# CONSISTENT SPECIFICATION TESTING VIA NONPARAMETRIC SERIES REGRESSION

By Yongmiao Hong and Halbert White[1]

This paper proposes two consistent one-sided specification tests for parametric regression models, one based on the sample covariance between the residual from the parametric model and the discrepancy between the parametric and nonparametric fitted values; the other based on the difference in sums of squared residuals between the parametric and nonparametric models. We estimate the nonparametric model by series regression. The new test statistics converge in distribution to a unit normal under correct specification and grow to infinity faster than the parametric rate $(n^{-1/2})$ under misspecification, while avoiding weighting, sample splitting, and non-nested testing procedures used elsewhere in the literature. Asymptotically, our tests can be viewed as a test of the joint hypothesis that the true parameters of a series regression model are zero, where the dependent variable is the residual from the parametric model, and the series terms are functions of the explanatory variables, chosen so as to support nonparametric estimation of a conditional expectation. We specifically consider Fourier series and regression splines, and present a Monte Carlo study of the finite sample performance of the new tests in comparison to consistent tests of Bierens (1990), Eubank and Spiegelman (1990), Jayasuriya (1990), Wooldridge (1992), and Yatchew (1992); the results show the new tests have good power, performing quite well in some situations. We suggest a joint Bonferroni procedure that combines a new test with those of Bierens and Wooldridge to capture the best features of the three approaches.

Keywords: Asymptotic normality for generalized quadratic forms, consistent testing, Fourier series, specification testing, regression splines.

## 1. INTRODUCTION

Not long after Hausman's (1978) landmark work on specification testing, Holly (1982) pointed out that Hausman's test fails to have unit power asymptotically against a range of misspecifications of potential concern in the regression context. Beginning with Bierens (1982), numerous authors have devoted attention to this problem by constructing consistent (asymptotic unit power) tests for misspecification. Particularly relevant is work of Bierens (1990), Eubank and Spiegelman (1990), Gozalo (1993), Lee (1988), Wooldridge (1992), and Yatchew (1992).

In this paper we continue this effort by proposing two new consistent one-sided specification tests for parametric regression models, one based on the sample covariance between the residual from the parametric model and the discrepancy between the parametric and nonparametric fitted values, and the other based on the difference in sums of squared residuals between the parametric and nonparametric models. Under correct specification, these two

statistics vanish faster than the parametric ($n^{1/2}$) rate, so a standard $n^{1/2}$-normalization leads to degenerate test statistics. With appropriate standardization, our test statistics converge in distribution to a unit normal under correct specification and grow to infinity faster than the parametric rate under misspecification, while avoiding weighting, sample-splitting, and non-nested testing procedures previously used to handle the degeneracy. We estimate the nonparametric model by series regressions, specifically Fourier series and regression splines. Asymptotically, our tests can be viewed as a test of the joint hypothesis that the "true parameters" of a series regression model are zero, where the dependent variable is the residual from the parametric model, and the series terms are functions of the explanatory variables, chosen so as to support nonparametric estimation of a conditional expectation. We present a Monte Carlo study comparing the finite sample performance of the new tests to tests of Bierens (1990), Eubank and Spiegelman (1990), Jayasuriya (1990), Wooldridge (1992), and Yatchew (1992); the results show that the new tests have good power, performing quite well in some situations. We suggest a joint Bonferroni procedure that combines our tests with those of Bierens and Wooldridge to capture the best features of the three approaches.

## 2. HEURISTICS

Let $\{Z_t \equiv (X_t', Y_t)' \in \mathbb{R}^{d+1}\}_{t=1}^{\infty}$ be a sequence of i.i.d. random vectors with $E|Y_t| < \infty$. Then there exists a measurable function $\theta_o$ such that $\theta_o(X_t) = E(Y_t|X_t)$ a.s. A standard procedure to approximate $\theta_o$ is to specify a parametric regression model, with typical element $f(\cdot, \alpha)$, $\alpha \in A$, where $A$ is a subset of a finite dimensional Euclidean space. We are interested in testing whether the model is correctly specified for $\theta_o$, as embodied by the null hypothesis

$$H_o: P[f(X_t, \alpha_o) = \theta_o(X_t)] = 1 \quad \text{for some } \alpha_o \in A.$$

The global alternative hypothesis is

$$H_A: P[f(X_t, \alpha) \neq \theta_o(X_t)] > 0 \quad \text{for all } \alpha \in A.$$

Our concern here is consistent specification testing for $H_o$ against $H_A$, i.e. testing procedures that will reject $H_o$ with asymptotic unit power whenever $H_o$ is false.

Let $\hat{\alpha}_n$ denote an estimator consistent for $\alpha_o$ under $H_o$; for example, $\hat{\alpha}_n$ could be the nonlinear least squares estimator, solving

$$\min_{\alpha \in A} n^{-1} \sum_{t=1}^{n} \{Y_t - f(X_t, \alpha)\}^2.$$

We denote the fitted value $\hat{f}_{nt} \equiv f(X_t, \hat{\alpha}_n)$ and the residual $\hat{\varepsilon}_{nt} \equiv Y_t - \hat{f}_{nt}$. A standard approach to specification testing is to regress $\hat{\varepsilon}_{nt}$ on certain functions of the conditioning variables $X_t$ to see if these additional regressors have any explanatory power. Under $H_o$, they should have no such power; a joint $F$ test

(asymptotically $\chi^2$) can be conducted to check this. If these additional regressors do exhibit statistically significant explanatory power, one has evidence of misspecification. However, such an $F$ test will miss alternatives orthogonal to these additional regressors and is thus not consistent against $H_A$.

We are motivated by the fact that in the presence of misspecification, the function of the conditioning variables most highly correlated with the regression error $\varepsilon_t \equiv Y_t - f(X_t, \alpha^*)$ is the specification error $\nu_t \equiv \theta_o(X_t) - f(X_t, \alpha^*)$, where $\alpha^*$ is the probability limit of $\hat{\alpha}_n$. Because $E\{\nu_t \varepsilon_t\} = E\{\theta_o(X) - f(X, \alpha^*)\}^2 = 0$ if and only if $H_o$ holds, a consistent test against $H_A$ can be based on the sample covariance

$$\hat{m}_n = n^{-1} \sum_{t=1}^{n} \hat{\nu}_{nt} \hat{\varepsilon}_{nt},$$

where $\hat{\nu}_{nt} = \hat{\theta}_n(X_t) - \hat{f}_{nt}$, and $\hat{\theta}_n$ is an appropriate nonparametric estimator of $\theta_o$. Various nonparametric estimators can be used. For example, $\hat{\theta}_n$ can be the ordinary least squares series estimator, solving

$$\min_{\theta \in \Theta_n} n^{-1} \sum_{t=1}^{n} \{Y_t - \theta(X_t)\}^2,$$

with

$$\Theta_n = \left\{ \theta \colon \mathbb{R}^d \to \mathbb{R} \,\middle|\, \theta(x) = \sum_{j=1}^{p_n} \beta_j \psi_j(x), \ \beta_j \in \mathbb{R} \text{ and } \psi_j \colon \mathbb{R}^d \to \mathbb{R} \right\},$$

where $\{\psi_j\}$ is a sequence of basis functions, and $p_n$ is the dimension of $\Theta_n$ chosen to grow at an appropriate rate with the sample size $n$. For concreteness, we focus our attention on use of Fourier series, Gallant's (1981) flexible Fourier form (FFF), Eubank and Speckman's (1990) polynomial-trigonometric series, and regression splines.

The challenge raised by considering $\hat{m}_n$ is that for the estimators $\hat{\theta}_n$ of interest to us, the usual standardization by $n^{1/2}$ is inappropriate: $n^{1/2}\hat{m}_n$ vanishes in probability under $H_o$, a type of degeneracy. To avoid this degeneracy, Wooldridge (1992), who also considers tests based on $\hat{m}_n$, requires that the nonparametric model delivering $\hat{\theta}_n$ be incapable of nesting the parametric model, thereby inducing sufficiently slow convergence for $\hat{\theta}_n$ to $\theta_o$. Heuristically, under $H_o$, $\hat{m}_n$ can be decomposed into two dominant conflicting effects: a variance effect and a bias effect. For each sample size $n$, the variance diverges as $p_n$ increases, while the bias converges to zero as $p_n$ increases. The non-nested approach of Wooldridge (1992) uses the bias to determine the limit distribution by controlling the variance so as to be negligible. This requires a slow growth of $p_n$ and excludes the possibility of nesting the parametric model in $\Theta_n$. Otherwise, the variance will dominate the bias, leading to overrejection of $H_o$.

A central contribution of the present work is to demonstrate that exploiting rather than avoiding the rapid convergence of $\hat{m}_n$ to zero under $H_o$ leads to

statistics that diverge more rapidly under $H_A$, with the consequent possibility of obtaining tests with better power. Our approach is to use the variance to determine the limit distribution by controlling the bias so as to be negligible. This is always possible by letting $p_n$ grow quickly or by nesting the parametric model in $\Theta_n$. Consequently, we can use straightforward choices for $\hat{\theta}_n$ without having to worry about their relationship to the parametric model.

Although our approach can be applied in the presence of heteroskedastic regression errors (see Theorem A.3 in the Appendix), we focus on testing $H_o$ under homoskedasticity (i.e. $E(\varepsilon_t^2|X_t) = \sigma_o^2$ a.s.) in order to keep our presentation succinct. Specifically, we prove that $M_n \overset{d}{\to} N(0,1)$ under $H_o$, where

$$(2.1) \qquad M_n = (n\hat{m}_n/\hat{\sigma}_n^2 - p_n)/(2p_n)^{1/2},$$

where $\hat{\sigma}_n^2$ is a variance estimator such as $n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_{nt}^2$. The form taken by $M_n$ can be understood heuristically by considering that $n\hat{m}_n/\hat{\sigma}_n^2$ behaves asymptotically like a $\chi_{p_n}^2$ statistic. Standardization toward normality involves subtracting the mean $p_n$ and dividing by the standard deviation $(2p_n)^{1/2}$. As $p_n \to \infty$, the standardized quantity becomes more nearly normal. (Our proofs, however, do not rely on this heuristic, as it is a bit too simplistic.)

To get some idea of the behavior of $M_n$ under $H_A$, consider the case in which $\hat{\theta}_n$ is obtained using Gallant's (1981) FFF. In this case a permissible choice is $p_n = \ln(n)$. (See Theorem 3.2 below.) Because $\hat{m}_n$ tends to a positive constant under $H_A$, $M_n$ behaves approximately like $Cn/(\ln(n))^{1/2}$, where $C$ is some constant. In contrast, Wooldridge's statistic behaves approximately like $C'n^{1/2}$, for a different constant $C'$. Thus, $M_n$ diverges under $H_A$ at a rate nearly the square of that of Wooldridge's statistic.

Our second test is closely related to those of Lee (1988) and Yatchew (1992), who consider basing specification testing on

$$\tilde{m}_n = n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_{nt}^2 - n^{-1}\sum_{t=1}^{n}\hat{\eta}_{nt}^2,$$

where $\hat{\eta}_{nt} = Y_t - \hat{\theta}_n(X_t)$ is the residual from the nonparametric estimation. Under $H_o$, $\tilde{m}_n$ converges to zero, while it will tend to a positive limit under $H_A$, as the first term will include the specification error. Like $\hat{m}_n$, $n^{1/2}\tilde{m}_n$ also vanishes in probability. Lee (1988) reweights $\hat{\eta}_{nt}^2$ to avoid this degeneracy. Such a weighting device is sensitive to heteroskedasticity, as the test may reject $H_o$ under heteroskedasticity even when $H_o$ is true. Alternatively, Yatchew (1992) proposes splitting the sample into two independent subsets, with the first subset used to estimate the parametric model and the second subset to estimate the nonparametric model. This approach is also used by Whang and Andrews (1993, Section 5) in testing semiparametric models. As pointed out by Wooldridge (1992), sample-splitting is rather costly.

Because we view degeneracy as a virtue rather than a vice, we obtain a viable test statistic by finding the proper standardization for $\tilde{m}_n$, instead of by modify-

ing $\tilde{m}_n$ so that the familiar $n^{1/2}$-normalization is appropriate. We consider the test statistic

$$(2.2) \qquad \bar{M}_n = (n\tilde{m}_n/\hat{\sigma}_n^2 - p_n)/(2p_n)^{1/2}$$

and show that under $H_o$ we have $M_n - \bar{M}_n = o_P(1)$. That is, $M_n$ and $\bar{M}_n$ are asymptotically equivalent. In addition to possible power improvements associated with faster divergence of $\bar{M}_n$ under $H_A$, we also avoid the drawbacks induced by weighting and sample-splitting. The statistic $\bar{M}_n$ is simple to compute because the sums of squared residuals are available in any standard regression package.

Another approach closely related to ours is taken by Eubank and Spiegelman (1990), who consider specification tests based on an orthogonal series regression using $\hat{\varepsilon}_{nt}$ as the dependent variable. The result is a nonparametric estimator of $E(\varepsilon_t|X_t)$; the idea is that this should be the zero function under $H_o$. Eubank and Spiegelman's statistic can be viewed as a joint $F$ test using the coefficients on all the $p_n$ included terms in the series regression. To obtain their results, Eubank and Spiegelman assume a linear model with a fixed single regressor and normally distributed errors $\varepsilon_t$. Jayasuriya (1990) generalizes Eubank and Spiegelman's results by dropping the normality assumption and permitting the linear model to be a fixed order polynomial (but still with a single fixed regressor). The tests we propose can be proven to be asymptotically equivalent to a test based on series regression with $\hat{\varepsilon}_{nt}$ as the dependent variable. Our results extend and complement those of Eubank and Spiegelman, and Jayasuriya (ES & J) by permitting nonorthogonal series, nonlinear parametric models, and random multiple regressors.

de Jong and Bierens (1991) use an approach similar to ES & J. Their test can be viewed as the heteroskedasticity-robust version of ES & J's test. Although they allow nonorthogonal series and test nonlinear parametric models, de Jong and Bierens' approach is specific to the Fourier series. Also, they show that their statistic grows at a rate of at least $(n/p_n^{3/2})$ under $H_A$, but do not deliver the exact rate $(n/p_n^{1/2})$. In contrast, our approach permits such nonparametric techniques such as kernel methods and smoothing splines (see White and Hong (1993) for use of kernel methods). We are also able to deliver the exact growth rate of $n/p^{1/2}$ for our test statistic under $H_A$. Our treatment of heteroskedastic errors (Theorem A.3, Appendix) is also different from theirs.

## 3. SPECIFICATION TESTING WITH FOURIER SERIES AND SPLINES

We work throughout with the following data generating process (DGP).

ASSUMPTION A.1: *For each $n \in \mathbb{N}$ the stochastic process $\{Z_t \equiv (X_t', Y_t)' \in \mathbb{R}^{d+1}$, $t = 1, 2, \ldots, n\}$, $d \in \mathbb{N}$, is independent and identically distributed with $E(Y_t^2) < \infty$. The distribution $\mu$ of $X_t$ has a continuous and positive density function $p$ on $\mathbb{X}_s$, where $\mathbb{X}_s$ is the compact support of $X_t$.*

The bounded support assumption facilitates use of Fourier series and splines. Unbounded support could be handled with appropriate choice of series (e.g. the Hermite polynomials). With Fourier series, we need to rescale $\mathbb{X}_s$ to either $\overline{\mathbb{X}} \equiv [0, 2\pi]^d$ or $\mathbb{X} \equiv [\nu, 2\pi - \nu]^d$ for some small $\nu > 0$. The condition on $\mu$ is restrictive, but is not uncommon in nonparametric estimation using Fourier series methods (e.g. Andrews (1991, Section 4), Gallant and Souza (1991), and Wooldridge (1992, Example 3.1)). It implies that the Lebesgue measure is absolutely continuous with respect to $\mu$, which permits application of Edmunds and Moscatelli's (1977) results on Fourier series.

ASSUMPTION A.2: *Put* $\varepsilon_t \equiv Y_t - E(Y_t|X_t)$. *(a)* $0 < E(\varepsilon_t^2|X_t) = \sigma_o^2$ *a.s.; (b)* $0 < E(\varepsilon^4) < \infty$ *and* $0 < \sup_{x \in \mathbb{X}_s} E(\varepsilon_t^4|X_t = x) < c^{-1} < \infty$.

The homoskedasticity assumption is a convenient but not vital condition. It greatly simplifies our test statistic. We treat the heteroskedastic case in Theorem A.3 of the Appendix. The moment condition on $\varepsilon_t$ helps ensure the asymptotic normality of our statistics.

Given $E(Y_t^2) < \infty$, there exists a measurable function $\theta_o$ such that $\theta_o(X_t) = E(Y_t|X_t)$ a.s. A parametric model for $\theta_o$ forms the basis of our null hypothesis.

ASSUMPTION A.3: *(a) Let* $A$ *be a subset of* $\mathbb{R}^q$, $q \in \mathbb{N}$. *For each* $n \in \mathbb{N}$ *the function* $f_n : \mathbb{X}_s \times A \to \mathbb{R}$ *is such that for each* $\alpha \in A$, $f_n(\cdot, \alpha)$ *is measurable, and* $f_n(X_t, \cdot)$ *is continuous a.s. on* $A$, *with* $f_n^2(X_t, \cdot) \leq D_n(X_t)$, *where* $D_n(X_t)$ *is integrable uniformly in* $n$; *(b) for each* $n$, $f_n(X_t, \cdot)$ *is twice continuously differentiable a.s. on* $A$ *with* $\|\nabla f_n(X_t, \cdot)\|^2$ *and* $\|\nabla^2 f_n(X_t, \cdot)\|$ *dominated by* $D_n(X_t)$.

The dependence of $f_n$ on $n$ will be used to generate local alternatives. Our next assumption specifies the behavior of the parametric estimator $\hat{\alpha}_n$.

ASSUMPTION A.4: $\{\hat{\alpha}_n\}$ *is a sequence of random q-vectors such that there exists a nonstochastic sequence* $\{\alpha_n^* \in \text{int } A\}$ *such that* $n^{1/2}(\hat{\alpha}_n - \alpha_n^*) = O_p(1)$.

It is not necessary to be more specific about the asymptotic behavior of $\hat{\alpha}_n$, as it will have no impact on the limit distribution of our test statistics. In other words, we can estimate $\alpha_o$ and proceed as if it were known in forming the test statistics. This greatly simplifies calculation of our test statistics. Such estimators as nonlinear least squares, generalized method of moment or adaptive efficient weighted least squares (e.g., White and Stinchcombe (1991)) estimators satisfy Assumption A.4.

To state our theorems, we define a class of local alternatives

$$H_{an} : f_n(X_t, \alpha_n^*) = \theta_o(X_t) + (p_n^{1/4}/n^{1/2})g(X_t),$$

where $g$ is square integrable on $\mathbb{X}_s$. The null hypothesis $H_o$ occurs if $g = 0$. As in Gallant and Jorgenson (1979) and Gallant and White (1988), we let the model

approach the DGP rather than vice versa. This leads to a much simpler analysis and delivers conclusions identical to those reached by fixing the model and moving the DGP appropriately.

A key consideration in constructing our tests is the number of series terms $p_n$. To gain insight into what determines the admissible rates for $p_n$, we decompose $\hat{m}_n$ under $H_o$ as follows:

$$n\hat{m}_n = \left\{ \sum_{t=1}^{n} \varepsilon_t \psi_{nt}' \right\} \{ \Psi_n' \Psi_n \}^{-1} \left\{ \sum_{t=1}^{n} \psi_{nt} \varepsilon_t \right\}$$
$$+ \sum_{t=1}^{n} \{ \theta_n^*(X_t) - \theta_o(X_t) \} \varepsilon_t + O_p(1),$$

where $\psi_{nt} = (\psi_1(X_t), \ldots, \psi_{p_n}(X_t))'$ is a $p_n \times 1$ vector, $\Psi_n = (\psi_{n1}, \ldots, \psi_{nn})'$ is an $n \times p_n$ matrix, and $\theta_n^*$ can be viewed as solving $\min_{\theta \in \Theta_n} E\{Y_t - \theta(X_t)\}^2$. The first term is a variance effect, growing as $p_n \to \infty$; the second term is a bias effect, vanishing as $p_n \to \infty$ sufficiently fast. We use the first term to determine the limit distribution by controlling the second term to be of smaller order. For this we need to control $\Theta_n$ so that $\rho(\theta_n^*, \theta_o)$ vanishes sufficiently fast, where $\rho(\theta_1, \theta_2) \equiv [\int_{\mathbb{X}_s} \{\theta_1(x) - \theta_2(x)\}^2 \mu(dx)]^{1/2}$ is an $L_2$ norm. This can be achieved by letting $p_n$ grow quickly or by nesting the parametric model in $\Theta_n$. We note that $\tilde{m}_n$ can also be decomposed similarly, although the bias term has a different expression.

A key condition for asymptotic normality of the first term (after proper standardization and recentering) is

$$\sup_{1 \le t \le n} \left\{ \psi_{nt}' (\Psi_n' \Psi_n)^{-1} \psi_{nt} \right\} \xrightarrow{P} 0.$$

(See Theorem A.1 in the Appendix.) Since $\sup_j \sup_{x \in \mathbb{X}_s} |\psi_j(x)| < \infty$ given our choices of $\{\psi_j\}$ specified below, a sufficient condition for this is $p_n / \lambda_{\min} \{\Psi_n' \Psi_n\} \xrightarrow{P} 0$. This restricts the growth of $p_n$, especially when $\lambda_{\min} \{E(\Psi_n' \Psi_n / n)\}$ is not bounded below. (We ensure this condition by relating $\lambda_{\min} \{\Psi_n' \Psi_n / n\}$ to $\lambda_{\min} \{E(\Psi_n' \Psi_n / n)\}$ using Gallant and Souza's (1991, Theorem 4) uniform strong law for $\lambda_{\min} \{\Psi_n' \Psi_n / n\}$.) Therefore, one must choose $p_n$ properly so that the first term becomes dominant and asymptotically normal. One of our objectives is to obtain an explicit rate for $p_n$ that implies asymptotic normality of our test statistics under $H_o$.

We first consider nonparametric estimators based on the trigonometric series on $\mathbb{X}_s \equiv \overline{\mathbb{X}} (= [0, 2\pi]^d)$:

$$(3.1) \qquad \Theta_n = \left\{ \theta : \overline{\mathbb{X}} \to \mathbb{R} \mid \theta(x) = u_0 + \sum_{i=1}^{I_n} \sum_{j=1}^{J_n} \left\{ u_{ij} \cos(jk_i'x) + v_{ij} \sin(jk_i'x) \right\} \right.$$
$$\left. = \sum_{j=1}^{p_n} \beta_j \psi_j(x) \right\},$$

where $I_n, J_n \in \mathbb{N}, u_0, u_{ij}$ and $v_{ij} \in \mathbb{R}, k_i \in \mathbb{K}_n \equiv \{k_i : i = 1, \ldots, I_n\}$ is an elementary multi-index, a $d \times 1$ vector of integers (for details of the construction

of $k_i$, see Gallant (1981)), $(\beta_1, \ldots, \beta_{p_n}) = (u_0, \beta_{(1)}, \ldots, \beta_{(I_n)})$, $\beta_{(i)} = (u_{i1}, v_{i1}, \ldots, u_{iJ_n}, v_{iJ_n})$, and the sequence $\{\psi_j(x)\}$ represents the constant function and the trigonometric functions $\cos(jk_i'x)$ and $\sin(jk_i'x)$ written in an order corresponding to the $\beta_j$ defined above. The number of series terms is $p_n = 1 + 2I_n J_n \simeq K_n^d$, where $K_n$ is the order of the Fourier series in (3.1).

First, suppose that $\theta_o$ is a periodic function defined on $\overline{\mathbb{X}}$ that is continuously differentiable up to order $r \in \mathbb{N}$ on a subset containing $\overline{\mathbb{X}}$. Edmunds and Moscatelli (1977, p. 15) show that there exists a sequence $\{\theta_n^* \in \Theta_n\}$ for $\Theta_n$ as in (3.1) such that

$$\rho(\theta_n^*, \theta_o) = O(K_n^{-r}) = O(p_n^{-r/d}).$$

On the other hand, $\lambda_{\min}\{E(\Psi_n'\Psi_n/n)\}$ is bounded below given that the density $p$ of $X_t$ is bounded away from zero on $\overline{\mathbb{X}}$. Applying Gallant and Souza's (1991, Theorem 4) uniform strong law for $\lambda_{\min}\{\Psi_n'\Psi_n/n\}$, we can ensure that $\lambda_{\min}\{\Psi_n'\Psi_n/n\}$ is also bounded away from zero from below. This permits a relatively fast rate for $p_n$.

In stating our specification testing results, we let

$$\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n \hat{\varepsilon}_{nt}^2 \quad \text{or} \quad \hat{\sigma}_n^2 = (n-p_n)^{-1} \sum_{t=1}^n \hat{\eta}_{nt}^2,$$

where $\hat{\varepsilon}_{nt} = Y_t - \hat{f}_{nt}$, $\hat{\eta}_{nt} = Y_t - \hat{\theta}_n(X_t)$, $\hat{f}_{nt} = f(X_t, \hat{\alpha}_n)$, and $\hat{\theta}_n = \text{argmin}_{\Theta_n} n^{-1} \sum_{t=1}^n \{Y_t - \theta(X_t)\}^2$ for some appropriate $\Theta_n$.

THEOREM 3.1: *Suppose Assumptions A.1–A.4 hold and $\theta_o$ is a periodic (in every coordinate) function defined on $\overline{\mathbb{X}} = [0, 2\pi]^d$ that is continuously differentiable up to order $r \in \mathbb{N}$ on $\overline{\mathbb{X}}$, where $0 < r < \infty$. Define $M_n$ and $\tilde{M}_n$ as in (2.1) and (2.2), and $\Theta_n$ as in (3.1). Suppose that $p_n^3/n \to 0$, $p_n^{4r+d}/n^{2d} \to \infty$, where $4r \geq 5d$. Then (i) under $H_{an}$, $M_n - \tilde{M}_n \xrightarrow{P} 0$ and*

$$M_n \xrightarrow{d} N(\delta, 1) \quad \text{and} \quad \tilde{M}_n \xrightarrow{d} N(\delta, 1),$$

*where $\delta = E\{g^2(X)\}/(2\sigma_o^4)^{1/2}$; (ii) under $H_A$ and for any sequence $\{C_n\}$, $C_n = o(n/p_n^{1/2})$,*

$$P[M_n > C_n] \to 1 \quad \text{and} \quad P[\tilde{M}_n > C_n] \to 1.$$

The rate $p_n^3/n \to 0$ helps ensure the asymptotic normality for $M_n$ and $\tilde{M}_n$. When $X_t$ is nonrandom, a faster rate for $p_n$ can be obtained. The slower rates here are due to imposing the strong law on $\lambda_{\min}\{\Psi_n'\Psi_n\}$ of Gallant and Souza (1991, Theorem 4). The rate $p_n^{4r+d}/n^{2d} \to \infty$, $4r \geq 5d$, ensures that the bias $\rho(\theta_n^*, \theta_o)$ vanishes sufficiently fast.

Theorem 3.1 shows that consistent tests based on $\hat{m}_n$ and $\tilde{m}_n$ can be obtained without using weighting, sample-splitting and non-nested testing procedures. Besides avoiding the features associated with these approaches, a benefit of our

approach is that it delivers test statistics diverging under $H_A$ at a rate of $n/p_n^{1/2}$, faster than the parametric rate $n^{1/2}$. (A parametric test statistic that has a null unit normal distribution typically diverges at $n^{1/2}$ under the alternative.) A cost for this is that our tests can only detect local alternatives of $O(p^{1/4}/n^{1/2})$, slightly slower than $n^{-1/2}$ due to $p_n \to \infty$. However, it can be shown that this rate is still faster than the local alternatives of Lee (1988), Yatchew (1992), and Wooldridge (1992) under standard regularity conditions. In addition, our admissible rates for $p_n$ are faster than those permitted by Wooldridge (1992) and Yatchew (1992). This suggests that our test may have superior power properties. We note that both our tests are one-sided, because asymptotically, negative values of our test statistics can occur only under $H_o$. Under $H_A$ our test statistics always tend to a positive increasing number.

Of course, the periodicity of $\theta_o$ is quite a restrictive assumption. Relaxing this will slow the convergence rate of the trigonometric series on $\mathbb{X}$, due to the Gibb's phenomenon (boundary effects). For example, when $p$ is uniform on $[0, 2\pi]$, the trigonometric series gives $\rho(\theta_n^*, \theta_o) = O(p_n^{-1/2})$ for $\theta_o(x) = \alpha_o x$, $x \in [0, 2\pi]$. Such a slow rate might make it difficult to have the bias negligible while maintaining asymptotic normality of our tests.

The boundary problem can be avoided and the same convergence rate maintained if we use Gallant's FFF series defined on $\mathbb{X} = [\nu, 2\pi - \nu]^d$ for some small $\nu > 0$:

$$(3.2) \qquad \Theta_n = \left\{ \theta : \mathbb{X} \to \mathbb{R} \quad \vartheta(x) = u_0 + \sum_{i=1}^{d} b_r x_i + \sum_{i=1}^{d} \sum_{j=1}^{i} c_{ij} x_i x_j \right.$$

$$\left. + \sum_{i=1}^{I_n} \sum_{j=1}^{J_n} \left\{ u_{ij} \cos(jk_i' x) + v_{ij} \sin(jk_i' x) \right\} = \sum_{j=1}^{p_n} \beta_j \psi_j(x) \right\},$$

where $(\beta_1, \ldots, \beta_{p_n}) = (u_0, \beta_{(0)}, \beta_{(1)}, \ldots, \beta_{(I_n)})$, $\beta_{(0)} = (b_1, \ldots, b_d, c_{11}, c_{12}, \ldots, c_{dd})$, $\beta_{(i)}, i = 1, \ldots, J_n$, is as in (3.1), and the sequence $\{\psi_j(x)\}$ now includes a constant, $x_i, x_i x_j$, and $\cos(jk_i' x)$ and $\sin(jk_i' x)$ written in an order corresponding to the $\beta_j$. A key difference of $\Theta_n$ in (3.2) from $\Theta_n$ in (3.1) is that $\Theta_n$ in (3.2) is defined on $\mathbb{X} = [\nu, 2\pi - \nu]^d$, which does not include the boundary. As pointed out by Gallant and Souza (1991), inclusion of $x_i$ and $x_i x_j$ helps improve finite sample performance and provides a means to test economic hypotheses. The same convergence rate is still obtained if only trigonometric series are included in (3.2).

If $\theta_o$ is continuously differentiable up to order $r$ on a subset containing $\mathbb{X}$, then there exists a periodic function defined on $\overline{\mathbb{X}} = [0, 2\pi]^d$ that is continuously differentiable up to $r$ order on a subset containing $\overline{\mathbb{X}}$ and coincides with $\theta_o$ on $\mathbb{X}$. As pointed out by Gallant (1981), Edmunds and Moscatelli's (1977) results can be applied to show that there is a sequence $\{\theta_n^* \in \Theta_n\}$ such that the same convergence rate is obtained as in the periodic case. A problem arises here

because $\lambda_{\min}\{E(\Psi_n'\Psi_n/n)\}$ is not bounded below for the FFF series on $\mathbb{X}$. (The same is true if $\Theta_n$ in (3.2) does not include $x_i$ and $x_r x_j$.) Gallant and Souza (1991, Section 5) show that for the FFF series on $\mathbb{X}$, $\lambda_{\min}\{E(\Psi_n'\Psi_n/n)\} = o(p_n^{-(s+\varepsilon)/d})$ for every $s \in \mathbb{N}$ and every $\varepsilon > 0$. Thus, it appears that $\lambda_{\min}\{\Psi_n'\Psi_n/n\}$ vanishes rapidly. This restricts the growth of $p_n$. Given such a slow growth of $p_n$, this also restricts the class of $\theta_o$ that permits quick reduction of the bias $\rho(\theta_n^*, \theta_o)$.

THEOREM 3.2: *Suppose Assumptions A.2–A.4 hold and that $\theta_o$ is continuously differentiable of order $\infty$ on a subset containing $\mathbb{X} = [\nu, 2\pi - \nu]^d$ for some small $\nu > 0$. Define $M_n$ and $\tilde{M}_n$ as in (2.1) and (2.2), and $\Theta_n$ as in (3.2). Suppose $p_n$ is such that $a(p_n^{1/d})\ln(p_n^{1/d}) \le \beta \ln(n)$ for some $0 < \beta < 1/2$ and some increasing function $a: \mathbb{R}_+ \to \mathbb{R}_+$ with $\lim_{p_n \to \infty} a(p_n^{1/d}) = \infty$. Then (i) under $H_{an}$, $M_n - \tilde{M}_n \overset{P}{\to} 0$ and*

$$M_n \overset{d}{\to} N(\delta, 1) \quad and \quad \tilde{M}_n \overset{d}{\to} N(\delta, 1);$$

(ii) *under $H_A$ and for any sequence $\{C_n\}$, $C_n = o(n/p_n^{1/2})$,*

$$P[M_n > C_n] \to 1 \quad and \quad P[\tilde{M}_n > C_n] \to 1.$$

We thus relax the periodicity assumption of $\theta_o$ at the cost of increasing smoothness and slowing the rate for $p_n$. As discussed in Gallant and Souza (1991, Section 5), $a(p_n^{1/d})\ln(p_n^{1/d}) \le \beta \ln(n)$ implies that $n^\gamma/p_n \to \infty$ for any $\gamma > 0$, so $p_n$ can only grow at a rate slower than any fractional power of $n$. For example, if we take $a(p_n^{1/d}) = \beta \ln(n)/\ln(p_n^{1/d})$, then $p_n \le \ln(n)$. Because under $H_A$ our statistics diverge at a rate of $n/p_n^{1/2}$, it follows that the slower $p_n$ is, the faster our statistics grow asymptotically. With $p_n = \ln(n)$, our statistics grow at almost the square of the parametric rate ($n^{1/2}$). Of course, the slow rate for $p_n$ may adversely affect the finite sample performance of our tests when the sample size is not sufficiently large.

To obtain a faster rate for $p_n$, an alternative to ensure that the bias vanishes under $H_o$ is to nest the parametric model in the nonparametric specification. While this approach is applicable to nonlinear parametric models, we focus for simplicity on linear models. We consider the FFF series on the cube $\overline{\mathbb{X}}$:

(3.3)     $\Theta_n$ *is as defined in (3.2) with $\overline{\mathbb{X}} = [0, 2\pi]^d$ replacing $\mathbb{X} = [\nu, 2\pi - \nu]^d$.*

Now $\lambda_{\min}\{E(\Psi_n'\Psi_n/n)\}$ of the FFF series on $\overline{\mathbb{X}}$ vanishes only at a polynomial rate. Direct calculation shows that if the density $p$ is bounded below, then $\lambda_{\min}\{E(\Psi_n'\Psi_n/n)\} = O(p_n^{-3/d})$. (The rate $O(p_n^{-1/d})$ appearing in Gallant and Souza (1991) is a typo.) Hence, a faster rate for $p_n$ is obtainable. In fact, if we choose

(3.4)     $\Theta_n$ *is as defined in (3.3) with $c_{ij} = 0$ for all $i, j = 1, \dots, d$,*

i.e. a combination of the trigonometric series with linear terms $x_i$ only, then an even faster rate for $p_n$ can be obtained because for $\Theta_n$ in (3.4), $\lambda_{\min}\{E(\Psi_n'\Psi_n/n)\}$ $= O(p_n^{-1/d})$. We give a formal result as follows.

THEOREM 3.3: *Suppose Assumptions A.1–A.2 hold and $E(X_t X_t')$ is nonsingular. Let $f_n(X_t, \alpha) = X_t'\alpha + (p_n^{1/4}/n^{1/2})g(X_t)$ for $\alpha \in$ int $A$, where $A$ is a subset of $\mathbb{R}^q$, $q \in \mathbb{N}$, and $g$ is square integrable with $E\{X_t g(x_t)\} = 0$. Let $\hat{\alpha}_n = \operatorname{argmin}_{\alpha \in A} n^{-1}\sum_{t=1}^n (Y_t - X_t'\alpha)^2$. Define $M_n$ and $\tilde{M}_n$ as in (2.1) and (2.2), and either (a) $\Theta_n$ is as in (3.3) with $p_n = o(n^{d/3(d+2)})$ or (b) $\Theta_n$ is as in (3.4) with $p_n = o(n^{d/(3d+2)})$. (i) Suppose $\theta_o(X_t) = X_t'\alpha_o$ a.s. for some $\alpha_o \in$ int $A$. Then $M_n - \tilde{M}_n \xrightarrow{\text{p}} 0$, and*

$$M_n \to N(\delta, 1) \quad \text{and} \quad \tilde{M}_n \to N(\delta, 1);$$

(ii) *suppose $\theta_o$ is square integrable on $\overline{\mathbb{X}}$ with respect to $\mu$. Then under $H_A$ and for any nonstochastic sequence $\{C_n\}$, $C_n = o(n/p_n^{1/2})$,*

$$P[M_n > C_n] \to 1 \quad \text{and} \quad P\big[\tilde{M}_n > C_n\big] \to 1.$$

For both $\Theta_n$ as in (3.3) and (3.4), admissible rates for $p_n$ are faster than those permitted by $\Theta_n$ in (3.2). This approach also applies to testing fixed order polynomial models by extending $\Theta_n$ in (3.3) to include the null polynomial model. One can expect that the rates for $p_n$ will be adversely affected by the order of polynomial model.

Both (3.3) and (3.4) belong to Eubank and Speckman's (1990) polynomial-trigonometric series. As pointed out by Eubank and Speckman, this series fits better than the pure trigonometric series on $\overline{\mathbb{X}}$ because the (low order) polynomials alleviate the boundary effects. Since ES & J's tests can be shown to be asymptotically equivalent to our tests when the series used for our tests does not include the null linear model (Hong and White (1991, Theorem 2.6)), we also expect that the tests of Theorem 3.3 may have better power than ES & J's tests. This is investigated in our simulation.

Although $\Theta_n$ as in (3.4) permits a faster rate for $p_n$ than $\Theta_n$ as in (3.3), this does not necessarily imply that (3.4) will give better power in finite samples, because under $H_A$, $\Theta_n$ as in (3.3) may fit $\theta_o$ better. By Eubank and Speckman (1990, Theorem 4.1), when $\theta_o$ is twice continuously differentiable, $\Theta_n$ as in (3.3) could fit $\theta_o$ better asymptotically than $\Theta_n$ as in 3.4 in terms of mean squared error if the optimal $p_n$ is chosen. We compare these two series in the simulation.

Finally, we turn to regression splines. For notational simplicity, we consider only the single regressor case, with $\mathbb{X}_t \equiv [0, 1]$. Let $\Delta = \{s_i\}_{i=1}^q$ with $0 = s_1 < \cdots < s_{q+1} = 1$ be a partition of $\mathbb{X}_s$ into $q$ subintervals

$$I_j = [s_j, s_{j+1}), \qquad j = 1, \ldots, q-1 \quad \text{and} \quad I_q = [s_q, s_{q+1}].$$

Let $\bar{h}$ be the mesh size and $s \in \mathbb{N}$. Then the space of polynomial splines of order $s$ with simple knots $s_1, \ldots, s_q$ can be defined as

$$\Theta(s, q, \Delta)$$
$$= \{\theta : \mathbb{X}_s \to \mathbb{R} \mid \theta(x) = p_j(x) \text{ for } x \in I_j, \text{ where } p_j(x) \text{ is a } s\text{th order}$$
$$\text{polynomial such that } p_j(x) \text{ is } C^{s-2} \text{ at knot } s_j, \quad j = 1, \ldots, q\}.$$

The dimension of $\Theta(s, q_n, \Delta_n)$ is $p_n \equiv s + q_n$. Let $\{\psi_j(x)\}$, $j = 1, \ldots, p_n$, be a basis spanning $\Theta(s, q_n, \Delta_n)$. Then $\Theta(s, q_n, \Delta_n)$ can be written as

$$(3.5) \qquad \Theta(s, q_n, \Delta_n) = \left\{ \theta : \mathbb{X}_s \to \mathbb{R} \,\middle|\, \theta(x) = \sum_{j=1}^{p_n} \beta_j \psi_j(x), \, \psi_j : \mathbb{X}_s \to \mathbb{R}, \, \beta_j \in \mathbb{R} \right\}.$$

It is well known (e.g., Schumaker (1981, Theorem 6.42)) that for any $\theta_o$ that is continuously differentiable up to order $r$ on a subset containing $\mathbb{X}_s$, there exists a sequence $\{\theta_n^* \in \Theta(s, p_n, \Delta_n)\}$ such that if $r \le s$, $\rho(\theta_n^*, \theta_o) = O(p_n^{-r})$. Thus, $p_n^{4r+1}/n^2 \to \infty$ suffices to ensure that the bias effect vanishes quickly under $H_{an}$. When $r > s$ and $\theta_o$ is not a polynomial of order $s$ or less, $\rho(\theta_n^*, \theta_o) = O(p_n^{-s})$, due to the "saturation effect" of the polynomial splines. In this case, $p_n^{4s+1}/n^2 \to \infty$ suffices.

A convenient choice for $\{\psi_j\}$ is the normalized $s$th order B-spline $\{N_j^s\}$ associated with knots $s_j, \ldots, s_{j+s}$ that satisfy the so-called partition of unity property, i.e. for all $s \ge 1$ and $j = 1, \ldots, p_n$,

$$\sum_{i=j+1-s}^{j} N_j^s(x) = 1 \quad \text{for all } s_j \le x < s_{j+1}.$$

From this we have $0 \le N_j^s(x) \le 1$ for all $x \in \mathbb{X}_s$ and all $j$. For more detailed discussion of $N_j^s(x)$, see e.g. Schumaker (1981, Ch. 4).

Put $N_{nt}^s \equiv (N_1^s(X_1), \ldots, N_{q_n}^s(X_t))'$. Then $\lambda_{\min}(EN_{nt}^s N_{nt}^{s'})$ is not bounded below. In the case of equally spaced knots with spacing $h_n \equiv 1/p_n = \bar{h}_n$ and $s$ given, it can be shown that $n^{-1}\sum_{t=1}^n E(N_{nt}^s N_{nt}^{s'})$ is a block-diagonal $p_n \times p_n$ symmetric matrix with $\lambda_{\min}(EN_{nr}^s N_{nt}^{s'}) = O(p_n^{-1})$. Again, this restricts the growth of $p_n$.

THEOREM 3.4: *Suppose Assumptions A.1–A.4 hold and $\theta_o$ is continuously differentiable up to order $r \in \mathbb{N}$ on a subset containing $\mathbb{X}_s = [0, 1]$. Define $M_n$ and $\tilde{M}_n$ as in (2.1) and (2.2) and $\Theta_n \equiv \Theta_n(s, q_n, \Delta_n)$ as in (3.5) with $h_n \equiv 1/p_n$ and $\{\psi_j\} = \{N_j^s\}$, the normalized $s$th order B-spline. Suppose either (a) $3 \le r \le s$, $p_n^{4r+1}/n^2 \to \infty$, $p_n^5/n \to 0$ or (b) $r > s \ge 3$, $p_n^{4s+1}/n^2 \to \infty$, $p_n^5/n \to 0$. Then (i) under $H_{an}$, $M_n - \tilde{M}_n \xrightarrow{p} 0$, and*

$$M_n \xrightarrow{d} N(\delta, 1) \quad \text{and} \quad \tilde{M}_n \xrightarrow{d} N(\delta, 1);$$

(ii) *under $H_A$ and for any sequence $\{C_n\}$, $C_n = o(n/p_n^{1/2})$,*

$$P[M_n > C_n] \to 1 \quad and \quad P[\tilde{M}_n > C_n] \to 1.$$

For unequal spacing, this result still holds if the spacing is appropriately chosen.

## 4. MONTE CARLO EVIDENCE

We compare the finite sample performance of our tests to consistent tests of Bierens (1990), ES&J, Wooldridge (1992), and Yatchew (1992) using Monte Carlo methods. Each test statistic is asymptotically unit normal under $H_o$, and except for Bierens' test, all are one-sided. We examine both size and power by testing correct specification for $E(Y_t|X_t)$ of a linear model against a number of alternatives.

Four DGP's are produced using the Gauss 386 random number generator:

$$DGP\ 1: \quad Y_t = 1 + X_{1t} + X_{2t} + \varepsilon_t \equiv X_t'\alpha_o + \varepsilon_t,$$

$$DGP\ 2: \quad Y_t = X_t'\alpha_o + 0.1(V_{1t} - \pi)(V_{2t} - \pi) + \varepsilon_t,$$

$$DGP\ 3: \quad Y_t = X_t'\alpha_o + X_t'\alpha_o \exp\left\{-0.01(X_t'\alpha_o)^{-2}\right\} + \varepsilon_t,$$

$$DGP\ 4: \quad Y_t = (X_t'\alpha_o)^{-0.5} + \varepsilon_t,$$

where $X_{1t} = (V_t + V_{1t})/2$, $X_{2t} = (V_t + V_{2t})/2$, $V_t$, $V_{1t}$, and $V_{2t}$ are i.i.d. $U[0, 2\pi]$, and $\varepsilon_t$ is i.i.d. $N(0, \sigma_o^2)$ with $\sigma_o^2 = 1$ or 4. A linear model $f_n(X_t, \alpha) = X_t'\alpha \equiv \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t}$ is correctly specified for $E(Y_t|X_t)$ under DGP 1 and misspecified under DGP's 2–4. DGP 2 originates from the alternative used by Bierens (1990), except that $V_t$, $V_{1t}$ and $V_{2t}$ are drawn from i.i.d. $U[0, 2\pi]$ instead of i.i.d. $N(0, 1)$. DGP 3 is similar to the alternative used by Eubank and Spiegelman (1990); and DGP 4 is one of the alternatives used by Wooldridge (1992).

Bierens' test involves choosing an increasing number of values of a nuisance parameter $\tau$ and a penalty term $\gamma n^\rho$, where $\gamma > 0$, $0 < \rho < 1$. We draw $\tau$ from a uniform distribution on $[1, 5]^2$, choose $(n/10) - 1$ for the number of $\tau$'s, and use two penalty rules: $(\gamma, \rho) = (0.25, 0.5)$ and $(0.25, 0.25)$.

Two nonparametric series are used for our tests of Theorem 3.3: a quadratic-trigonometric (QT) series with $p_n = [n^{0.15} \ln(n)]$, and a linear-trigonometric (LT) series with $p_n = [n^{0.24} \ln(n)]$, where $[a]$ denotes the integer closest to the real number $a$. These rates, given $d = 2$, satisfy the conditions of Theorem 3.3. Since both series include a linear model and $\hat{\alpha}_n$ is estimated by OLS, $M_n$ and $\tilde{M}_n$ are numerically identical, as is easily verified. Therefore, we need not distinguish them. The above two series with their rates are also applied to Yatchew's test with equal sample-splitting. For our tests, we use $\hat{\sigma}_n^2 \equiv (n - p_n)^{-1} \sum_{t=1}^n \hat{\eta}_{nt}^2$ to estimate $\sigma_o^2$. Use of $\hat{\sigma}_n^2 \equiv n^{-1} \sum_{t=1}^n \hat{\varepsilon}_{nt}^2$ gives similar results, with slightly less power. (See Hong and White (1991).) For Yatchew's test, we

use $\hat{\varepsilon}_{nt}$ to form an estimator for $\mathrm{var}(\varepsilon_t^2)$. Use of $\hat{\eta}_{nt}$ to form an estimator for $\mathrm{var}(\varepsilon_t^2)$ leads to strong overrejection.

For our multivariate version of ES&J's tests, a pure trigonometric series without the constant is used to satisfy the appropriate conditions. For comparison with our tests, we also choose $p_n = [n^{0.15} \ln(n)]$ and $[n^{0.24} \ln(n)]$ for ES&J's tests and use numerically identical estimators for $\sigma_o^2$.

The QT or LT series estimator is prohibited in Wooldridge's non-nested test. Like Wooldridge, we use a trigonometric series with a constant. We use two rates: $p_n = [\ln(n)]$ and $p_n = 5, 7, 9$ for $n = 100, 300, 500$ respectively. The latter is used by Wooldridge.

We use the following notations: NEW1 and NEW2 denote our tests with the QT and LT series estimators respectively; BT1 and BT2 denote Bierens' tests with $(\gamma, \rho) = (0.25, 0.5)$ and $(0.25, 0.25)$ respectively; ESJ1 and ESJ2 denote ES&J's test with $p_n = [n^{0.15} \ln(n)]$ and $[n^{0.24} \ln(n)]$ respectively; WT1 and WT2 denote Wooldridge's test with $p_n = [\ln(n)]$ and $p_n = 5, 7, 9$ for $n = 100, 300, 500$, respectively; and finally, YT1 and YT2 denote Yatchew's tests with the QT and LT series estimators respectively.

All the tests are based on the same seeds and 1000 replications, with a different seed for each $n$. We use both the asymptotic critical value (ACV) and the empirical critical value (ECV) at the 5% level. The ECV's are generated using 5000 replications, and are reported in Table I. In all cases, the ECV's of our tests and those of ES&J are less than 1.65, the 5% ACV. This differs from the finding of ES&J, who report an ECV larger than 1.65 for the classical linear model with a single fixed regressor and $n = 100$. We first examine size. Table II shows rejection rates using both the ACV and ECV of all test statistics under DGP 1. In all cases, our tests are conservative, and so are ES&J's tests. BT1 has reasonable size, but BT2 has a tendency toward overrejection. This warns against choosing too small a penalty for Bierens' test. WT1 and YT1 exhibit some overrejection; WT2 and YT2 exhibit strong overrejection, i.e., the faster $p_n$ grows, the stronger the overrejection. Therefore, both Wooldridge and Yatchew's tests exhibit potential "overfit" problems when $p_n$ grows too fast. In contrast, our tests are still conservative even if $p_n$ grows faster than the rates permitted by Theorem 3.3 (not reported here). When $\sigma_o^2$ increases, all the tests except Wooldridge's tests remain unchanged; Wooldridge's tests become worse.

Tables III–V report power performances under DGP's 2–4. Under DGP 2, misspecification of a linear model is "small" in the sense that the neglected nonlinearity is uncorrelated with $X_t$; the OLS estimator $\hat{\alpha}_n$ is consistent for $\alpha_o$. Our tests and ES&J's tests are powerful against DGP 2. NEW1 is a little more powerful than NEW2, i.e. the QT series performs better than the LT series, although the rate of $p_n$ for the QT series is slower. BT1 and BT2 have some power, but are much less powerful than our tests and those of ES&J. WT1, YT1, and YT2 have little power against DGP2. WT2 gains power when $n = 500$ and $\sigma_o^2 = 4$. When $\sigma_o^2$ increases, the power of most tests suffers, but the ranking of relative performances of all the tests remains unchanged.

TABLE I

EMPIRICAL CRITICAL VALUES AT THE 5% LEVEL

| $\sigma_o^2$ | 1 | | | 4 | | |
|---|---|---|---|---|---|---|
| $n$: | 100 | 300 | 500 | 100 | 300 | 500 |
| NEW1 | 1.02 | 1.13 | 1.15 | 1.02 | 1.13 | 1.15 |
| NEW2 | 1.25 | 1.33 | 1.37 | 1.25 | 1.33 | 1.37 |
| BT1 | 1.96 | 1.96 | 1.97 | 1.96 | 1.96 | 1.97 |
| BT2 | 2.09 | 2.22 | 2.28 | 2.09 | 2.22 | 2.28 |
| ESJ1 | 1.09 | 1.23 | 1.14 | 1.09 | 1.23 | 1.14 |
| ESJ2 | 1.24 | 1.34 | 1.35 | 1.24 | 1.34 | 1.35 |
| WT1 | 1.88 | 1.82 | 1.83 | 2.08 | 2.00 | 2.04 |
| WT2 | 1.95 | 2.03 | 2.04 | 2.21 | 2.37 | 2.42 |
| YT1 | 1.70 | 1.75 | 1.80 | 1.70 | 1.75 | 1.80 |
| YT2 | 1.89 | 1.96 | 2.04 | 1.89 | 1.96 | 2.04 |

NOTES: (a) DGP: $Y_t = X_t'\alpha_o + \varepsilon_t$, where $X_t = (1, X_{1t}, X_{2t})'$, $\alpha_o = (1,1,1)'$, $X_{1t} = (V_t + V_{1t})/2$, $X_{2t} = (V_t + V_{2t})/2$, $V_t, V_{1t}, V_{2t}$ are i.i.d. $U[0,2\pi]$, and $\varepsilon_t$ is i.i.d. $N(0, \sigma_o^2)$.
(b) 5000 replications.
(c) $n$ = sample size.

Our tests and Bierens' tests are powerful against DGP 3. Although our tests are competitive with those of Bierens, they are less robust to increase in $\sigma_o^2$. The tests of ES&J are also powerful, but are a little less powerful than our tests, in particular for small $n$ and/or $\sigma_o^2 = 4$. This seems to confirm the theoretical expectation that our tests are more powerful than ES&J's when we combine low order polynomials with trigonometric series. WT1 and WT2 have some power, but are less powerful than the preceding tests. WT2 is more powerful than WT1. Yatchew's tests are the least powerful. Again, all the tests become less powerful when $\sigma_o^2$ increases.

TABLE II

REJECTION RATES (%) AT THE 5% ASYMPTOTIC AND EMPIRICAL LEVELS UNDER DGP 1

| $\sigma_o^2$ | 1 | | | | | | 4 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$: | 100 | | 300 | | 500 | | 100 | | 300 | | 500 | |
| | acv | ecv | acv | ecv | acv | ecv | acv | ecv | acv | ecv | acv | ecv |
| NEW1 | 1.6 | 5.1 | 2.0 | 4.8 | 2.0 | 4.7 | 1.6 | 5.1 | 2.0 | 4.8 | 2.0 | 4.7 |
| NEW2 | 2.8 | 4.8 | 2.8 | 5.1 | 2.7 | 4.3 | 2.8 | 4.8 | 2.8 | 5.1 | 2.7 | 4.3 |
| BT1 | 4.8 | 4.7 | 3.6 | 3.6 | 6.1 | 5.7 | 4.8 | 4.7 | 3.6 | 3.6 | 6.1 | 5.7 |
| BT2 | 6.7 | 4.9 | 7.5 | 4.1 | 10.7 | 5.4 | 6.7 | 4.9 | 7.5 | 4.1 | 10.1 | 5.4 |
| ESJ1 | 2.0 | 5.2 | 1.3 | 4.5 | 2.2 | 4.5 | 2.0 | 5.2 | 1.3 | 4.5 | 2.2 | 4.5 |
| ESJ2 | 2.7 | 5.4 | 3.2 | 4.8 | 3.0 | 4.1 | 2.7 | 5.4 | 3.2 | 4.8 | 3.0 | 4.1 |
| WT1 | 8.7 | 5.1 | 5.3 | 3.6 | 7.5 | 5.3 | 13.2 | 5.0 | 8.4 | 3.7 | 10.1 | 5.3 |
| WT2 | 10.0 | 4.7 | 9.7 | 4.4 | 10.5 | 5.1 | 15.3 | 4.3 | 15.0 | 3.9 | 20.6 | 5.3 |
| YT1 | 7.2 | 6.2 | 6.2 | 4.9 | 8.5 | 6.4 | 7.2 | 6.2 | 6.2 | 4.9 | 8.5 | 6.4 |
| YT2 | 10.4 | 6.1 | 9.9 | 5.0 | 13.1 | 6.9 | 10.4 | 6.1 | 9.9 | 5.0 | 13.1 | 6.9 |

NOTES: (a) DGP; $Y_t = X_t'\alpha_o + \varepsilon_t$, where $X_t = (1, X_{1t}, X_{2t})'$, $\alpha_o = (1,1,1)'$, $X_{1t} = (V_t + V_{1t})/2$, $X_{2t} = (V_t + V_{2t})/2$, $V_t, V_{1t}, V_{2t}$ are i.i.d. $U[0,2\pi]$, and $\varepsilon_t$ is i.i.d. $N(0, \sigma_o^2)$.
(b) 1000 replications.
(c) acv = asymptotic critical value; ecv = empirical critical value.
(d) $n$ = sample size.

TABLE III

REJECTION RATES (%) AT THE 5% ASYMPTOTIC AND EMPIRICAL LEVELS UNDER DGP 2

| | 1 | | | | | | 4 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_o^2$ | 100 | | 300 | | 500 | | 100 | | 300 | | 500 | |
| n: | acv | ecv | acv | ecv | acv | ecv | acv | ecv | acv | ecv | acv | ecv |
| NEW1 | 32.2 | 46.5 | 85.2 | 91.6 | 99.4 | 99.9 | 7.5 | 13.7 | 17.5 | 26.5 | 37.2 | 49.9 |
| NEW2 | 28.1 | 36.8 | 75.1 | 81.5 | 96.2 | 97.9 | 8.6 | 12.4 | 12.0 | 19.5 | 28.6 | 35.1 |
| BT1 | 12.3 | 12.1 | 31.6 | 31.6 | 49.3 | 48.7 | 6.9 | 6.9 | 9.9 | 9.9 | 16.9 | 16.6 |
| BT2 | 16.0 | 10.5 | 42.8 | 31.3 | 42.8 | 31.3 | 8.2 | 6.4 | 16.8 | 9.9 | 27.1 | 17.4 |
| ESJ1 | 30.5 | 43.8 | 83.4 | 88.9 | 98.6 | 99.5 | 7.0 | 12.3 | 17.2 | 24.5 | 37.2 | 48.7 |
| ESJ2 | 26.4 | 36.2 | 72.9 | 79.1 | 95.8 | 97.6 | 8.0 | 11.7 | 13.9 | 19.4 | 27.3 | 35.5 |
| WT1 | 9.3 | 6.2 | 5.7 | 2.8 | 10.1 | 5.3 | 13.3 | 4.9 | 7.8 | 3.0 | 8.1 | 5.9 |
| WT2 | 9.3 | 4.7 | 9.9 | 4.1 | 20.6 | 5.3 | 14.8 | 3.7 | 16.2 | 4.2 | 82.6 | 70.3 |
| YT1 | 9.9 | 8.6 | 11.3 | 9.3 | 19.7 | 14.4 | 8.1 | 7.4 | 7.2 | 6.0 | 9.7 | 7.9 |
| YT2 | 14.2 | 8.4 | 18.0 | 10.4 | 27.5 | 14.2 | 11.0 | 6.6 | 11.2 | 5.8 | 15.6 | 7.8 |

NOTES: (a) DGP: $Y_t = X_t'\alpha_o + 0.1(V_{1t} - \pi)(V_{2t} - \pi) + \varepsilon_t$, where $X_t = (1, X_{1t}, X_{2t})'$, $\alpha_o = (1,1,1)'$, $X_{1t} = (V_t + V_{1t})/2$, $X_{2t} = (V_t + V_{2t})/2$, $V_t, V_{1t}, V_{2t}$ are i.i.d. $U[0, 2\pi]$, and $\varepsilon_t$ is i.i.d. $N(0, \sigma_o^2)$.
(b) 1000 replications.
(c) acv = asymptotic critical value; ecv = empirical critical value.
(d) $n$ = sample size.

Interestingly, WT1 and WT2 are the only tests that have some power against DGP 4. When $\sigma_o^2$ increases, they become slightly less powerful at the ECV. All other tests do not have power against this DGP.

To summarize our findings: (i) The sizes of our tests are conservative and robust to increase in $\sigma_o^2$; so are ES&J's tests. Bierens' test has a reasonable size when the penalty term is not chosen too small. The tests of Wooldridge and Yatchew may overreject when the nonparametric estimator "overfits" the data.

TABLE IV

REJECTION RATES (%) AT THE 5% ASYMPTOTIC AND EMPIRICAL LEVELS UNDER DGP 3

| | 1 | | | | | | 4 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\sigma_o^2$ | 100 | | 300 | | 500 | | 100 | | 300 | | 500 | |
| n: | acv | ecv | acv | ecv | acv | ecv | acv | ecv | acv | ecv | acv | ecv |
| NEW1 | 79.6 | 86.9 | 99.9 | 100.0 | 100.0 | 100.0 | 17.1 | 28.0 | 60.7 | 71.1 | 86.9 | 92.7 |
| NEW2 | 73.0 | 80.2 | 99.9 | 100.0 | 100.0 | 100.0 | 14.9 | 21.4 | 50.6 | 59.5 | 86.0 | 92.2 |
| BT1 | 87.9 | 87.9 | 100.0 | 100.0 | 100.0 | 100.0 | 38.7 | 38.6 | 83.7 | 83.7 | 96.5 | 96.4 |
| BT2 | 93.2 | 91.6 | 100.0 | 100.0 | 100.0 | 100.0 | 49.3 | 43.0 | 93.5 | 88.9 | 99.3 | 98.0 |
| ESJ1 | 70.9 | 79.8 | 99.9 | 99.9 | 100.0 | 100.0 | 12.8 | 22.8 | 50.4 | 61.3 | 78.3 | 86.8 |
| ESJ2 | 64.7 | 73.8 | 99.6 | 99.8 | 100.0 | 100.0 | 13.6 | 20.0 | 47.1 | 54.9 | 73.3 | 79.9 |
| WT1 | 44.2 | 36.8 | 71.9 | 65.3 | 88.8 | 84.1 | 32.5 | 19.5 | 42.0 | 29.3 | 56.9 | 43.5 |
| WT2 | 42.0 | 29.7 | 92.8 | 87.2 | 99.4 | 99.0 | 32.1 | 15.4 | 67.0 | 38.4 | 85.7 | 65.2 |
| YT1 | 16.1 | 14.1 | 27.6 | 24.1 | 44.5 | 38.5 | 9.0 | 7.5 | 9.8 | 8.2 | 14.3 | 11.2 |
| YT2 | 23.4 | 14.2 | 37.6 | 24.0 | 54.9 | 36.1 | 13.5 | 7.7 | 14.8 | 7.8 | 21.7 | 10.7 |

NOTES: (a) DGP: $Y_t = X_t'\alpha_o + (X_1'\alpha_o)\exp\{-0.01(X_t'\alpha_o)^{-2}\} + \varepsilon_t$, where $X_t = (1, X_{1t}, X_{2t})'$, $\alpha_o = (1,1,1)'$, $X_{1t} = (V_t + V_{1t})/2$, $X_{2t} = (V_t + V_{2t})/2$, $V_1, V_{1t}, V_{2t}$ are i.i.d. $U[0, 2\pi]$, and $\varepsilon_t$ is i.i.d. $N(0, \sigma_o^2)$.
(b) 1000 replications.
(c) acv = asymptotic critical value; ecv = empirical critical value.
(d) $n$ = sample size.

TABLE V

REJECTION RATES (%) AT THE 5% ASYMPTOTIC AND EMPIRICAL LEVELS UNDER DGP 4

| $\sigma_o^2$ | 1 | | | | | | 4 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 100 | | 300 | | 500 | | 100 | | 300 | | 500 | |
| n: | acv | ecv | acv | ecv | acv | ecv | acv | ecv | acv | ecv | acv | ecv |
| NEW1 | 1.6 | 5.4 | 2.2 | 5.0 | 2.5 | 5.2 | 1.6 | 5.3 | 2.2 | 4.7 | 2.2 | 4.8 |
| NEW2 | 3.1 | 5.4 | 2.6 | 5.2 | 3.1 | 4.8 | 2.8 | 5.0 | 2.6 | 5.3 | 2.7 | 4.7 |
| BT1 | 4.4 | 4.8 | 4.6 | 4.6 | 7.9 | 7.8 | 4.8 | 4.8 | 4.1 | 4.1 | 6.8 | 6.4 |
| BT2 | 6.5 | 5.1 | 9.4 | 4.7 | 12.9 | 7.1 | 6.6 | 4.9 | 8.9 | 3.7 | 11.5 | 5.6 |
| ESJ1 | 2.1 | 5.3 | 2.0 | 4.6 | 2.6 | 5.1 | 2.0 | 5.3 | 1.9 | 4.6 | 2.3 | 4.8 |
| ESJ2 | 2.7 | 5.7 | 3.0 | 5.3 | 3.4 | 4.9 | 2.7 | 5.4 | 2.9 | 5.0 | 3.1 | 4.3 |
| WT1 | 31.3 | 19.8 | 38.1 | 30.8 | 50.8 | 40.9 | 30.8 | 14.1 | 38.7 | 22.3 | 52.7 | 32.0 |
| WT2 | 38.6 | 22.9 | 65.3 | 43.6 | 85.5 | 67.9 | 38.3 | 14.8 | 66.2 | 27.0 | 85.7 | 46.8 |
| YT1 | 7.5 | 6.5 | 6.2 | 5.0 | 8.5 | 6.2 | 7.5 | 6.3 | 6.0 | 4.9 | 8.5 | 6.4 |
| YT2 | 10.2 | 6.2 | 9.9 | 5.2 | 13.3 | 6.9 | 10.3 | 6.2 | 9.8 | 5.0 | 13.0 | 6.9 |

NOTES: (a) DGP: $Y_t = (X_t'\alpha_o)^{-0.5} + \varepsilon_t$, where $X_t = (1, X_{1t}, X_{2t})'$, $\alpha_o = (1,1,1)'$, $X_{1t} = (V_t + V_{1t})/2$, $X_{2t} = (V_t + V_{2t})/2$, $V_t, V_1, V_{2t}$ are i.i.d. $U[0, 2\pi]$, and $\varepsilon_t$ is i.i.d. $N(0, \sigma_o^2)$.
(b) 1000 replications.
(c) acv = asymptotic critical value; ecv = empirical critical value.
(d) $n$ = sample size.

(ii) Our tests are powerful in some cases. They are often a little more powerful than ES&J's tests, and much more powerful than Yatchew's test. They are also competitive with or more powerful than Bierens's test. (iii) No one test dominates the others in power. (iv) The powers of all the tests suffer from increase in $\sigma_o^2$, as should be expected.

Because no one test dominates the others, we explore the extent to which combining these tests can capture the best features of each. We use a simple Bonferroni procedure, which gives an upper bound on the joint $p$-value of several test statistics despite their possible dependence. Let $P_1, \ldots, P_k$ be the ordered $p$-values corresponding to $k$ test statistics, with $P_1$ being the smallest. The Bonferroni procedure says to reject $H_o$ at level $\alpha$ if $P_1 \leq \alpha/k$.

Table VI reports the rejection rates of some Bonferroni procedures with various combinations of the above tests. We first consider a procedure combining all the tests. Bonferroni 1 combines NEW1, BT1, ESJ1, WT1, and YT1. The reason for choosing BT1, WT1, and YT1 is to avoid possible overrejections. This procedure has a reasonable size and is powerful against DGP's 2–3. It has a little power against DGP 4. Because Yatchew's tests are always dominated by some other test, we drop YT1 in obtaining Bonferroni 2, which now consists of NEW1, BT2, ESJ1, and WT1. There are some improvements in power, in particular against DGP 4. Since NEW1 is slightly more powerful than ESJ1 against DGP's 2–3, we drop ESJ1 in obtaining Bonferroni 3, which consists of NEW1, BT1, and WT1. There are again some improvements in power against DGP 4, with reasonable size. We also report Bonferroni 4, consisting of BT1, ESJ1, and WT1. It is slightly less powerful than Bonferroni 3. Of the four procedures, we prefer Bonferroni 3.

TABLE VI

REJECTION RATES (%) OF BONFERRONI PROCEDURES AT THE 5% ASYMPTOTIC LEVELS

| | $\sigma_o^2$ | 1 | | | 4 | | |
|---|---|---|---|---|---|---|---|
| DGP | $n$: | 100 | 300 | 500 | 100 | 300 | 500 |
| 1 | Bonferroni 1 | 4.0 | 3.1 | 5.1 | 5.1 | 3.6 | 6.1 |
| | Bonferroni 2 | 3.6 | 3.0 | 4.7 | 5.1 | 3.3 | 5.6 |
| | Bonferroni 3 | 4.6 | 3.2 | 6.1 | 6.1 | 4.4 | 7.3 |
| | Bonferroni 4 | 4.4 | 3.3 | 5.9 | 5.9 | 4.5 | 7.1 |
| 2 | Bonferroni 1 | 26.3 | 79.8 | 98.3 | 7.8 | 15.6 | 33.9 |
| | Bonferroni 2 | 27.3 | 81.1 | 98.5 | 8.3 | 16.6 | 34.7 |
| | Bonferroni 3 | 27.1 | 80.3 | 98.4 | 9.6 | 16.7 | 33.6 |
| | Bonferroni 4 | 25.0 | 76.2 | 97.6 | 9.1 | 16.2 | 32.0 |
| 3 | Bonferroni 1 | 81.7 | 99.9 | 100.0 | 27.5 | 70.7 | 93.3 |
| | Bonferroni 2 | 82.5 | 99.9 | 100.0 | 29.2 | 73.4 | 94.0 |
| | Bonferroni 3 | 84.8 | 99.9 | 100.0 | 33.1 | 76.1 | 94.2 |
| | Bonferroni 4 | 81.8 | 99.9 | 100.0 | 31.9 | 74.4 | 93.7 |
| 4 | Bonferroni 1 | 9.4 | 12.3 | 21.2 | 9.8 | 13.4 | 19.6 |
| | Bonferroni 2 | 10.5 | 14.3 | 23.9 | 10.5 | 14.9 | 23.2 |
| | Bonferroni 3 | 12.9 | 19.3 | 27.8 | 12.5 | 18.1 | 28.9 |
| | Bonferroni 4 | 12.9 | 19.3 | 27.7 | 12.6 | 18.1 | 28.8 |

NOTES: (a) Bonferroni 1 is a Bonferroni procedure consisting of NEW1, BT1, ESJ1, WT1, and YT1; Bonferroni 2 consists of NEW1, BT1, ESJ1, and WT1; Bonferroni 3 consists of EW1, BT1, and WT1; Bonferroni 4 consists of ESJ1, BT1, and WT1.
(b) 1000 replications.

## 5. CONCLUSION AND DIRECTIONS FOR FURTHER RESEARCH

This paper proposes two consistent specification tests for nonlinear parametric models via nonparametric series regressions. The test statistics grow at a rate faster than the parametric rate under misspecification, while avoiding weighting, sample-splitting and non-nested testing procedures previously used in the literature. Our approach can be viewed as a nested testing complement to Wooldridge's (1992) non-nested testing approach. It permits more flexible nonparametric estimation. Our results suggest possibly better size and better power in finite samples, as confirmed by simulation experiments, which also compare the relative performance of some related consistent tests. We examine a Bonferroni procedure to capture the best features of several alternative tests.

While our results are stated under the homoskedasticity assumption, our approach applies to heteroskedastic errors as well, with proper modification of the test statistics (see Theorem A.3 in the Appendix). Furthermore, our treatment of degenerate statistics is not restricted to regression contexts and to the two statistics studied here. For example, because of its appealing intuitive interpretation, relative entropy has been proposed and used to test certain interesting nonparametric hypotheses (e.g. Robinson (1991)). Since the nonparametric entropy estimator also vanishes faster than the parametric rate under the null hypothesis, our approach can be used to derive a well-defined distribution,

thus providing an alternative to the weighting device previously used in the literature. See Hong and White (1995) and White and Hong (1995) for details.

*Dept. of Economics, Cornell University, Uris Hall, Ithaca, NY 14853, U.S.A.*

*and*

*Dept. of Economics, University of California—San Diego, 9500 Gilman Dr., La Jolla, CA 92093, U.S.A.*

## MATHEMATICAL APPENDIX

We first state and prove Theorems A.1–A.2. Theorems 3.1–3.4 then follow as corollaries of Theorem A.2. Theorem A.3 treats heteroskedastic errors. First, we state the following conditions:

ASSUMPTION B.1: $\Theta_n = \{\theta\colon \theta(x) = \sum_{j=1}^{p_n} \beta_j \psi_j(x), \ \beta_j \in \mathbb{R} \text{ and } \psi_j\colon \mathbb{R}^d \to \mathbb{R}\}$ *is such that*
(a) $\Psi_n'\Psi_n$ *is nonsingular for all $n$ sufficiently large a.s.;*
(b) $\sup_t \psi_{nt}'(\Psi_n'\Psi_n)^{-1}\psi_{nt} \to 0$ *a.s.;*
(c) *there exists a sequence $\{\theta_n^* \in \Theta_n\}$ such that $\rho(\theta_n^*, \theta_o) = o(p_n^{1/4}/n^{1/2})$ under $H_{an}$ and $\rho(\theta_n^*, \theta_o) = o(1)$ under $H_A$.*

ASSUMPTION B.2: $\hat{\sigma}_n^2$ *is measurable such that $\hat{\sigma}_n^2 - \sigma_o^{*2} = o_P(p_n^{-1/2})$ under $H_{an}$, and $\hat{\sigma}_n^2 - \sigma_n^{*2} = o_P(1)$ under $H_A$ for some $\sigma_n^{*2}$, $0 < c \le \sigma_n^{*2} \le c^{-1} < \infty$, $n = 1, 2, \ldots$ .*

THEOREM A.1: *Suppose Assumptions A.1–A.2 and B.1(a, b)–B.2 hold. Define $W_n = \{\sum_t \varepsilon_t \psi_{nt}'\}\{\Psi_n'\Psi_n\}^{-1}\{\sum_t \psi_{nt}\varepsilon_t\}$. Let $p_n \to \infty$ as $n \to \infty$. Then $(W_n/\hat{\sigma}_n^2 - p_n)/(2p_n)^{1/2} \xrightarrow{d} N(0, 1)$.*

PROOF: Throughout this Appendix, we denote $\sum_t \equiv \sum_{t=1}^n$, $\sum\sum_{s<t} \equiv \sum_{t=2}^n \sum_{s=1}^{t-1}$, $\sum\sum\sum_{k<s<t} \equiv \sum_{t=3}^n \sum_{s=2}^{t-1} \sum_{k=1}^{s-1}$ and $\sum\sum\sum\sum_{i<k<s<t} \equiv \sum_{t=4}^n \sum_{s=3}^{t-1} \sum_{k=2}^{s-1} \sum_{i=1}^{k-1}$. For notational simplicity and without loss of generality, we set $\sigma_o^2 = 1$. Put $\varphi_{nt} \equiv (\Psi_n'\Psi_n)^{-1/2}\psi_{nt}$, $A_n \equiv \sum_t \varphi_{nt}'\varphi_{nt}\varepsilon_t^2 - p_n$ and $U_n \equiv \sum\sum_{s<t} U_{nts}$, where $U_{nts} \equiv 2\varepsilon_t \varphi_{nt}'\varphi_{ns}\varepsilon_s$. Then $W_n = p_n + U_n + A_n$. We first show that $p_n^{-1/2}A_n = o_P(1)$. Given Assumptions A.1–A.2 and the identity $\sum_t \varphi_{nt}'\varphi_{nt} = p_n$, we have $E(A_n) = 0$ and

$$E(A_n^2) = E\left\{\sum_t \varphi_{nt}'\varphi_{nt}(\varepsilon_t^2 - 1)\right\}^2$$

$$= E\left\{\sum_t \sum_s (\varphi_{nt}'\varphi_{nt})(\varphi_{ns}'\varphi_{ns})E[(\varepsilon_t^2 - 1)(\varepsilon_s^2 - 1)\mid X_1, \ldots, X_n]\right\}$$

$$= E\sum_t (\varphi_{nt}'\varphi_{nt})^2 E\left[(\varepsilon_t^2 - 1)^2 \mid X_1, \ldots, X_n\right] \le c^{-1} E\sum_t (\varphi_{nt}'\varphi_{nt})^2$$

$$\le c^{-1} p_n E \sup_t (\varphi_{nt}'\varphi_{nt}).$$

Hence, $p_n^{-1/2}A_n = o_P(1)$ by Chebyshev's inequality given Assumption B.1(b). Therefore,

(A.1)    $p_n^{-1/2}(W_n - p_n) = p_n^{-1/2}U_n + o_P(1).$

We next consider $U_n$. Because $E(U_{nts}|Z_t) = E(U_{nts}|Z_s) = 0$ for $t \neq s$, where $Z_t = (Y_t, X_t')'$, Lemma 2.1 of de Jong (1987) holds and we can use his CLT's for generalized quadratic forms. By de Jong (1987, Proposition 3.2), $U_n/S_n \xrightarrow{d} N(0,1)$ if $G_I$, $G_{II}$, and $G_{IV}$ are $o(S_n^4)$, where $S_n^2 \equiv EU_n^2$, and

$$G_I \equiv \sum_{s<t}\sum EU_{nts}^4, \quad G_{II} \equiv \sum_{k<s<t}\sum\sum (EU_{nts}^2 U_{nik}^2 + EU_{nst}^2 U_{nsk}^2 + EU_{nkt}^2 U_{nks}^2), \quad \text{and}$$

$$G_{IV} = \sum_{i<k<s<t}\sum\sum\sum (EU_{nik}U_{nis}U_{ntk}U_{nts} + EU_{nik}U_{nit}U_{nsk}U_{nst} + EU_{nis}U_{nit}U_{nks}U_{nkt})$$

$$= (1/2)\sum_{i<k}\sum \sum_{s<t}\sum E(U_{si}U_{sk}U_{nti}U_{ntk})$$

(cf. de Jong (1987, pp. 266–267)). We now verify these conditions. First, given Assumptions A.2(a) (and $\sigma_o^2 = 1$), B.1(b), and the identity $\sum_t \varphi_{nt}\varphi_{nt}' = I_n$, we have

$$S_n^2 = \sum_{s<t}\sum EU_{nts}^2 = 4\sum_{s<t}\sum E\{(\varphi_{nt}'\varphi_{ns})^2 \varepsilon_t^2\varepsilon_s^2\}$$

$$= 2E\sum_t \varphi_{nt}'\left(\sum_s \varphi_{ns}\varphi_{ns}'\right)\varphi_{nt} - 2E\sum_t (\varphi_{nt}'\varphi_{nt})^2$$

$$= 2E\sum_t \varphi_{nt}'\varphi_{nt} - 2E\sum_t (\varphi_{nt}'\varphi_{nt})^2 = 2p_n\left\{1 - E\sup_t (\varphi_{nt}'\varphi_{nt})\right\} = 2p_n(1 + o(1)).$$

Next, we compute the orders of magnitude for $G_j$, $j = I$, II, and IV. Given Assumptions A.1–A.2, B.1(b), and the two identities for $\varphi_{nt}$, we have

$$G_I = 16E\sum_{s<t}\sum (\varphi_{nt}'\varphi_{ns})^4 \varepsilon_t^4\varepsilon_s^4 \leq 16c^{-2}E\left\{\left[\sup_t (\varphi_{nt}'\varphi_{nt})\right]^2 \sum_{s<t}\sum (\varphi_{nt}'\varphi_{ns})^2\right\}$$

$$\leq 16c^{-2}E\left\{\left[\sup_t (\varphi_t'\varphi_{nt})\right]^2 \sum_t \varphi_{nt}'\varphi_{nt}\right\} = o(p_n);$$

$$G_{II} = 16E\left\{\sum_{t<s<k}\sum\sum (\varphi_{nt}'\varphi_{ns})^2(\varphi_{nt}'\varphi_{nk})^2 \varepsilon_t^4\varepsilon_s^2\varepsilon_k^2 + (\varphi_{ns}'\varphi_{nt})^2(\varphi_{ns}'\varphi_{nk})^2 \varepsilon_s^4\varepsilon_t^2\varepsilon_k^2\right.$$

$$\left. + (\varphi_{nk}'\varphi_{nt})^2(\varphi_{nk}'\varphi_{ns})^2 \varepsilon_k^4\varepsilon_t^2\varepsilon_s^2\right\} \leq 48c^{-1}E\sum_t\sum_s\sum_k (\varphi_{nt}'\varphi_{ns})^2(\varphi_{nt}'\varphi_{nk})^2$$

$$= 48c^{-1}E\sum_t \varphi_{nt}'\left(\sum_s \varphi_{ns}\varphi_{ns}'\right)\varphi_{nt}\varphi_{nt}'\left(\sum_k \varphi_{nk}\varphi_{nk}'\right)\varphi_{nt} \leq 48c^{-1}\sum_t (\varphi_{nt}'\varphi_{nt})^2 = o(p_n);$$

$$G_{IV} = 8E\sum_{i<k}\sum \sum_{s<t}\sum (\varphi_{ns}'\varphi_{ni})(\varphi_{ns}'\varphi_{nk})(\varphi_{nt}'\varphi_{ni})(\varphi_{nt}'\varphi_{nk})\varepsilon_t^2\varepsilon_s^2\varepsilon_k^2\varepsilon_i^2$$

$$\leq 2E\sum_i\sum_k\sum_s\sum_t (\varphi_{ns}'\varphi_{ni})(\varphi_{ns}'\varphi_{nk})(\varphi_{nt}'\varphi_{ni})(\varphi_{nt}'\varphi_{nk})$$

$$= 2E\sum_t \varphi_{nt}'\sum_k (\varphi_{nk}\varphi_{nk}')\sum_s (\varphi_{ns}\varphi_{ns}')\sum_i (\varphi_{ni}\varphi_{ni}')\varphi_{nt} = 2\sum_t \varphi_{nt}'\varphi_{nt} = 2p_n.$$

Hence, $G_j/S_n^4 = o(p_n^{-1})$ for $j = I$, II and $O(p_n^{-1})$ for $j = IV$. Hence, $G_j/S_n^4 \to 0$ given $p_n \to \infty$. It follows from de Jong (1987, Proposition 3.2) that $U_n/(2p_n)^{1/2} \xrightarrow{d} N(0,1)$. We then have $(W_n/\hat{\sigma}_n^2 - p_n)/(2p_n)^{1/2} = (W_n - p_n)/(2p_n)^{1/2} + (\hat{\sigma}_n^2 - 1)W_n/(2p_n)^{1/2} = U_n/(2p_n)^{1/2} + o_P(1) \xrightarrow{d} N(0,1)$ given (A.1) and Assumption B.2. Q.E.D.

THEOREM A.2: *Suppose Assumptions A.1–A.4 and B.1–B.2 hold. Define $M_n$ and $\tilde{M}_n$ as in (2.1) and (2.2). Let $p_n \to \infty$ as $n \to \infty$. Then (i) under $H_{an}$, $M_n - \tilde{M}_n \xrightarrow{P} 0$ and*

$$M_n \to N(\delta, 1) \quad \text{and} \quad \tilde{M}_n \to N(\delta, 1),$$

*where* $\delta = Eg^2(X)/\sqrt{2}\,\sigma_o^2$, *(ii) under $H_A$ and for any nonstochastic sequence $\{C_n\}$, $C_n = o(n/p^{1/2})$,*

$$P[M_n > C_n] \to 1 \quad and \quad P[\tilde{M}_n > C_n] \to 1.$$

PROOF: (i) Asymptotic normality: Put $\hat{\theta}_{nt} = \hat{\theta}_n(X_t)$, $\theta_{nt}^* = \theta_n^*(X_t)$, $\theta_t^o = \theta_o(X_t)$, $\hat{f}_{nt} = f_n(X_t, \hat{\alpha}_n)$, $f_{nt}^* = f_n(X_t, \alpha_n^*)$, and $g_t = g(X_t)$. We first consider $M_n$. Noting $\hat{\varepsilon}_{nt} = \varepsilon_t - (\hat{f}_{nt} - \theta_t^o)$ and $\hat{\theta}_{nt} - \theta_{nt}^* = \psi_{nt}'(\hat{\beta}_n - \beta_n^*) = \psi_{nt}'(\psi_n'\psi_n)^{-1}\Sigma_t\psi_{nt}\{\varepsilon_t - (\theta_{nt}^* - \theta_t^o)\}$, we decompose

(A.2) $\quad \hat{m}_n = n^{-1}\sum_t (\hat{\theta}_{nt} - \hat{f}_{nt})\hat{\varepsilon}_{nt} = n^{-1}\sum_t \{(\hat{\theta}_{nt} - \theta_{nt}^*)\hat{\varepsilon}_{nt} + (\theta_{nt}^* - \theta_t^o)\hat{\varepsilon}_{nt} - (\hat{f}_{nt} - \theta_t^o)\hat{\varepsilon}_{nt}\}$

$$= n^{-1}\left(\sum_t \varepsilon_t\psi_{nt}'\right)(\Psi_n'\Psi_n)^{-1}\left(\sum_t \psi_{nt}\varepsilon_t\right)$$

$$- n^{-1}\left(\sum_t (\theta_{nt}^* - \theta_t^o)\psi_{nt}'\right)(\Psi_n'\Psi_n)^{-1}\left(\sum_t \psi_{nt}\varepsilon_t\right)$$

$$- n^{-1}\sum_t (\hat{\theta}_{nt} - \theta_{nt}^*)(\hat{f}_{nt} - \theta_t^o) + n^{-1}\sum_t (\theta_{nt}^* - \theta_t^o)\varepsilon_t - n^{-1}\sum_t (\theta_{nt}^* - \theta_t^o)(\hat{f}_{nt} - \theta_t^o)$$

$$- n^{-1}\sum_t (\hat{f}_{nt} - \theta_t^o)\varepsilon_t + n^{-1}\sum_t (\hat{f}_{nt} - \theta_t^o)^2$$

$$\equiv n^{-1}W_n - A_{n1} - A_{n2} + A_{n3} - A_{n4} - A_{n5} + A_{n6}, \quad \text{say},$$

where the first two terms come from $n^{-1}\Sigma_t(\hat{\theta}_{nt} - \theta_{nt}^*)\varepsilon_t$. We first show $A_{nj} = o_P(p_n^{1/2}/n)$ for $1 \le j \le 6$ and then use Theorem A.1. Given Assumptions A.1–A.2(a), we have

$$E(A_{n1}) = n^{-2}\sigma_o^2 E\left\{\left(\sum_t (\theta_{nt}^* - \theta_t^o)\psi_{nt}'\right)(\Psi_n'\Psi_n)^{-1}\left(\sum_t \psi_{nt}(\theta_{nt}^* - \theta_t^o)\right)\right\}$$

$$\le \sigma_o^2 n^{-2}\sum_t (\theta_{nt}^* - \theta_t^o)^2,$$

where the last inequality follows from the basic projection inequality (PI): $(\Sigma_t h_t\psi_{nt}')(\Psi_n'\Psi_n)^{-1}$ $(\Sigma_t\psi_{nt}h_t) \le \Sigma_t h_t^2$ for any measurable function $h_t \equiv h(X_t)$. It follows that $A_{n1} = O_P(n^{-1/2}\rho(\theta_n^*, \theta_o))$ $= o_P(p_n^{1/4}/n)$ by Chebyshev's inequality and Assumption B.1(c). Given B.1(c), we also have $A_{n3} = O_P(n^{-1/2}\rho(\theta_n^*, \theta_o)) = o_P(p_n^{1/4}/n)$ by Chebyshev's inequality. Next, noting $\hat{f}_{nt} - \theta_t^o = \hat{f}_{nt} - f_{nt}^*$ $+ (p_n^{1/4}/n^{1/2})g_t$ under $H_{an}$ and using a two term Taylor expansion, we have

(A.3) $\quad A_{n2} = (\hat{\alpha}_n - \alpha_n^*)'\left(n^{-1}\sum_t (\hat{\theta}_{nt} - \theta_{nt}^*)\nabla_\alpha f_{nt}^*\right)$

$$+ (1/2)(\hat{\alpha}_n - \alpha_n^*)'\left(n^{-1}\sum_t (\hat{\theta}_{nt} - \theta_{nt}^*)\nabla_\alpha^2\ddot{f}_{nt}\right)(\hat{\alpha}_n - \alpha_n^*)$$

$$+ (p_n^{1/4}/n^{1/2})n^{-1}\sum_t (\hat{\theta}_{nt} - \theta_{nt}^*)g_t,$$

where $\nabla_\alpha^2\ddot{f}_{nt} \equiv \nabla_\alpha^2 f_n(X_t, \ddot{\alpha}_n)$ with a different $\ddot{\alpha}_n$ such that $\|\ddot{\alpha}_n - \alpha_n^*\| \le \|\hat{\alpha}_n - \alpha_n^*\|$ appearing in each row of $\nabla_\alpha^2 f_{nt}$. For the first term in (A.3), we have

(A.4) $\quad n^{-1}\sum_t (\hat{\theta}_{nt} - \theta_{nt}^*)\nabla_\alpha f_{nt}^* = n^{-1}\left(\sum_t (\theta_{nt}^* - \theta_t^o)\psi_{nt}'\right)(\Psi_n'\Psi_n)^{-1}\left(\sum_t \psi_{nt}\nabla_\alpha f_{nt}^*\right)$

$$+ n^{-1}\left(\sum_t \varepsilon_t\psi_{nt}'\right)(\Psi_n'\Psi_n)^{-1}\left(\sum_t \psi_{nt}\nabla_\alpha f_{nt}^*\right)$$

$$= O_P(\rho(\theta_n^*, \theta_o) + n^{-1/2}) = o_P(p_n^{1/4}/n^{1/2}).$$

by the Cauchy-Schwarz inequality (for the first term), Chebyshev's inequality (for the second term), the projection inequality (PI) and Assumption B.1(c). Similarly, for the last term of (A.3), we have

$$(A.5) \qquad n^{-1} \sum_t (\hat{\theta}_{nt} - \theta_{nt}^*) g_t = O_P(\rho(\theta_n^*, \theta_o) + n^{-1/2}) = o_P(p_n^{1/4}/n^{1/2}).$$

Finally, for the second term of (A.3), we have

$$(A.6) \qquad \left| n^{-1} \sum_t (\hat{\theta}_{nt} - \theta_{nt}^*) \nabla_\alpha^2 \ddot{f}_{nt} \right| \le \left( n^{-1} \sum_t (\hat{\theta}_{nt} - \theta_{nt}^*)^2 \right)^{1/2} \left( n^{-1} \sum_t \|\nabla_\alpha^2 \ddot{f}_{nt}\|^2 \right)^{1/2}$$

$$= O_P(p_n^{1/2}/n^{1/2})$$

given Assumptions A.1–A.3 and B.1(c). Combining (A.3)–(A.6), we have $A_{n2} = o_P(p_n^{1/2}/n)$. Next, we consider the remaining terms. Given Assumptions A.3–A.4 and B.1(c),

$$A_{n4} = (\hat{\alpha}_n - \alpha_n^*)' n^{-1} \sum_t (\theta_{nt}^* - \theta_t^o) \nabla_\alpha \ddot{f}_{nt} + (p_n^{1/4} n^{1/2}) n^{-1} \sum_t (\theta_{nt}^* - \theta_t^o) g_t$$

$$= o_P(p_n^{1/2}/n)$$

by the Cauchy-Schwarz inequality, where $\nabla_\alpha \ddot{f}_{nt} \equiv \nabla_\alpha f_n(X_t, \ddot{\alpha}_n)$, with a different $\ddot{\alpha}_n$ appearing in each row of $\nabla_\alpha \ddot{f}_{nt}$. Next, given Assumptions A.1–A.4, we have

$$A_{n5} = (\hat{\alpha}_n - \alpha_n^*)' n^{-1} \sum_t \nabla_\alpha f_{nt}^* \varepsilon_t + (1/2)(\hat{\alpha}_n - \alpha_n^*)' \left( n^{-1} \sum_t \nabla^2 \ddot{f}_{nt} \varepsilon_t \right) (\hat{\alpha}_n - \alpha_n^*)$$

$$+ (p_n^{1/4}/n^{1/2}) n^{-1} \sum_t g_t \varepsilon_t$$

$$= O_P(n^{-1}) + O_P(n^{-1}) + O_P(p_n^{1/4}/n) = O_P(p_n^{1/4}/n)$$

by Chebyshev's inequality (for the first and last terms) and the Cauchy-Schwarz inequality (for the second term), where $\nabla_\alpha^2 \ddot{f}_{nt} \equiv \nabla_\alpha^2 f_n(X_t, \ddot{\alpha}_n)$, with a different $\ddot{\alpha}_n$ appearing in each row. Finally, we have

$$A_{n6} = n^{-1} \sum_t (\hat{f}_{nt} - f_{nt}^*)^2 + 2(p_n^{1/4}/n^{1/2}) n^{-1} \sum_t (\hat{f}_{nt} - f_{nt}^*) g(X_t)$$

$$+ (p_n^{1/2}/n) \left( n^{-1} \sum_t g_t^2 \right)$$

$$= O_P(n^{-1}) + O_P(p_n^{1/4}/n) + (p_n^{1/2}/n) Eg^2(X)\{1 + o_P(1)\}$$

by Assumptions A.3–A.4 and the weak law of large numbers, where $n^{-1} \sum_t (\hat{f}_{nt} - f_{nt}^*)^2 = (\hat{\alpha}_n - \alpha_n^*)'(n^{-1} \sum_t \nabla_\alpha \ddot{f}_{nt} \nabla_\alpha' f_{nt})(\hat{\alpha}_n - \alpha_n^*) = O_P(n^{-1})$. It follows from (A.2) that $\hat{m}_n = (p_n^{1/2}/n) Eg^2(X) + n^{-1} W_n + o_P(p_n^{1/2}/n)$. The desired result follows from Theorem A.1 by proper standardization.

We now treat $\bar{M}_n$. We decompose

$$(A.7) \qquad \bar{m}_n = 2n^{-1} \sum_t (\hat{\theta}_{nt} - \theta_t^o) \varepsilon_t - n^{-1} \sum_t (\hat{\theta}_{nt} - \theta_t^o)^2 - n^{-1} \sum_t (\hat{f}_{nt} - \theta_t^o) \varepsilon_t$$

$$+ n^{-1} \sum_t (\hat{f}_{nt} - \theta_t^o)^2$$

$$= 2n^{-1} \sum_t (\hat{\theta}_{nt} - \theta_t^o) \varepsilon_t - n^{-1} \sum_t (\hat{\theta}_{nt} - \theta_t^o)^2 - A_{n5} + A_{n6}.$$

For the first term of (A.7),

$$(A.8) \quad n^{-1} \sum_t (\hat{\theta}_{nt} - \theta_t^o)\varepsilon_t = n^{-1} \sum_t (\hat{\theta}_{nt} - \theta_{nt}^*)\varepsilon_t + n^{-1} \sum_t (\theta_{nt}^* - \theta_t^o)\varepsilon_t = n^{-1}W_n - A_{n1} + A_{3n}.$$

For the second term of (A.7), after some manipulation, we can obtain

$$(A.9) \quad n^{-1} \sum_t (\hat{\theta}_{nt} - \theta_t^o)^2 = n^{-1} \sum_t (\hat{\theta}_{nt} - \theta_{nt}^*)^2 + n^{-1} \sum_t (\theta_{nt}^* - \theta_t^o)^2$$

$$+ 2n^{-1} \sum_t (\hat{\theta}_{nt} - \theta_{nt}^*)(\theta_{nt}^* - \theta_t^o)$$

$$= n^{-1}W_n - n^{-1}\left(\sum_t (\theta_{nt}^* - \theta_t^o)\psi_{nt}'\right)(\Psi_n'\Psi_n)^{-1}\left(\sum_t \psi_{nt}(\theta_{nt}^* - \theta_t^o)\right)$$

$$+ n^{-1} \sum_t (\theta_{nt}^* - \theta_t^o)^2$$

$$= n^{-1}W_n + O_P(\rho^2(\theta_n^*, \theta_o)) = n^{-1}W_n + o_P(p_n^{1/2}/n)$$

given Assumption B.1(c), where the last two terms are $o_P(p_n^{1/2}/n)$ by the projection inequality (PI) and Markov's inequality. Combining (A.7)–(A.9) and noting $A_{n6} = (p_n^{1/2}/n)Eg^2(X) + o_P(p_n^{1/2}/n)$ and $A_{nj} = o_P(p_n^{1/2}/n)$ for $j \neq 6$, we obtain $\tilde{m}_n = (p_n^{1/2}/n)Eg^2(X) + n^{-1}W_n + o_P(p_n^{1/2}/n)$. It follows that $\tilde{m}_n - \hat{m}_n = o_P(p_n^{1/2}/n)$. Therefore, we have $\tilde{M}_n - M_n = o_P(1)$, and $\tilde{M}_n \xrightarrow{d} N(\delta, 1)$.

(ii) Following the analogous reasoning of part (i), it can be shown that under $H_A$, $\hat{m}_n = n^{-1}\Sigma_t(f_{nt}^* - \theta_t^o)^2 + o_P(1) = E(f_{nt}^* - \theta_t^o)^2 + o_P(1)$ by the weak law of large numbers (e.g., Andrews (1988)) given Assumptions A.1 and A.3(a). The proof for $\tilde{M}_n$ is similar. Therefore, consistency follows.    Q.E.D.

In proving Theorems 3.1–3.4, we repeatedly use the following two lemmas.

LEMMA A.1 (Uniform Strong Law for $\lambda_{\min}(\Psi_n'\Psi_n/n)$): *Define* $B(p_n) \equiv \sup_t(\psi_{nt}'\psi_{nt})$. *Suppose* $p_n$ *satisfies* $B(p_n)/\lambda_{\min}\{E(\Psi_n'\Psi_n/n)\} \leq n^\beta$, $0 \leq \beta < 1/2$ *and* $p_n \leq n^\alpha$, $0 \leq \alpha < 1 - 2\beta$. *Then*

$$P[B(p_n)/\lambda_{\min}(\Psi_n'\Psi_n/n) > 2n^\beta \text{ infinitely often } (i.o.)] = 0.$$

PROOF: See Gallant and Souza (1991, Theorem 4).

LEMMA A.2: *Define* $\hat{\sigma}_n^2 = n^{-1}\Sigma_{t=1}^n \hat{\varepsilon}_{nt}^2$. *Let* $p_n/n \to 0$. *Then* $\hat{\sigma}_n^2 - \sigma_o^2 = O_P(n^{-1/2})$ *under* $H_{an}$, *and* $\hat{\sigma}_n^2 - \sigma_n^{*2} = o_P(1)$ *under* $H_A$ *for some* $0 < c \leq \sigma_n^{*2} \leq c^{-1} < \infty$, $n = 1, 2, \ldots$.

PROOF: Write $\hat{\sigma}_n^2 - \sigma_o^2 \equiv n^{-1}\Sigma_t(\varepsilon_t^2 - \sigma_o^2) - 2n^{-1}\Sigma_t \varepsilon_t(\hat{f}_{nt} - \theta_t^o) + n^{-1}\Sigma_t(\hat{f}_{nt} - \theta_t^o)^2$. For the first term, $n^{-1}\Sigma_t(\varepsilon_t^2 - \sigma_o^2) = O_P(n^{-1/2})$ by Chebyshev's inequality and Assumptions A.1–A.2. Next, we consider the two remaining terms. (i) Under $H_{an}$: in the proof of Theorem A.2 we have shown that $n^{-1}\Sigma_t \varepsilon_t(\hat{f}_{nt} - \theta_t^o) = A_{n5} = O_P(p_n^{1/4}/n)$ and $n^{-1}\Sigma_t(\hat{f}_{nt} - \theta_t^o)^2 = A_{n6} = O_P(p_n^{1/2}/n)$. It follows that $\hat{\sigma}_n^2 - \sigma_o^2 = O_P(n^{-1/2})$ given $p_n/n \to 0$. (ii) Under $H_A$ it can be shown that $n^{-1}\Sigma_t \varepsilon_t(\hat{f}_{nt} - \theta_t^o) = o_P(1)$ by the mean value expansion and Chebyshev's inequality given Assumption A.3(b) and $n^{-1}\Sigma_t(\hat{f}_{nt} - \theta_t^o)^2 = E\{f_n(X, \alpha_n^*) - \theta_o(X_t)\}^2 + o_P(1)$ by the law of large numbers given Assumptions A.1 and A.3(a). The result follows immediately with $\sigma_n^{*2} = \sigma_o^2 + E\{f_n(X, \alpha_n^*) - \theta_o(X_t)\}^2$.    Q.E.D.

PROOF OF THEOREM 3.1: We verify the conditions of Theorem A.2. (i) Assumptions A.1–A.4 are imposed directly. Assumption B.2 holds given either $\hat{\sigma}_n^2 \equiv n^{-1}\Sigma_{t=1}(Y_t - \hat{f}_{nt})^2$ (by Lemma A.2) or $\hat{\sigma}_n^2 \equiv (n - p_n)^{-1}\Sigma_{t=1}^n(Y_t - \hat{\theta}_{nt})^2$ (the proof is deferred to the end). Next, we verify the key assumption B.1. We use Lemma A.1 to verify Assumption B.1(a, b). Given Assumption A.1 ($p(x) \geq c > 0$ for all $x \in \mathbb{X}$),

$$E\{\psi_i(X)\psi_j(X)\} = \int_{\mathbb{X}} \psi_i(x)\psi_j(x)p(x)\,dx \geq c\int_{\mathbb{X}} \psi_i(x)\psi_j(x)\,dx = c\delta_{ij},$$

where $\delta_{ii} = 1$ and $\delta_{ij} = 0$ for $i \neq j$ by orthonormality of $\{\psi_j\}$. Hence, $E(\Psi_n'\Psi_n/n) = n^{-1}\sum_t E(\psi_{nt}\psi_{nt}')$ $\geq cI_{p_n}$ and $\lambda_{\min}\{E(\Psi_n'\Psi_n/n)\} \geq c > 0$ for all $n$ and $p_n$. Given $p_n = o(n^{1/3})$ and $B(p_n) \leq$ $p_n \max_{1 \leq j \leq p_n} \{\sup_{x \in \overline{\mathbb{X}}}|\psi_j(x)|\}^2 = c^{-1}p_n$ for the trigonometric series, we can set $\alpha = (1/3) - \varepsilon$ for some arbitrary small $\varepsilon > 0$ and $\beta = 1/3$ in Lemma A.1 such that $\alpha < 1 - 2\beta$. It follows that the conditions of Lemma A.1 hold for $B(p_n)/\lambda_{\min}(\Psi_n'\Psi_n/n)$ with $\beta = 1/3$. As a result, Assumption B.1(b) holds since $\sup_t \psi_{nt}(\Psi_n'\Psi_n)^{-1}\psi_{nt} \leq n^{-1} \sup_t (\psi_{nt}'\psi_{nt})/\lambda_{\min}(\Psi_n'\Psi_n/n) = o(n^{-2/3})$ a.s. Assumption B.1(a) also holds because $\lambda_{\min}(\Psi_n'\Psi_n)/(n^{1-\beta}/p_n) \to \infty$ a.s. and $n^{1-\beta}/p_n \to \infty$. Finally, given $\theta_o \in C^r(\overline{\mathbb{X}})$ is periodic ($\overline{\mathbb{X}}$ is a subset containing $\overline{\mathbb{X}}$), we have $\rho(\theta_n^*, \theta_o) = O(p_n^{-r/d})$ as argued in the text. Hence, B.1(c) holds given $p_n^{4r+d}/n^{2d} \to \infty$, $4r > 5d$. Because the conditions of Theorem A.2(i) are satisfied, asymptotic normality follows for both $M_n$ and $\tilde{M}_n$.

It remains to show $\hat{\sigma}_n^2 \equiv (n - p_n)^{-1}\sum_{t=1}^n (Y_t - \hat{\theta}_{nt})^2$ satisfies B.2. Put $n' = n - p_n$. Then

$$\hat{\sigma}_n^2 - \sigma_o^2 = n'^{-1}\sum_t (\varepsilon_t^2 - \sigma_o^2) - 2n'^{-1}\sum_t \varepsilon_t(\hat{\theta}_{nt} - \theta_t^o)$$

$$+ n'^{-1}\sum_t (\hat{\theta}_{nt} - \theta_t^o)^2 + (p_n/n')\sigma_o^2.$$

Given Assumptions A.1–A.2, $n'^{-1}\sum_t(\varepsilon_t^2 - \sigma_o^2) = O_P(n^{-1/2})$ by Chebyshev's inequality. From the proof of Theorem A.2, we have shown (cf. (A.8) and (A.9)) that $n^{-1}\sum_t \varepsilon_t(\hat{\theta}_{nt} - \theta_t^o) = O_P(p_n/n)$ and $n^{-1}\sum_t(\hat{\theta}_{nt} - \theta_t^o)^2 = O_P(p_n/n)$. It follows that $\hat{\sigma}_n^2 - \sigma_o^2 = O_P(n^{-1/2})$ under $H_{an}$ and $H_A$ given Assumption B.1(c) and $p_n = o(n^{1/3})$. (ii) Consistency follows immediately from Theorem A.2(ii).
Q.E.D.

PROOF OF THEOREM 3.2: (i) The only difference from Theorem 3.1 is that now $\lambda_{\min}\{E(\Psi_n'\Psi_n/n)\}$ $= O(p_n^{-(s+\varepsilon)/d})$ for every positive integer $s \in \mathbb{N}$ and any $\varepsilon > 0$ (see Gallant and Souza (1991, Section 5)), so we only need verify Assumption B.1. Gallant and Souza show that for such a rapidly decreasing sequence as $\lambda_{\min}\{E(\Psi_n'\Psi_n/n)\} = O(p_n^{-(s+\varepsilon)/d})$ for any integer $s > 0$ and any $\varepsilon > 0$, there exists an equivalent characterization $\ln(\lambda_{\min}\{E(\Psi_n'\Psi_n/n)\}) = -a(p_n^{1/d})\ln(p_n^{1/d})$ for some function $a$ with $\lim_{p_n \to \infty} a(p_n^{1/d}) = \infty$. It follows that $a(p_n^{1/d})\ln(p_n^{1/d}) \leq \beta \ln(n)$ for some $\beta, 0 \leq \beta < 1/2$, ensures $B(p_n)/\lambda_{\min}\{E(\Psi_n'\Psi_n/n)\} \leq n^\beta$. Because $a(p_n^{1/d})\ln(p_n^{1/d}) \leq \beta \ln(n)$ implies that $p_n$ grows slower than any fractional power of $n$ (i.e., $n^\alpha/p_n \to \infty$ for any $\alpha > 0$), the condition $p_n \leq n^\alpha$ for some $\alpha$, $0 < \alpha < 1 - 2\beta$, also holds. Hence, the conditions of Lemma A.1 hold for $B(p_n)/\lambda_{\min}(\Psi_n'\Psi_n/n)$. It follows that Assumption B.1(b) holds by Lemma A.1. Assumption B.1(a) also holds because $\lambda_{\min}(\Psi_n'\Psi_n)/(n^{1-\beta}/p_n) \to \infty$ a.s. and $n^{1-\beta}/p_n \to \infty$. Since $\theta_o$ is infinitely differentiable, Assumption B.1(c) also holds for the choice of $p_n$, following reasoning analogous to that of Gallant and Souza (1991, Section 5). Therefore, all conditions of Theorem A.2 hold, and asymptotic normality for $M_n$ and $\tilde{M}_n$ follows by Theorem A.2(i). (ii) Similar to Theorem 3.1(ii).
Q.E.D.

PROOF OF THEOREM 3.3: Given $f_n(X, \alpha) = X'\alpha + (p_n^{1/4}/n^{1/2})g(X)$ and $\alpha \in A$, $A$ a subset of $\mathbb{R}^q$, Assumption A.3 holds. Put $\alpha_n^* = \alpha^* \equiv E(XX')^{-1}E(XY)$. Then Assumption A.4 also holds because $\hat{\alpha}_n - \alpha^* = (n^{-1}\sum_t X_t X_t')^{-1} n^{-1}\sum_t X_t Y_t - \alpha^* = O_P(n^{-1/2})$. We now verify B.1–B.2. (i) Given $\theta_o(X) = X'\alpha_o$ and that $\Theta_n$ in either (3.3) or (3.4) contains the linear model, we have $\rho(\theta_n^*, \theta_o) = 0$ for all $p_n \geq d$. Hence, Assumption B.1(c) holds under $H_{an}$. Given either (a) $\Theta_n$ in (3.4) or (b) $\Theta_n$ in (3.5) and their corresponding rates for $\lambda_{\min}\{E(\Psi_n'\Psi_n/n)\}$ (see the text), it follows that either (a) $p_n = o(n^{d/3(d+2)})$ or (b) $p_n = o(n^{d/(3d+2)})$ suffices for Assumption B.1(b) by invoking Lemma A.1. Assumption B.1(a) also holds. Next, Assumption B.2 holds also either given $\hat{\sigma}_n^2 = n^{-1}\sum_t(Y_t - \hat{f}_{nt})^2$ (by Lemma A.2) or $\hat{\sigma}_n^2 = (n - p_n)^{-1}\sum_t(Y_t - \hat{\theta}_{nt})^2$ (as shown the proof of Theorem 3.1). Asymptotic normality follows from Theorem A.2(i). (ii) Since $\theta_o$ is square integrable with respect to $\mu$, it is also square integrable with respect to Lebesgue measure given Assumption A.1. It follows that Assumption B.1(c) holds under $H_A$. Consistency then follows from Theorem A.2(ii).
Q.E.D.

PROOF OF THEOREM 3.4: We have argued in the text that $\lambda_{\min}\{E(\Psi_n'\Psi_n/n)\} = O(p_n^{-1})$ and $\sup_{x \in \mathbb{X}_s}(N_j^m(x)) < c$ for all $j$. The proof is similar to Theorem 3.1.
Q.E.D.

In the following theorem, we show that our approach applies in the presence of heteroskedastic errors. We make the following assumption on the error term:

ASSUMPTION A.2': *Suppose that* $\varepsilon_t = \sigma(X_t)u_t$, *where* $P[0 < \inf_t\{\sigma(X_t)\} \leq \sup_t\{\sigma(X_t)\} < \infty] = 1$ *and* $\{u_t\}$ *is an i.i.d. sequence with* $E(u_t) = 0$, $E(u_t^2) = 1$, *and* $E(u_t^4) < \infty$. *Furthermore,* $\{u_t\}$ *is independent of* $\{X_t\}$.

THEOREM A.3: *Suppose that Assumptions A.1, A.2', A.3–A.4, and B.1 hold. Define* $M_n = (n\hat{m}_n - \hat{R}_n)/\hat{S}_n$ *and* $\bar{M}_n = (n\tilde{m}_n - \hat{R}_n)/\hat{S}_n$, *where* $\hat{m}_n$ *and* $\tilde{m}_n$ *are as in Theorem A.2,* $\hat{R}_n = \sum_t \varphi'_{nt}\varphi_{nt}\hat{\varepsilon}^2_{nt}$, $\hat{S}^2_n = 2\sum_t\sum_s(\varphi'_{nt}\varphi_{ns})^2\hat{\varepsilon}^2_{nt}\hat{\varepsilon}^2_{ns}$, $\varphi_{nt} = (\Psi'_n\Psi_n)^{-1/2}\psi_{nt}$ *and* $\hat{\varepsilon}_{nt} = Y_t - \hat{f}_{nt}$. *Let* $p_n \to \infty$ *as* $n \to \infty$. *Then (i) under* $H_o$, $M_n - \bar{M}_n = o_P(1)$, *and*

$$M_n \xrightarrow{d} N(0,1) \quad and \quad \bar{M}_n \xrightarrow{d} N(0,1);$$

*(ii) under* $H_A$ *and for any nonstochastic sequence* $\{C_n\}$, $C_n = o(n/p_n^{1/2})$,

$$P[M_n > C_n] \to 1 \quad and \quad P[\bar{M}_n > C_n] \to 1.$$

PROOF: We give a proof for $M_n$ only. (i) Following an analogous reasoning of the proof of Theorem A.2, we can obtain

$$(A.10) \qquad (n\hat{m}_n - R_n)/p_n^{1/2} = (W_n - R_n)/p_n^{1/2} + o_P(1) = U_n/p_n^{1/2} + o_P(1)$$

under $H_o$ given Assumptions A.1, A.2', A.3–A.4, and B.1, where $U_n = \sum\sum_{s<t}2\varphi'_{nt}\varphi_{nt}\varepsilon_t\varepsilon_s$ and $R_n = \sum_t \varphi'_{nt}\varphi_{nt}\varepsilon_t^2$. Put $R_n^* = \sum_t \varphi'_{nt}\varphi_{nt}\sigma_t^2$, where $\sigma_t^2 \equiv \sigma^2(X_t)$. Then $p_n^{-1/2}(R_n - R_n^*) = p_n^{-1/2}\sum_t \varphi'_{nt}\varphi_{nt}(\varepsilon_t^2 - \sigma_t^2) = o_P(1)$ by Chebyshev's inequality and $E\{\sum_t \varphi'_{nt}\varphi_{nt}(\varepsilon_t^2 - \sigma_t^2)\}^2 \leq c^{-1}E\{\sum_t(\varphi'_{nt}\varphi_{nt})^2\} = o(p_n)$. It follows from (A.10) that

$$(n\hat{m}_n - R_n^*)/S_n = U_n/S_n + o_P(1),$$

where $S_n^2 = 2\sum_t\sum_s(\varphi'_{nt}\varphi_{ns})^2\sigma_t^2\sigma_s^2$ and hence $cp_n \leq S_n^2 \leq c^{-1}p_n$ a.s. given Assumption A.2'. We now show $U_n/S_n \xrightarrow{d} N(0,1)$. For this, we first show that conditional on $X^n \equiv \{X_1,\ldots,X_n\}$, $U_n/S_n \xrightarrow{d} N(0,1)$ and then apply the Dominated Convergence Theorem to prove the unconditional normality. Since $E(U_{nts}|X^n;u_t) = E(U_{nts}|X^n;u_s) = 0$ for $t \neq s$, we apply de Jong's (1987) CLT to $U_n|X^n$. (de Jong's result applies to independent but not necessarily identical distributions.) First, we compute $\text{var}(U_n|X^n)$:

$$\text{var}(U_n|X^n) = \sum_{s<t}\sum E(U^2_{nts}|X^n) = 4\sum_{s<t}\sum(\varphi'_{nt}\varphi_{ns})^2\sigma_t^2\sigma_s^2 = S_n^2 - 2\sum_t(\varphi'_{nt}\varphi_{nt})^2\sigma_t^4$$

$$= S_n^2\{1 + o(1)\},$$

where $\sum_t(\varphi'_{nt}\varphi_{nt})^2\sigma_t^4 = o(p_n) = o(S_n^2)$. Following reasoning analogous to the proof of Theorem A.1, we have $G_I = o(p_n)$, $G_{II} = o(p_n)$, and $G_{IV} = O(p_n)$. It follows that $G_j/S_n^4 \to 0$, $j = $ I, II, and IV given $p_n \to \infty$. Therefore, we have $(U_n/S_n)|X^n \xrightarrow{d} N(0,1)$ by de Jong's (1987, Proposition 3.2). That is, the conditional probability $P[U_n/S_n \leq \xi|X^n]$ converges to the probability that a unit normal is less than $\xi$. Since the unconditional probability is the expectation of the conditional probability with respect to the distribution of $X^n$, the Dominated Convergence Theorem implies that $U_n/S_n \xrightarrow{d} N(0,1)$ unconditionally. It follows that $(n\hat{m}_n - R_n^*)/S_n \xrightarrow{d} N(0,1)$ under $H_o$.

To show $M_n \xrightarrow{d} N(0,1)$, it remains to show that $p_n^{-1/2}(\hat{R}_n - R_n) = o_P(1)$ and $p_n^{-1}(\hat{S}_n^2 - S_n^2) = o_P(1)$ (the latter implies that $p_n^{-1/2}(\hat{S}_n - S_n) = o_P(1)$). Noting that $\hat{\varepsilon}_{nt} = \varepsilon_t - (\hat{f}_{nt} - \theta_t^o)$, we have

$$p_n^{-1/2}(\hat{R}_n - R_n) = p_n^{-1/2}\sum_t \varphi'_{nt}\varphi_{nt}(\varepsilon_t^2 - \sigma_t^2) - p_n^{-1/2}\sum_t \varphi'_{nt}\varphi_{nt}\varepsilon_t(\hat{f}_{nt} - \theta_t^o)$$

$$+ p_n^{-1/2}\sum_t \varphi'_{nt}\varphi_{nt}(\hat{f}_{nt} - \theta_t^o)^2$$

$$= o_P(1),$$

where the first term is $o_P(1)$ by Chebyshev's inequality given Assumption B.1(b); for the third term, $p_n^{-1/2}\sum_t \varphi'_{nt}\varphi_{nt}(\hat{f}_{nt} - \theta_t^o)^2 \leq p_n^{-1/2}\sup_t\{\varphi'_{nt}\varphi_{nt}\}\sum_t(\hat{f}_{nt} - \theta_t^o)^2 = o_P(1)$ given B.1(b) and $\sum_t(\hat{f}_{nt} - \theta_t^o)^2$

$= O_P(1)$ under $H_o$; for the second term, $p_n^{-1/2}\sum_t \varphi'_{nt}\varphi_{nt}\varepsilon_t(\hat{f}_{nt} - \theta_t^o) = o_P(1)$ by the Cauchy-Schwarz inequality. Similarly, we can also show that

$$p_n^{-1}(\hat{S}_n^2 - S_n^2) = p_n^{-1}\sum_t\sum_s(\varphi'_{nt}\varphi_{ns})^2(\hat{\varepsilon}_{nt}^2\hat{\varepsilon}_{ns}^2 - \sigma_t^2\sigma_s^2) = o_P(1).$$

It follows that $M_n \xrightarrow{d} N(0,1)$ under $H_o$. (ii) Under $H_A$, we have $\hat{m}_n \xrightarrow{p} E\{\theta_o(X) - f_n(X, \alpha_n^*)\}^2$. In addition, it can be shown that $cp_n \leq \hat{S}_n^2 \leq c^{-1}p_n$ with probability approaching 1. The consistency then follows as in the proof of Theorem A.2(ii). Q.E.D.

## REFERENCES

ANDREWS, D. W. K. (1988): Laws of Large Numbers for Dependent Non-identically Distributed Random Variables," *Econometric Theory*, 4, 458–467.

——— (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59, 307–345.

BIERENS, H. J. (1982): "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134.

——— (1990): "A Consistent Conditional Moment Test of Functional Forms," *Econometrica*, 58, 1443–1458.

DE JONG, P. (1987): "A Central Limit Theorem for Generalized Quadratic Forms," *Probability Theory and Related Fields*, 75, 261–277.

DE JONG, R. M., AND H. J. BIERENS (1991): "On the Limit Behavior of a Chi-square Type Test if the Number of Conditional Moments Tested Approaches Infinity," Free University, Department of Econometrics Working Paper.

EDMUNDS, D. E., AND V. B. MOSCATELLI (1977): "Fourier Approximation and Embeddings of Sobolev Space," *Dissertationes Mathematicae*. Warsaw: Polish Scientific Publishers.

EUBANK, R., AND S. SPECKMAN (1990): "Curve Fitting by Polynomial-Trigonometric Regression," *Biometrika*, 77, 1–9.

EUBANK, R., AND C. SPIEGELMAN (1990): "Testing the Goodness of Fit of a Linear Model via Nonparametric Regression Techniques," *Journal of the American Statistical Association*, 85, 387–392.

GALLANT, A. R. (1981): "Unbiased Determination of Production Technologies," *Journal of Econometrics*, 20, 285–323.

GALLANT, A. R., AND D. JORGENSON (1979): "Statistical Inference for a System of Simultaneous, Nonlinear Implicit Equations in the Context of Instrumental Variables Estimation," *Journal of Econometrics*, 11, 275–302.

GALLANT, A. R., AND G. SOUZA (1991): "On the Asymptotic Normality of Fourier Flexible Form Estimates," *Journal of Econometrics*, 50, 329–353.

GALLANT, A. R., AND H. WHITE (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Oxford: Basil Blackwell.

GOZALO, P. (1993): "A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models," *Econometric Theory*, 9, 451–477.

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

HOLLY, A. (1982): "A Remark on Hausman's Specification Test," *Econometrica*, 50, 749–759.

HONG, Y., AND H. WHITE (1991): "Consistent Specification Testing Via Nonparametric Series Regression," University of California, San Diego, Department of Economics Discussion Paper.

——— (1995): "Consistent Nonparametric Entropy-Based Testing," University of California, San Diego, Department of Economics Discussion Paper.

JAYASURIYA, B. R. (1990): "Testing for Polynomial Regression Using Nonparametric Regression Techniques," Texas A&M University Department of Statistics Ph.D. Dissertation.

LEE, B.-J. (1988): "A Model Specification Test Against the Nonparametric Alternative," Dissertation, University of Wisconsin Department of Economics.

ROBINSON, P. M. (1991): "Consistent Nonparametric Entropy-based Testing," *Review of Economic Studies*, 58, 437–453.

SCHUMAKER, L. (1981): *Spline Functions: Basic Theory*. New York: John Wiley.

WHANG, YOON-JAE, AND D. W. K. ANDREWS (1993): "Tests of Specification for Parametric and Semiparametric Models," *Journal of Econometrics*, 57, 277–318.

WHITE, H., AND Y. HONG (1995): "*m*-Testing Using Finite and Infinite Dimensional Parameter Estimators," University of California, San Diego, Department of Economics Discussion Paper.

WHITE, H., AND M. STINCHCOMBE (1991): "Adaptive Efficient Weighted Least Squares with Dependent Observations," in *Directions in Robust Statistics and Diagnostics, Part II*, ed. by W. Stahel and S. Weisberg. New York: Springer Verlag, pp. 337–364.

WOOLDRIDGE, J. (1992): "A Test for Functional Form Against Nonparametric Alternatives," *Econometric Theory*, 8, 452–475.

YATCHEW, A. J. (1992): "Nonparametric Regression Tests Based on an Infinite Dimensional Least Squares Procedure," *Econometric Theory*, 8, 435–451.