deviations of x_t and z_t.

TABLE 3

ESTIMATED POWER FOR THE OMITTED VARIABLE MODEL 5 PERCENT LEVEL TESTS

Experiment	$\rho_{xz}(+1)$	$\rho_{xz}(0)$	$\rho_{xz}(-1)$	ρ_x	T=30				T=100				T=200			
					Δ	*Wu*	*d*	*RESET*	Δ	*Wu*	*d*	*RESET*	Δ	*Wu*	*d*	*RESET*
I	0	.5	0	0	.054*	.052†	.069	.034	.053*	.059†	.066	.054	.054*	.052†	.065	.039
				.3	.116	.067	.051	.041	.202	.119	.079	.055	.361	.190	.057	.045
				.5	.188	.114	.055	.053	.537	.292	.073	.060	.853	.545	.054	.043
				.7	.472	.241	.063	.057	.963	.703	.079	.063	.999	.948	.091	.056
II	.25	.5	0	0	.063	.079	.074	.039	.140	.153	.073	.058	.261	.254	.061	.046
				.3	.056	.069	.071	.047	.043	.103	.071	.045	.063	.107	.064	.051
				.5	.091	.051†	.069	.039	.189	.048†	.059	.057	.303	.052†	.054	.034
				.7	.230	.061	.050	.050	.621	.102	.070	.048	.913	.161	.049	.052
III	0	.5	.25	0	.075	.057†	.068	.034	.135	.042†	.077	.047	.266	.054†	.058	.041
				.3	.060	.070	.062	.030	.052	.100	.072	.043	.063	.183	.070	.061
				.5	.090	.119	.060	.049	.173	.282	.061	.041	.317	.529	.045	.042
				.7	.246	.258	.058	.048	.673	.736	.066	.045	.913	.950	.058	.049
IV	.25	.5	.25	0	.142	.067	.060	.042	.425	.127	.082	.039	.718	.249	.048	.042
				.3	.071	.054	.061	.042	.122	.091	.080	.048	.195	.101	.048	.053
				.5	.052*	.041†	.073	.052	.059*	.057†	.088	.052	.060*	.058†	.052	.043
				.7	.078	.079	.072	.042	.175	.118	.082	.053	.309	.181	.050	.039
				.9	.435	.235	.059	.065	.933	.656	.105	.081	1.00	.932	.105	.059
V	.5	.5	0	0	.141	.158	.060	.041	.430	.418	.071	.054	.746	.642	.053	.055
				.3	.062	.127	.063	.034	.103	.421	.074	.047	.183	.762	.046	.054
				.5	.040*	.120	.059	.044	.042*	.332	.083	.049	.032*	.551	.052	.051
				.7	.077	.093	.077	.038	.180	.176	.096	.050	.308	.333	.064	.039

	0	.5	.5	0	.141	.052†	.051	.042	.417	.047†	.069	.051	.745	.050†	.042	.050
VI				.3	.066	.062	.067	.043	.122	.101	.077	.050	.185	.204	.058	.043
				.5	.040*	.118	.066	.046	.046*	.325	.085	.036	.056*	.540	.052	.042
				.7	.068	.266	.061	.037	.153	.688	.072	.034	.283	.957	.061	.050

NOTE: The true model is: $y_t = \beta x_t + \gamma z_t + u_t$, where $\beta = \gamma = 1$, $\sigma_x^2 = \sigma_z^2$, and $\sigma_u^2 = \sigma_x^2 + \sigma_z^2 + 2\sigma_x\sigma_z\rho_{xz}(0)$, so the R^2 for the true model is .50. ρ_x is the first-order autocorrelation coefficient for x_t, corr (x_t, x_{t-1}), and $\rho_{xz}(k) =$ corr (x_t, z_{t+k}) is the cross-correlation coefficient between x_t and z_{t+k} for $k = 1, 0, -1$. The model which is estimated is: $y_t = \beta x_t + \varepsilon_t$, so z_t is the omitted variable. Δ is the test statistic based on the differences regression, and *Wu* is the test statistic based on the instrumental variables regression using x_{t-1} as the instrument. The Durbin-Watson statistic, *d*, tests for residual autocorrelation. The rejection frequencies are based on the lower bound value of the tabulated distribution. *RESET* is the specification test due to Ramsey and is based on adding $\hat{y}^2$, $\hat{y}^3$, and $\hat{y}^4$ to the estimated model. *T* is the number of observations used to estimate the undifferenced regression. The power estimates are based on 1000 replications.

* This is a case where plim $(\tilde{\beta}_T - \hat{\beta}_T) = 0$, so the differencing test should not have power.

† This is a case where plim $(\beta_T^{IV} - \hat{\beta}_T) = 0$, so the *Wu* instrumental variables test should not have power.

power equal to the size of the test in experiments IV, V, and VI when $\rho_x = .5$, because plim $(\tilde{\beta}_T - \hat{\beta}_T) = 0$. Thus, it is necessary to consider the autocorrelation of the included variable in comparison with the cross-correlations of the included variable with the excluded variable to determine whether the differencing test has power against a particular omitted variable.

As suggested by equation (15), the Wu test is most powerful when $[\rho_{xz}(+1) - \rho_x \rho_{xz}(0)]$ is furthest from zero. In this case, it is only the correlation between z_t and the instrumental variable x_{t-1} that determines the power; the cross-correlation between z_t and x_{t+1} does not matter. This explains why the results for the Wu test are essentially the same for experiments, I, III, and VI, and also for experiments II and IV in Table 3. The Wu test has power equal to the size of the test when the cross-correlation function between x and z decays at the same rate as the autocorrelation function for x; that is, when $\rho_{xz}(+1) = \rho_x \rho_{xz}(0)$. This occurs for one set of parameters in each of the experiments, except for experiment V.

The comparison of the power estimates for the differencing and Wu tests in Table 3 shows how different these tests are from one another. In virtually all instances at least one of these tests rejects significantly more than 5 percent of the time. Thus, in typical situations where the alternative hypothesis is not well-specified in terms of the cross-correlations between the included variable and the excluded variable, it may be worthwhile to compute both the differencing and the Wu tests. Since the tests are not independent, the size of the joint test is difficult to determine precisely. Nevertheless, Bonferroni bounds are easy to apply. To obtain a joint test of size less than or equal to α, set the nominal size of each test individually at $\alpha/2$.

Table 3 also indicates that for the range of parameters considered neither the RESET test nor the Durbin-Watson test are very useful for detecting the omitted variable. However, it is useful to recognize that the power of all the tests reported in Table 3 can be increased by reducing σ_u^2 and thus increasing the R^2 for the true model. In some limited experimentation we have found that by raising the R^2 of the true model to above .90, we get rejection frequencies for the RESET test that are comparable to those reported in Ramsey and Gilbert [1972]. Of course the estimated power of the differencing test and the Wu test also rises substantially and remains significantly larger than that of the RESET and Durbin-Watson tests.

5. CONCLUSION

This paper introduces and analyzes a simple specification test for linear time series regression models. This test formalizes the procedure advocated by Plosser and Schwert [1978] of comparing the least squares estimators from the undifferenced and differenced forms of the model.

We show that specification errors that lead to regressors that are correlated with the disturbances also lead to least squares estimators from the undifferenced and differenced forms of the model that have different probability limits. This suggests that the differencing test will have power to detect such specification errors. We examine in detail the asymptotic local power of the differencing test under the alternative hypotheses of measurement error and omitted variables.

The small sample properties of the differencing test are investigated through some Monte Carlo simulation experiments. Based on these experiments it seems that the asymptotic properties of the differencing test are reasonably valid for samples as small as 30 observations. Also, for the range of models investigated, the differencing test performs well in comparison with an instrumental variables specification test proposed by Wu [1973], the frequently employed Durbin-Watson test and the RESET test proposed by Ramsey [1969] and Ramsey and Schmidt [1976].

The differencing test developed in this paper is general in the sense that it is designed to test against a broad class of alternative hypotheses. This is appropriate and desirable if the potential sources of specification error are unknown. Thus, the differencing test can be viewed as a "diagnostic check" on the model specification. This interpretation raises the question of what should be done if the differencing test leads to a rejection of the maintained model specification. One obvious strategy would be to perform some additional tests that are designed to test against specific alternative hypotheses; for example, additional variables could be added to the model or alternative functional forms could be estimated.[13] However, even if there are no obvious candidates for alternative model specifications, the rejection of the specification using the differencing test should be a clear warning not to proceed with inference or prediction as though the maintained model specification was adequate.

Thus, it seems that the differencing test could have many fruitful applications in situations where there is reason to think that the regression disturbance is not orthogonal to the regressors. The test is both easy to apply and is likely to be robust against a wide range of misspecifications commonly found in practical situations.

University of Rochester,
University of Rochester,
University of California, San Diego.

MATHEMATICAL APPENDIX

All symbols and definitions are as given in the text. For convenience, we restate the first assumption.

A1. The model can be written in mean-deviation form as

[13] Of course, the usual pre-testing problems arise when a sequence of related tests are used to arrive at an "adequate" model specification.

$$y_t = X_t\beta_0 + \varepsilon_t, \qquad t = 1,\ldots, T,$$

where $\{X_t\}$ is a weakly stationary sequence of random $1 \times k$ vectors such that plim $T^{-1} \sum_{t=1}^{T} X_t'X_t = M_{XX} \equiv E(X_t'X_t)$ is finite and nonsingular; $\{\varepsilon_t\}$ is a pure white noise sequence, such that $E(\varepsilon_t)=0$, $E(\varepsilon_t^2)=\sigma_o^2$, $0<\sigma_o^2<\infty$; ε_t is independent of X_τ for all τ; and β_0 is an unknown finite $k\times 1$ vector of constants.

To guarantee that $\sqrt{T}(\hat\beta_T-\beta_0)\overset{A}{\sim} N(0, \sigma_o^2 M_{XX}^{-1})$, A1 and the following assumption are sufficient.

A2. There exist $\delta>0$ and $\eta<\infty$ such that $E|X_{it}X_{jt}|^{1+\delta}<\eta$ and $E|\varepsilon_t^2|^{1+\delta}<\eta$ for $i, j=1,\ldots, k$ and all t, and

$$\lim_{t\to\infty} E|E(X_{i\tau+t}X_{j\tau+t}\|X_1,\ldots, X_\tau) - m_{ij} = 0,$$

uniformly in τ, where $m_{ij}\equiv E(X_{i\tau}X_{j\tau})$ for $i, j=1,\ldots, k$.

Next, we introduce the $T\times k$ instrument matrix Z with rows Z_t, and impose the following condition.

A3. $\{Z_t\}$ is a weakly stationary sequence of random $1\times k$ vectors such that plim $T^{-1} \sum_{t=1}^{T} \dot Z_t'\dot Z_t = M_{\dot Z\dot Z} \equiv E(\dot Z_t'\dot Z_t)$ is finite and nonsingular; the sequence $\{Z_t, \dot X_t\}$ is jointly weakly stationary such that plim $T^{-1} \sum_{t=1}^{T} Z_t'\dot X_t = M_{Z\dot X}\equiv E(Z_t'\dot X_t)$ is finite and nonsingular, and Z_t is independent of ε_τ for all τ.

A4. There exist $\delta>0$ and $\eta<\infty$ such that $E|\dot Z_{it}\dot Z_{jt}|^{1+\delta}<\eta$ for $i, j=1,\ldots, k$ and all t, and

$$\lim_{t\to\infty} E|E(\dot Z_{i\tau+t}\dot Z_{j\tau+t}\|\dot Z_1,\ldots, \dot Z_\tau)-l_{ij}| = 0,$$

uniformly in τ, where $l_{ij}\equiv E(\dot Z_{it}\dot Z_{jt})$, for $i, j=1,\ldots, k$.

A3 and A4 are analogous to A1 and A2. Note, however, that A4 is in terms of the differences of the instruments, $\dot Z_t$. Conditions A1, A3 and A4 are sufficient to ensure the asymptotic normality of $\tilde\beta_T$. Specifically, $\sqrt{T}(\tilde\beta_T-\beta_0)\overset{A}{\sim} N(0, \sigma_0^2 M_{Z\dot X}^{-1} M_{\dot Z\dot Z} M_{Z\dot X}^{-1\prime})$.

To ensure the asymptotic normality of $\sqrt{T}(\tilde\beta_T-\hat\beta_T)$, which requires the joint asymptotic normality of $\sqrt{T}(\hat\beta_T-\beta_0)$ and $\sqrt{T}(\tilde\beta_T-\beta_0)$, it is sufficient to assume in addition to A1–A4:

A5. There exist $\delta>0$ and $\eta<\infty$ such that $E|\dot Z_{it+1}X_{jt}|^{1+\delta}<\eta$ for $i, j=1,\ldots, k$ and all t, and

$$\lim_{t\to\infty} E|E(\dot Z_{it+\tau+1}X_{jt+\tau}\|X_1,\ldots, X_\tau, \dot Z_1,\ldots, \dot Z_{\tau+1})-K_{ij}|=0,$$

uniformly in τ, where $K_{ij}=E(\dot Z_{it+1}X_{jt})=-E(Z_{it}\dot X_{jt})$ for $i, j=1,\ldots, k$.

With this structure, we can derive the properties of the differencing statistic Δ in a straightforward manner. Since Theorem 1 is actually a simple corollary of

Theorem 2, we prove Theorem 2 first.

THEOREM 2. *Given A1–A5,* $\Delta \stackrel{A}{\sim} \chi^2_k$ *where*

$$\hat{V} = \hat{\sigma}^2_T \left((Z'\dot{X}/T)^{-1}(\dot{Z}'\dot{Z}/T)(\dot{X}'Z/T)^{-1} - (X'X/T)^{-1}\right)$$

and

$$\hat{\sigma}_T = (y - X\hat{\beta}_T)'(y - X\hat{\beta}_T)/T,$$

provided plim $\hat{V}$ *is nonsingular.*

SKETCH OF PROOF: The desired result is obtained by deriving the asymptotic distribution of

$$\text{(a.1)} \quad \sqrt{T}(\tilde{\beta}_T - \hat{\beta}_T) = \sqrt{T}(Z'\dot{X})^{-1}Z'\dot{\varepsilon} - \sqrt{T}(X'X)^{-1}X'\varepsilon$$
$$= (Z'\dot{X}/T)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z'_t\dot{\varepsilon}_t - (X'X/T)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X'_t\varepsilon_t .$$

Since we may write

$$\text{(a.2)} \qquad \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z'_t\dot{\varepsilon}_t = \frac{-1}{\sqrt{T}}\sum_{t=1}^{T} \dot{Z}'_{t+1}\varepsilon_t + \frac{1}{\sqrt{T}}(Z'_{T+1}\varepsilon_T - Z'_1\varepsilon_0),$$

consider the asymptotic distribution of

$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \lambda'(M^{-1}_{Z\dot{X}}\dot{Z}'_{t+1} + M^{-1}_{XX}X'_t)\varepsilon_t ,$$

where λ is any real $k \times 1$ vector. The weakly stationary random variables $\lambda'(M^{-1}_{Z\dot{X}}\dot{Z}'_{t+1} + M^{-1}_{XX}X'_t)\varepsilon_t$ are easily shown to be elements of a sequence of martingale differences.

Asymptotic normality then follows upon application of the appropriate martingale central limit theorem. For the present case, corollary 6.1.1 of Serfling [1968] is applicable given A1–A5. Thus,

$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \lambda'(M^{-1}_{Z\dot{X}}\dot{Z}'_{t+1} + M^{-1}_{XX}X'_t)\varepsilon_t$$
$$\stackrel{A}{\sim} N(0, \sigma^2_0\lambda'(M^{-1}_{Z\dot{X}}M_{\dot{Z}\dot{Z}}M^{-1}_{Z\dot{X}} - M^{-1}_{XX})\lambda).$$

That $\sqrt{T}\lambda'(\tilde{\beta}_T - \hat{\beta}_T)$ has the identical asymptotic distribution follows from Rao [1973] 2c.4 $(x.d)$ provided that

$$\left| \sqrt{T}\lambda'(\tilde{\beta}_T - \hat{\beta}_T) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \lambda'(M^{-1}_{Z\dot{X}}Z'_{t+1} + M^{-1}_{XX}X'_t)\varepsilon_t \right| \xrightarrow{p} 0.$$

This is straightforward to verify given A1–A5, so that for any real λ,

$$\sqrt{T}\lambda'(\tilde{\beta}_T - \hat{\beta}_T) \stackrel{A}{\sim} N(0, \sigma^2_0\lambda'(M^{-1}_{Z\dot{X}}M_{\dot{Z}\dot{Z}}M^{-1\prime}_{Z\dot{X}} - M^{-1}_{XX})\lambda).$$

It follows from Rao [1973] 2c. 4(xi) that

$$\sqrt{T}(\tilde{\beta}_T - \hat{\beta}_T) \stackrel{A}{\sim} N(0, \sigma^2_0(M^{-1}_{Z\dot{X}}M_{\dot{Z}\dot{Z}}M^{-1\prime}_{Z\dot{X}} - M^{-1}_{XX})).$$

It then follows from White [1980] lemma 3.3 that

$$\Delta \equiv T(\tilde{\beta}_T - \hat{\beta}_T)' \hat{V}^{-1}(\tilde{\beta}_T - \hat{\beta}_T) \overset{A}{\sim} \chi_k^2,$$

provided that

$$\hat{V} = \hat{\sigma}_T((Z'\dot{X}/T)^{-1}(\dot{Z}'\dot{Z}/T)(\dot{X}'Z/T)^{-1} - (X'X/T)^{-1})$$

is consistent for $V = \sigma_0^2(M_{Z\dot{X}}^{-1} M_{\dot{Z}\dot{Z}} M_{Z\dot{X}}^{-1\prime} - M_{XX}^{-1})$, where V is nonsingular. This follows easily given A1 and A3 and the desired result follows. Q. E. D.

To prove Theorem 1, we simply replace Z_t by $\dot{X}_t$ in the preceding result. In this situation, additional restrictions must be placed on the behavior of X_t. The most compact way to impose these restrictions is to assume

A6. A3–A5 are satisfied for $Z_t = \dot{X}_t$.

Obviously, for A3 to be satisfied, X_t cannot contain lagged dependent variables.

REFERENCES

Box, G. E. P. and M. E. Müller, "A Note on the Generation of Random Normal Deviates," *Annals of Mathematical Statistics*, 29 (1958), 610–611.

Durbin, J. and G. S. Watson, "Testing for Serial Correlation in Least-squares Regression, I," *Biometrika*, 37 (December, 1950), 409–428.

———, "Testing for Serial Correlation in Least-squares Regression, II," *Biometrika*, 38 (June, 1951), 159–178.

Goldberger, A. S., *Econometric Theory*, (New York: John Wiley and Sons, 1964).

Hausman, J. A., "Specification Tests in Econometrics," *Econometrica*, 46 (1978), 1251–1272.

Pierce, D. A., "Least Squares Estimation in Dynamic-Disturbance Time Series Models," *Biometrika*, 59 (April, 1972), 73–78.

Plosser, C. I., "Serial Correlation Corrections and Distributed Lag Models in the Presence of Measurement Error," Working Paper Series, No. 7830, Graduate School of Management, University of Rochester (1981).

——— and G. W. Schwert, "Estimation of a Non-invertible Moving Average Process: The Case of Overdifferencing," *Journal of Econometrics*, 6 (September, 1977), 199–224.

——— "Money, Income, and Sunspots: Measuring Economic Relationships and the Effects of Differencing," *Journal of Monetary Economics*, 4 (November, 1978), 637–660.

Ramsey, J. B., "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis", *Journal of the Royal Statistical Society, Ser. B*, 31 (1969), 350–371.

——— and R. Gilbert, "A Monte Carlo Study of Some Small Sample Properties of Tests for Specification Error," *Journal of the American Statistical Association*, 67 (March, 1972), 180–186.

——— and P. Schmidt, "Some Further Results on the Use of OLS and BLUS Residuals in Specification Error Tests," *Journal of the American Statistical Association*, 71 (June, 1976), 389–390.

Rao, C. R., *Linear Statistical Inference and Its Applications*, (New York: John Wiley and Sons, 1973).

Serfling, R. J., "Contributions to Central Limit Theory for Dependent Variables," *Annals of Mathematical Statistics*, 39 (August, 1968), 1158–1175.

White, H., "Nonlinear Regression on Cross-Section Data," *Econometrica*, 48 (April, 1980), 721–746.

Wu, D., "Alternative Tests of Independence Between Stochastic Regressors and Disturbances," *Econometrica*, 41 (July, 1973), 733–750.