UC San Diego
Recent Work

Title
M-Testing Using Finite and Infinite Dimensional Parameter Estimators

Permalink
https://escholarship.org/uc/item/9qz123ng

Authors
White, Halbert
Hong, Yongmiao

Publication Date
1999

eScholarship.org
Powered by the California Digital Library
University of California

93-01R

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

M-TESTING USING FINITE AND INFINITE DIMENSIONAL PARAMETER ESTIMATORS

BY

HALBERT WHITE

AND

YONGMIAO HONG

DISCUSSION PAPER 93-01R
JANUARY 1999

M-TESTING USING FINITE AND INFINITE DIMENSIONAL PARAMETER ESTIMATORS*

HALBERT WHITE
Project for Econometric Analysis
Department of Economics
University of CaliforniaΓSan Diego

and

YONGMIAO HONG
Department of Economics and Department of Statistical Science
Cornell University

* This is a revised version of our UCSD Department of Economics Discussion Paper 1993-1. We wish to thank Randy Eubank, Clive Granger, Kurt Honik, Sing-Chee Lee, Whitney Newey, Yosef Rinott, Max Stinchcombe, Jeff Wooldridge, and two anonymous referees for useful comments. All remaining errors are ours. White's participation was supported by NSF grants SES-8921382 and SES-9209023.

ABSTRACT

The m-testing approach provides a general and convenient framework in which to view and construct specification tests for econometric models. Previous m-testing frameworks only consider test statistics that involve finite dimensional parameter estimators and infinite dimensional parameter estimators affecting the limit distribution of the m-test statistics. In this paper we propose a new m-testing framework using both finite and infinite dimensional parameter estimatorsΓwhere the latter may or may not affect the limit distribution of the m-test. This greatly extends the potential and flexibility of m-testing. The new m-testing framework can be used to test hypotheses on parametricΓsemiparametric and nonparametric models. Some examples are given to illustrate how to use it to develop new specification tests.

Key Words: Consistent specification testΓInfinite dimensional parameterΓNonparametric estimationΓm-testing.

1. INTRODUCTION, MOTIVATION, AND HEURISTICS

A prominent theme recurring throughout Clive Granger's extensive body of work is his concern with the adequacy of econometric models. Granger [1990] gives this concern the status of an axiom in his general introduction to his volume *Modelling Economic Series*. In factΓit is his first axiom:

Axiom A: Any model will only be an approximation to the generating mechanismΓwith there being a preference for the best available approximation.

Granger's second axiom is:

Axiom B: The basic objective of a modelling exercise is to affect the beliefs — and hence the behavior — of other research workers.

Taken togetherΓ Axioms A and B require us not merely to report the results of our econometric modelingΓbut to evaluate the models as wellΓproviding "comparisons with other modelsΓthe results of specification testsΓout-of-sample evaluationΓand so forth" [GrangerΓ 1990Γp. 3].

IndeedΓ one of the most useful approaches to specification testing involves the direct comparison of the results of two different models of the same phenomenon. This approach to specification testingΓ pioneered by Durbin [1954]ΓWu [1973]Γand Hausman [1978]Γhas undergone substantial evolution and extension. Specification tests have progressed from purely parametric contextsΓas in Durbin [1954]ΓWu [1973]ΓHausman [1978]ΓNewey [1985]Γ and Tauchen [1985]Γto contexts involving both nonparametric and parametric approachesΓ as in Whang and Andrews [1993] and Hong and White [1991Γ1995]. Our purpose here is to extend and unify these approaches in a way that permits hypothesis testing about parametricΓ semi-parametricΓand nonparametric models in a manner not previously possibleΓproviding new tools to aid in achieving the objective of Granger's Axiom B.

To illustrate the issues involvedΓ consider the consistent test for the correctness of a parametric regression model $f(X_t, \alpha)$ for the conditional expectation $\theta_o(X_t) = E(Y_t|X_t)$ given by Hong and White [1991Γ1995]Γwhere Y_t is the dependent variableΓX_t is the vector of explanatory variables and α is a finite dimensional parameter vector. The test is based on the sample covariance

$$\hat{m}_n = n^{-1} \sum_{t=1}^{n} (\hat{\theta}_n(X_t) - f(X_t, \hat{\alpha}_n))(Y_t - f(X_t, \hat{\alpha}_n)). \tag{1.1}$$

Here $\hat{\alpha}_n$ is an appropriate estimatorΓsuch as the nonlinear least squares estimatorΓand $\hat{\theta}_n$ is a nonparametric series estimator for θ_o. The statistic $\hat{m}_n$ estimates the covariance

$$m_o = E[(\theta_o(X_t) - f(X_t, \alpha^*))(Y_t - f(X_t, \alpha^*))],$$

where $\alpha^* = p\lim \hat{\alpha}_n$. Under the null hypothesis of correct specification (and only then) we have $m_o = 0$, as correct specification implies $\theta_o(X_t) = f(X_t, \alpha_o)$ $a.s.$ for some $\alpha^* = \alpha_o$. ThusΓ a test based on $\hat{m}_n$ has asymptotic power one whenever $f(X_t, \alpha)$ is misspecified.

Although Hong and White [1995ΓTheorem A.3] give a version of their statistic that does not suffer from the effects of neglecting heteroskedasticity of unknown formΓthis immunity is achieved essentially by use of a heteroskedasticity-consistent covariance matrix estimator. An attractive alternative is to correct for heteroskedasticity of unknown form directlyΓusing a consistent nonparametric estimator for the conditional varianceΓas this may deliver better power. Letting $\pi_o(X_t) = (var(Y_t|X_t))^{1/2}$ and $\theta_o = \mu_o/\pi_o$, where $\mu_o(X_t) = E(Y_t|X_t)$, we now estimate θ_o nonparametrically by $\hat{\theta}_n$ and π_o nonparametricallyΓsay by $\hat{\pi}_n$ (e.g.Γas in Robinson [1987]). The statistic of interest is now

$$\hat{m}_n = n^{-1} \sum_{t=1}^{n} (\hat{\theta}_n(X_t) - f(X_t, \hat{\alpha}_n)/\hat{\pi}_n(X_t))(Y_t - f(X_t, \hat{\alpha}_n))/\hat{\pi}_n(X_t). \tag{1.2}$$

This estimates

$$m_o = E[(\mu_o(X_t) - f(X_t, \alpha^*))^2/\pi_o^2(X_t)],$$

which is again zero only under correct specification. We see that $\hat{m}_n$ is a particular value of

$$m_n(\alpha, \theta, \pi) = n^{-1} \sum_{t=1}^{n} m(Z_t, \alpha, \theta, \pi),$$

where $Z_t = (X_t', Y_t)'$ and $m(Z_t, \alpha, \theta, \pi) = (\theta(X_t) - f(X_t, \alpha)/\pi(X_t))(Y_t - f(X_t, \alpha))/\pi(X_t)$.

We distinguish between the two infinite dimensional parameters θ and π, as it turns out that the effects of replacing them with $\hat{\theta}_n$ and $\hat{\pi}_n$ are quite different: $\hat{\theta}_n$ plays a key role in determining the asymptotic distribution of a suitably scaled version of $\hat{m}_n$Γwhile $\hat{\pi}_n$ plays essentially no role in determining this distribution.

Although we are motivated by consideration of (1.1) and (1.2)Γit is conceptually simpler and notationally much simpler to work first with the general statistic $m_n(\hat{\alpha}_n, \hat{\theta}_n, \hat{\pi}_n)$ and then specialize. A not inconsiderable additional benefit to this is that many other interesting specification testing procedures fall into the same frameworkΓsaving a great deal of effort that might otherwise be required in treating them. We discuss several new applications in considerable detailΓincluding a new test for regression error normality in a nonparametric regression and a new test for omitted variables in nonparametric regressionΓas well as two new consistent regression specification tests based on (1.1) and (1.2).

To make clear the contribution of our approach and its relation to prior workΓwe recall that Newey [1985] and Tauchen [1985] treated the case of "m-testing" based on the statistic

$$m_n(\hat{\alpha}_n),$$

while White [1987Γ1994] treated the case

$$m_n(\hat{\alpha}_n, \hat{\pi}_n),$$

where $\hat{\alpha}_n$ is a parametric estimator that affects the asymptotic distribution of the test statisticΓand $\hat{\pi}_n$ is a finite dimensional parametric estimator not affecting the asymptotic distribution. Whang and Andrews [1993] achieved a substantial advance by letting $\hat{\pi}_n$ be a nonparametric estimator not affecting the asymptotic distribution. This framework can be used to test parametric and semiparametric models (see Whang and Andrews [1993]).

To handle the specific examples aboveΓwe introduce the nonparametric estimator $\hat{\theta}_n$, leading to

$$m_n(\hat{\alpha}_n, \hat{\theta}_n, \hat{\pi}_n).$$

This m-statistic cannot be handled within any previously studied framework (e.g.ΓWhang and Andrews [1993])Γbecause $\hat{\theta}_n$ plays a key role in determining its asymptotic distribution. Our goalΓthereforeΓis to develop appropriate theory to permit us to test hypotheses based on such statistics. A major consequence of introducing $\hat{\theta}_n$ is that the distribution and power theory for the tests of interest can differ substantially from that previously developed. IndeedΓthe joint presence of $\hat{\alpha}_n, \hat{\pi}_n$ and $\hat{\theta}_n$ introduces the potential for a variety of interesting possibilities. Our framework can be used to test hypotheses about parametricΓ semiparametric and nonparametric models.

Although the notation in the sections that follows is unavoidably complicated by the need to keep separate track of α, θ, and π, the basic underlying idea for developing our distribution theory is straightforward: essentiallyΓwe just take a Taylor series expansion appropriate to the situation at hand.

To see what is involvedΓwe first replace $\hat{\pi}_n$ with its limitΓsay π_o : at each step we will impose conditions ensuring that this has no effect on the asymptotic distribution of interest. Now take a first order Taylor expansion around α_o and θ_o :

$$m_n(\hat{\alpha}_n, \hat{\theta}_n, \pi_o) = m_n(\alpha_o, \theta_o, \pi_o) + \nabla'_\alpha m_n^o(\hat{\alpha}_n - \alpha_o) + \nabla'_\theta m_n^o(\hat{\theta}_n - \theta_o) + r_n.$$

In the second termΓ $\nabla'_\alpha m_n^o$ denotes the Jacobian of m_n with respect to α evaluated at $(\alpha_o, \theta_o, \pi_o)$. The third term is a flagrant abuse of notationΓbut it greatly helps us to see what is going on. If θ were finite dimensionalΓthen the Jacobian $\nabla'_\theta m_n^o$ of m_n with respect to θ at $(\alpha_o, \theta_o, \pi_o)$ could multiply $(\hat{\theta}_n - \theta_o)$ as we have written. Because θ is infinite dimensionalΓ what we have written is invalid; howeverΓby using the Frechet differentialΓwe get a term that behaves essentially just as $\nabla'_\theta m_n^o(\hat{\theta}_n - \theta_o)$ does. LaterΓthe Frechet notation δm_n^o appears in its place. For nowΓwe stick with our abuse. The final term (r_n) is a remainder.

To obtain the desired null distributionsΓwe need to find the orders of the different terms under the nullΓrescale by the rate for the slowest converging (i.e.Γdominant) term(s)Γand apply appropriate central limit results. ClearlyΓdifferent cases may arise in which the orders of the various terms bear different relationships to each other.

A particularly interesting possibility is that r_n dominatesΓi.e.Γthe first three terms in the first order expansion vanish under the null hypothesis at a rate faster than r_n vanishes. In particularΓthis occurs for our motivating case $m(Z_t, \alpha_o, \theta_o) = (\theta_o(X_t) - f(X_t, \alpha_o))(Y_t - f(X_t, \alpha_o))$ because $\theta_o(X_t) = f(X_t, \alpha_o)$ *a.s*, causing the first term to vanish for all n. The terms involving $\nabla'_\alpha m_n^o$ and $\nabla'_\theta m_n^o$ essentially vanish in probability at rates fast enough to overwhelm the more slowly converging $(\hat{\alpha}_n - \alpha_o)$ and $(\hat{\theta}_n - \theta_o)$. We refer to cases in which this does *not* happen as "first order"Γbecause the analysis can be based on the first order Taylor expansion. Cases in which we *do* have this sort of degeneracy will be called "second order"Γbecause it turns out that a second order Taylor expansion works.

The second order cases involve an approximation that acts like

$$m_n(\hat{\alpha}_n, \hat{\theta}_n, \pi_o) = \tfrac{1}{2}(\hat{\theta}_n - \theta_o)' \nabla_\theta^2 m_n^o (\hat{\theta}_n - \theta_o) + r_n.$$

Again we abuse notation. The term on the right involving $\nabla_\theta^2 m_n^o$ is really a second order Frechet derivative later denoted $\delta^2 m_n^o$. All but the dominant term have been placed in the remainder r_n. Analysis of the dominant term turns out to be straightforward using the distribution theory for U- or V-statistics. As might be expectedΓthe dominant term has non-zero expectation and so must be recentered properly; estimation of the requisite recentering is usually straightforward. InterestinglyΓthe rate of convergence of the leading term is typically quite rapid. In the pastΓthis has often been viewed as a form of degeneracyΓwith a variety of special measures introduced to avoid it. (See Section 2.3 below and Hong and White [1995] for a discussion.) We view this "degeneracy" as a potential advantage to be exploited: the rapid convergence rate leads to re-scalings that deliver statistics with better power under both local and global alternatives.

The preceding discussion suggests that for the first order case we will obtain conditions ensuring that $n^{1/2}\hat{m}_n$ converges in distribution to a normal random vector with mean zero under the nullΓas is usual; from this we can construct asymptotic χ^2 statistics in the usual way (i.e.Γby forming an appropriate quadratic form in $\hat{m}_n$). For the second order caseΓwe find thatΓafter recentering by R_n (say) and scaling by a_n (say)Γwhere a_n grows faster than $n^{1/2}$, $a_n(\hat{m}_n - R_n)$ converges to a normal random vector with mean zero under the null. AgainΓwe can construct asymptotic χ^2 statistics.

With this heuristic picture of what we are going to do and whyΓwe can now turn to a rigorous development of our theoryΓtreating first and second order cases separately.

2. THE BASIC FRAMEWORK

2.1 Fundamentals of M-testing

To beginΓwe describe the data generating process (DGP) and the estimators of interest.

Assumption A.1: $(\Omega, \mathcal{F}, P)$ is a complete probability space on which is defined the stochastic process $\{Z_{nt} : \Omega \to \mathbb{R}^\nu\}$, $t = 1, ..., n, n = 1, 2, ...$, $\nu \in \mathbb{N}$, where P is such that for each n $\{Z_{nt}\}$ is independently but not necessarily identically distributed (i.n.i.d.).

Assumption A.2: For pseudo-metric spaces (Θ, ρ_Θ) and (Π, ρ_Π) suppose $\hat{\theta}_n : \Omega \to \Theta$ and $\hat{\pi}_n : \Omega \to \Pi$, $n = 1, 2, ...$, are measurable such that $\rho_\Theta(\hat{\theta}_n, \theta_o) \to^p 0$ for $\theta_o \in \Theta$ and $\rho_\Pi(\hat{\pi}_n, \pi_o) \to^p 0$ for $\pi_o \in \Pi$. FurthermoreΓ$\hat{\alpha}_n : \Omega \to \mathbb{A} \subset \mathbb{R}^p$, $p \in \mathbb{N}$, is measurable with $\hat{\alpha}_n - \alpha_n^o \to^p 0$ for some nonstochastic sequence $\{\alpha_n^o \in \mathbb{A}\}$.

For notational simplicityΓbelow we let the dependence of Z_{nt} on n be implicit. Put $\Gamma = \mathbb{A} \times \Theta \times \Pi$. We consider a measurable "moment" function $m_{nt} : \mathbb{R}^\nu \times \Gamma \to \mathbb{R}^q$, $q \in \mathbb{N}$, that satisfies

$$E[m_{nt}(Z_t, \alpha_o, \theta_o, \pi_o)] = 0 \text{ for some } \alpha_o \in \mathbb{A} \text{ and all } t = 1, ..., n,\ n \geq 1$$

when the model is correctly specified. Under model misspecificationΓsuch a moment condition does not hold generallyΓgiving the test its power. The specific form taken by m_{nt} will be dictated by the null hypothesis of interest and the alternatives against which power is desired. Sections 3 and 4 provide a variety of examples illustrating choice of m_{nt}.

ThroughoutΓwe put $m_t(\gamma) = m_{nt}(Z_t, \gamma)$ and $\bar{m}_n(\gamma) = n^{-1}\sum_{t=1}^n Em_t(\gamma)$. Given the i.n.i.d. assumption and that $\hat{\theta}_n$ may affect the convergence rate of our statisticsΓwe define the null hypothesis based on $\{m_t\}$ to be

$$H_o : a_n \bar{m}_n(\alpha_o, \theta_o, \pi) \to 0 \text{ for some } (\alpha_o, \theta_o) \in \mathbb{A} \times \Theta \text{ and all } \pi \in \Pi_o \subseteq \Pi$$

for a nonstochastic sequence $\{a_n : a_n \to \infty, a_n/n \to 0\}$. Local alternatives are

$$H_{an} : a_n \bar{m}_n(\alpha_n^o, \theta_o, \pi) = O(1) \text{ for some } (\alpha_n^o, \theta_o) \in \mathbb{A} \times \Theta \text{ and all } \pi \in \Pi_o \subseteq \Pi.$$

We specify the global alternative as

$$H_A : \|\bar{m}_n(\gamma)\| \geq c > 0 \text{ for all } \gamma \in \Gamma \text{ and all } n \text{ sufficiently large.}$$

Note that H_{an} can be generated by the functional form of $\{m_t\}$ and/or sequence $\{\alpha_n^o\}$. The factor a_n is determined by $\{m_t\}$ and $\hat{\theta}_n$. In first order m-testingΓ$a_n = n^{1/2}$; for second orderΓ $a_n = n^{1/2+\epsilon}$ for some $\epsilon > 0$. These hypotheses may or may not coincide with the null hypothesis originally of interest (say H_o^*) and its alternatives. In first order m-testingΓ there is often a discrepancy between H_o and H_o^*. In second orderΓ howeverΓ H_o generally coincides with H_o^*, thus delivering consistent tests. These issues are addressed further in the applications of Sections 3 and 4.

Stochastic equicontinuity plays a key role in ensuring that $\hat{\pi}_n$ has no asymptotic effect.

Definition 2.1 [*Stochastic Equicontinuity*]: *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* (Π, ρ_Π) *be a pseudo-metric space. The stochastic process* $\{Q_n : \Omega \times \Pi \to \mathbb{R}^q\}$, $n = 1, 2, ...$, $q \in \mathbb{N}$, *is stochastically* ρ_Π*-equicontinuous at* $\pi_o \in \Pi$ *if for each* $\epsilon > 0$ *there exists* $\delta > 0$ *such that*

$$\lim_{n\to\infty} P^* \left[\sup_{\pi \in \mathbb{B}(\pi_o, \delta)} \|Q_n(\cdot, \pi) - Q_n(\cdot, \pi_o)\| > \epsilon\right] < \epsilon,$$

where P^* *is outer probability and* $\mathbb{B}(\pi_o, \delta) = \{\pi \in \Pi : \rho_\Pi(\pi, \pi_o) \leq \delta\}$.

Primitive conditions can be found in Andrews [1994]; Theorem 3.6 below also provides an alternative method to ensure stochastic equicontinuity.

Assumption A.3: (a) Given $(\alpha_n^o, \theta_o) \in \mathbb{A} \times \Theta$ and a nonstochastic sequence $\{a_n : a_n \to \infty, a_n/n \to 0\}$, $a_n(m_n(\alpha_n^o, \theta_o, \cdot) - \bar{m}_n(\alpha_n^o, \theta_o, \cdot))$ is stochastically ρ_Π-equicontinuous at $\pi_o \in \Pi$; and (b) $a_n \bar{m}_n(\alpha_n^o, \theta_o, \hat{\pi}_n) = a_n \bar{m}_n(\gamma_n^o) + o_P(1)$, where $\gamma_n^o = (\alpha_n^o, \theta_o, \pi_o)$.

We also make use of the concept of uniform equicontinuity (cf. Billingsley [1986]).

Definition 2.2 [*Uniform Equicontinuity*]: *Let* (Γ, ρ_Γ) *be a product pseudo-metric space. For each* n, *let* $Q_n : \Gamma \to \mathbb{R}^q \times \mathbb{R}^k,\ q, k \in \mathbb{N}$, *be a given mapping. Then* $\{Q_n\}$ *is uniformly equicontinuous on* Γ *with respect to* ρ_Γ *if for each* $\epsilon > 0$ *there exists* $\delta > 0$ *such that*

$$\limsup_n \sup_{(\gamma_1,\gamma_2)\in\mathbb{B}_n(\delta)} \|Q_n(\gamma_1) - Q_n(\gamma_2)\| < \epsilon,$$

where $\mathbb{B}_n(\delta) = \{(\gamma_1, \gamma_2) \in \Gamma \times \Gamma : \rho_\Gamma(\gamma_1, \gamma_2) \leq \delta\}$.

Assumption A.4: For each n denote $\hat{\gamma}_n = (\hat{\alpha}_n, \hat{\theta}_n, \hat{\pi}_n)$. Let ρ_Γ be a product pseudo-metric on Γ such that $\rho_\Gamma(\hat{\gamma}_n, \gamma_n^o) \to^p 0$. (a) For each $\gamma \in \Gamma$, $\bar{m}_n(\gamma)$ is $O(1)$ and is uniformly equicontinuous on Γ with respect to ρ_Γ; (b) $\{m_t(\gamma)\}$ obeys a weak uniform law of large numbers (ULLN) on ΓΓi.e.Γ$\sup_{\gamma\in\Gamma} \|m_n(\gamma) - \bar{m}_n(\gamma)\| \to^p 0$.

Weak ULLN's are given by Andrews [1991a]Γ Newey [1991]Γ and White and Wooldridge [1991].

2.2 First Order M-testing

We now treat the case in which the first order terms of a Taylor expansion determine the behavior of our test statistics. The next two assumptions permit a first order expansion.

Assumption B.1: (a) For each $(\theta, \pi) \in \Theta \times \Pi$, $m_t(\cdot, \theta, \pi)$ is continuously differentiable *a.s.* on $\mathbb{A}$ and $\bar{m}_n(\cdot, \theta, \pi)$ is continuously differentiable on $\mathbb{A}$; (b) for each $\gamma \in \Gamma$Γ $\nabla_\alpha \bar{m}_n(\gamma) = n^{-1}\sum_{t=1}^n E\nabla_\alpha m_t(\gamma)$ is $O(1)$ and is uniformly equicontinuous on Γ with respect to a product pseudo-metric ρ_Γ such that $\rho_\Gamma(\hat{\gamma}_n, \gamma_n^o) \to^p 0$; and (c) $\{\nabla_\alpha m_t(\gamma)\}$ obeys a weak ULLN on Γ.

Assumption B.2: (a) For each $\pi \in \Pi$, $m_t(\alpha_n^o, \cdot, \pi)$ is Frechet differentiable with respect to ρ_Θ *a.s.* on a neighborhood Θ_o of θ_o such that $E(\delta m_t(\theta - \theta_o; \gamma_n^o)) < \infty$ for all $\theta \in \Theta_o$; (b) there exist some $\lambda > 0$ and $D_{nt} : \mathbb{R}^\nu \to \mathbb{R}^+$, $n^{-1}\sum_{t=1}^n ED_{nt}(Z_t) = O(1)$, such that

$$\|m_t(\alpha_n^o, \theta, \pi) - m_t(\alpha_n^o, \theta_o, \pi) - \delta m_t(\theta - \theta_o; \alpha_n^o, \theta_o, \pi)\| \leq D_{nt}(Z_t)\rho_\Theta(\theta, \theta_o)^{1+\lambda} \ \ a.s.$$

for all $\theta \in \Theta_o$ and all $\pi \in \Pi$; (c) $\rho_\Theta(\hat{\theta}_n, \theta_o) = o_P(n^{-1/2(1+\lambda)})$.

The product pseudo-metric ρ_Γ in B.1 may differ from that of A.4. In B.2Γ$\delta m_t(\theta - \theta_o; \alpha_n^o, \theta_o, \pi)$ is the Frechet differential of $m_t(\alpha_n^o, \cdot, \pi)$ with respect to ρ_Θ at θ_o with increment $\theta - \theta_o$Γ corresponding to $\nabla'_\theta m_n^o(\hat{\theta}_n - \theta_o)$ in Section 1. The inequality in B.2(b) controls the Taylor series remainder. For $\{m_t\}$ linear in θ, set $\lambda = \infty$.

Assumption B.3: For each n denote $\delta m_n^o(\theta - \theta_o; \pi) = n^{-1}\sum_{t=1}^n \delta m_t(\theta - \theta_o; \alpha_n^o, \theta_o, \pi)$ and $\delta \bar{m}_n^o(\theta - \theta_o; \pi) = E\delta m_n^o(\theta - \theta_o; \pi)$. (a) $n^{1/2}\sup_{\pi \in \Pi} \|\delta m_n^o(\hat{\theta}_n - \theta_o; \pi) - \delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi)\| \to^p 0$; (b) $n^{1/2}(\delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) - \delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi_o)) \to^p 0$; and (c) $n^{1/2}\|\delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi_o) - V_n\| = o_P(1)$, where $V_n = n^{-1}\sum_{t=1}^n v_{nt}(Z_t, \gamma_n^o)$ and $v_{nt} : \mathbb{R}^\nu \times \Gamma \to \mathbb{R}^q$ is measurable with $n^{1/2}EV_n \to 0$.

For each $\pi \in \Pi$Γ$\delta m_n^o(\hat{\theta}_n - \theta_o; \pi)$ is asymptotically a second order V-statisticΓso B.3(a) is a uniform V-statistic projection. B.3(b) ensures that replacing $\hat{\pi}_n$ with π_o does not affect the limiting distribution of $\delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n)$. Ensuring B.3(c) typically involves an "under-smoothing" procedure to make the bias of $\hat{\theta}_n$ vanish faster than its variance.

Assumption B.4: $n^{1/2}(\hat{\alpha}_n - \alpha_n^o) = n^{1/2}S_n + o_P(1)$, where $S_n = n^{-1}\sum_{t=1}^n s_{nt}(Z_t, \alpha_n^o)$, $s_{nt} : \mathbb{R}^\nu \times \mathbb{A} \to \mathbb{R}^p$ is measurableΓand $n^{1/2}ES_n \to 0$.

This includes most parametric and semiparametric estimators that are $n^{1/2}$-consistent and asymptotically normal. In parametric maximum likelihood estimationΓfor exampleΓs_{nt} is the score function premultiplied by the inverse of the information matrix.

Assumption B.5: $J_n^{o-1/2} n^{1/2} W_n \to^d N(0, I_q)$, where $W_n = m_n(\gamma_n^o) - \bar{m}_n(\gamma_n^o) + \nabla'_\alpha \bar{m}_n(\gamma_n^o)S_n + V_n$ and J_n^o is a $q \times q$ nonstochastic $O(1)$ uniformly positive definite matrix.

This ensures that $n^{1/2}W_n$ is nondegenerate. It occurs when $\delta \bar{m}_n^o(\theta - \theta_o; \pi_o) = O_P(n^{-1/2})$, so that this functional of $\hat{\theta}_n$ achieves the parametric rate. HereΓ$n^{1/2}V_n$, $n^{1/2}(m_n(\gamma_n^o) - \bar{m}_n(\gamma_n^o))$ and $n^{1/2}\nabla'_\alpha \bar{m}_n(\gamma_n^o)S_n$ jointly determine the limiting distribution of $n^{1/2}\hat{m}_n$. This possibility arises when $\delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi_o)$ can be approximated asymptotically as a weighted integral of $\hat{\theta}_n - \theta_o$, providing additional smoothing. See Andrews [1991bΓSection 4]ΓGoldstein and Messer [1992]ΓHärdle and Stoker [1989]ΓLavergne and Vuong [1996]ΓNewey [1994]ΓPowellΓStock and Stoker [1989]ΓRobinson [1988] and Stoker [1989].

Assumption B.6: For each n there exists a measurable $\hat{J}_n : \Omega \to \mathbb{R}^q \times \mathbb{R}^q$ such that $\hat{J}_n - J_n \to^p 0$, where J_n is a $q \times q$ nonstochastic $O(1)$ uniformly positive definite matrix with $J_n = J_n^o$ under H_{an}, where J_n^o is as in B.5.

We now state the first main resultΓa substantive extension of Whang and Andrews [1993].

Theorem 2.3: *Define* $M_n = n\hat{m}_n' \hat{J}_n^- \hat{m}_n$*, where* $\hat{m}_n = m_n(\hat{\gamma}_n)$*. (i) Suppose A.1-A.3 (with* $a_n = n^{1/2}$*) and B.1-B.6 hold. Then under* H_{an} *with* $a_n = n^{1/2}$,

$$M_n \to^d \chi_q^2(\zeta_n^o),$$

where $\chi_q^2(\zeta_n^o)$ *is a chi-square distribution with* q *degrees of freedom and noncentrality* $\zeta_n^o = n\bar{m}_n'(\gamma_n^o) J_n^{o-1} \bar{m}_n(\gamma_n^o)$*; (ii) Suppose A.1,A.2,A.4 and B.6 hold. Then under* H_A *and for any nonstochastic sequence* $\{C_n = o(n)\}$,

$$P[M_n > C_n] \to 1.$$

When the limiting random variable depends on n as in (i) aboveΓthe convergence in distribution is as defined by White [1994ΓDefinition 8.3]. Theorem 2.3 implies that M_n is able to detect the class of local alternatives converging to the null at the parametric rate $n^{-1/2}$. Compared to Whang and Andrews [1993]Γwho consider only the infinite dimensional parameter estimators that do not affect the limit distribution of the m-test statisticΓwe permit use of infinite dimensional parameter estimators that may or may not affect the limit distribution of interest. This extends the scope of m-testing to test parametricΓsemiparametricΓand nonparametric models against various alternativesΓas illustrated by the examples in Sections 3 and 4 below.

2.3 Second Order M-testing

We now consider the case in which second order terms dominate in our Taylor approximation. To characterize the relevant casesΓwe use the following definition.

Definition 2.4 [*Degenerate Moment Function*]*: Let A.1, B.1(a) and B.2(a) hold. Then* $\{m_t\}$ *is* a_n*-degenerate at* $\gamma_n^o = (\alpha_n^o, \theta_o, \pi_o) \in \Gamma$ *if there exists a nonstochastic sequence* $\{a_n : a_n/n^{1/2} \to \infty, a_n/n \to 0\}$ *such that (a)* $a_n(m_n(\gamma_n^o) - \bar{m}_n(\gamma_n^o)) \to^p 0$*; (b)* $(a_n/n^{1/2})\nabla_\alpha \bar{m}_n(\gamma_n^o) \to 0$*; and (c)* $a_n \delta \bar{m}_n^o(\theta - \theta_o; \pi_o) \to 0$ *for all* $\theta \in \Theta_o \subseteq \Theta$*, where* Θ_o *contains a neighborhood of* θ_o.

Under H_o, $m_t(\gamma_o) = 0$ *a.s*, $\nabla_\alpha \bar{m}_n(\gamma_o) = 0$ and $\delta \bar{m}_n^o(\theta - \theta_o; \pi_o) = 0$ for all $\theta \in \Theta_o$ and all t, n. Assumption B.5 thus fails. ConsequentlyΓTheorem 2.3 does not apply to DMF's.

The examples at the outset of Section 1 are DMF's. Hong and White [1995] give numerous other examples relevant for testing specification hypotheses about models of conditional densities or expectations. In the pastΓthe standard response to degeneracy has been to remove itΓe.g.Γby sample splitting (Yatchew [1992]ΓWhang and Andrews [1993ΓSection 5])Γuse of nonparametric estimators not nesting the parametric model (Wooldridge [1992])Γor special weightings (Fan and Gencay [1993]ΓLee [1988] and Robinson [1991]). These approaches base the limiting distribution of the test statistics essentially on modified first order terms. As it turns outΓthese approaches do not fully exploit the possible efficiency gains provided by the degeneracy. In additionΓeach has features one may consider drawbacks: sample-splitting uses relatively inefficient nonparametric estimators; non-nested approaches require slow convergence of the nonparametric estimator to the true function; and weighting may introduce unnecessary noise or make the asymptotic covariance matrix depend on a nuisance parameterΓthe choice of which may affect size and power in finite samples. FurtherΓthese procedures may work only in certain cases. For exampleΓnon-nested testing and deterministic weighting may not apply when θ_o is constant under the nullΓas in testing heteroskedasticity.

We therefore part with tradition and avoid these drawbacks by basing tests on the dominant second order terms. As our statistics are quadratic formsΓCLT's for generalized quadratic forms (e.g.Γde Jong [1987]) or degenerate U-statistics (e.g.ΓHall [1984]) apply. In addition to being straightforwardΓa main advantage of our approach is that it improves asymptotic power under both local and global alternativesΓas will be seen below.

We now introduce two conditions that permit a two term Taylor expansion.

Assumption C.1: (a) For each $(\theta, \pi) \in \Theta \times \Pi$, $m_t(\cdot, \theta, \pi)$ is twice continuously differentiable $a.s.$ on $\mathbb{A}$, with $\|\nabla'_\alpha m_t(\cdot)\|$ and $\|\nabla^2_\alpha m_t(\cdot)\|$ dominated by $D_{nt} : \mathbb{R}^\nu \to \mathbb{R}^+$, $n^{-1}\sum_{t=1}^n ED_{nt}(Z_t) = O(1)$; (b) with $\{a_n\}$ as in Assumption A.3Γ$a_n n^{-1/2}(\nabla'_\alpha m_n(\alpha^o_n, \hat{\theta}_n, \hat{\pi}_n) - \nabla'_\alpha \bar{m}_n(\alpha^o_n, \hat{\theta}_n, \hat{\pi}_n)) \to^p 0$; and (c) $a_n n^{-1/2}(\nabla'_\alpha \bar{m}_n(\alpha^o_n, \hat{\theta}_n, \hat{\pi}_n) - \nabla'_\alpha \bar{m}_n(\gamma^o_n)) \to^p 0$.

This ensures that the first two terms in the Taylor expansion of $a_n \hat{m}_n$ around α^o_n do not affect its limit distribution.

Assumption C.2: (a) For each $\pi \in \Pi$, $m_t(\alpha^o_n, \cdot, \pi)$ is twice Frechet differentiable with respect to ρ_Θ $a.s.$ on a neighborhood Θ_o of θ_o and there exist some $\lambda > 0$ and $D_{nt} : \mathbb{R}^\nu \to$

$\mathbb{R}^+$, $n^{-1}\sum_{t=1}^{n} ED_{nt}(Z_t) = O(1)$, such that for all $\theta \in \Theta_o$ and $\pi \in \Pi$,

$$\|m_t(\alpha_n^o, \theta, \pi) - m_t(\alpha_n^o, \theta_o, \pi) - \delta m_t(\theta - \theta_o; \alpha_n^o, \theta_o, \pi) - \delta^2 m_t(\theta - \theta_o; \alpha_n^o, \theta_o, \pi)\| \\ \leq D_{nt}(Z_{nt})\rho_\Theta(\theta; \theta_o)^{2+\lambda} \ a.s.;$$

(b) with $\{a_n\}$ as in Assumption A.3 and λ as in (a)Γ$\rho_\Theta(\hat{\theta}_n, \theta_o) = o_P(a_n^{-1/(2+\lambda)})$.

HereΓ$\delta^2 m_t(\theta - \theta_o; \alpha_n^o, \theta_o, \pi)$ corresponds to $\frac{1}{2}(\theta - \theta_o)'\nabla_\theta^2 m_n^o(\theta - \theta_o)$ in Section 1. The inequality imposes a rate condition on the remainder term of the Taylor expansion.

Assumption C.3: For $(\theta, \pi) \in \Theta_o \times \Pi$ denote $\delta^2 m_n^o(\theta - \theta_o; \pi) = n^{-1}\sum_{t=1}^{n} \delta^2 m_t(\theta - \theta_o; \alpha_n^o, \theta_o, \pi)$ and $\delta^2 \bar{m}_n^o(\theta - \theta_o; \pi) = E\delta^2 m_n^o(\theta - \theta_o; \pi)$. Let $\{a_n\}$ be as in Assumption A.3. (a) $a_n \sup_{\pi\in\Pi}\|\delta^2 m_n^o(\hat{\theta}_n - \theta_o; \pi) - \delta^2 \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi)\| \to^p 0$; (b) $\delta m_n^o(\hat{\theta}_n - \theta_o; \pi) - \delta \bar{m}_n^o(\theta - \theta_o; \pi) + \delta^2 \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi) = W_n(\pi) + o_P(a_n^{-1})$ uniformly in $\pi \in \Pi$, where $W_n(\pi) = n^{-2}\sum_{t=1}^{n}\sum_{s=1}^{n} W_{nts}(Z_t, Z_s; \pi)$ and $W_{nts} : \mathbb{R}^\nu \times \mathbb{R}^\nu \times \Pi \to \mathbb{R}^q$ is measurable; (c) $a_n(W_n(\hat{\pi}_n) - EW_n(\hat{\pi}_n)) = a_n(W_n(\pi_o) - EW_n(\pi_o)) + o_P(1)$; (d) $a_n EW_n(\hat{\pi}_n) = a_n EW_n(\pi_o) + o_P(1)$ and $a_n \delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) = a_n \delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi_o) + o_P(1)$.

For each $\pi \in \Pi$, $\delta^2 m_n^o(\hat{\theta}_n - \theta_o; \pi)$ is asymptotically a third order V-statisticΓso C.3(a) is a uniform V-statistic projection. C.3(b) says that $\delta m_n^o(\hat{\theta}_n - \theta_o; \pi) - \delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi) + \delta^2 \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi)$ is asymptotically a generalized quadratic form $W_n(\pi)$ (see de Jong [1987]). When $W_{nts}(Z_t, Z_s; \pi) = W_n(Z_t, Z_s; \pi)$, $W_n(\pi)$ is a second order V-statistic. In first order m-testingΓthe term $\delta m_n^o(\hat{\theta}_n - \theta_o; \pi) - \delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi)$ vanishes (cf. B.3(a))Γbut here it matters. GenericallyΓit is of the same order as the second term in C.3(b). In specific casesΓthe limiting distribution of the m-test statistic may be determined by: (i) the first term only (when $\{m_t\}$ is linear in θ, as in (1.1) and (1.2)); (ii) the second term only; or (iii) both jointly (e.g.ΓHong and White [1993Γ1995]). C.3(b) provides a useful decomposition of DMF's; there may exist alternative decompositionsΓleading to different tests. For exampleΓHong and White [1991] show that different decompositions for (1.1) lead to a nested test (Hong and White [1991]) and a non-nested test (Wooldridge [1992]). C.3(c) ensures that replacing $\hat{\pi}_n$ with π_o does not matter asymptotically. While $a_n EW_n(\pi_o)$ dominates $a_n(W_n(\pi_o) - EW_n(\pi_o))$Γit can be subtracted from $a_n \hat{m}_n$ so that $a_n(W_n(\pi_o) - EW_n(\pi_o))$ becomes dominantΓa "recentering" procedure. This can have an appealing interpretation. For exampleΓwith $\hat{\theta}_n$ a nonparametric series estimatorΓHong and White [1995] interpret recentering as subtracting the degrees of freedom from a χ^2 random variable.

Assumption C.4: $n^{1/2}(\hat{\alpha}_n - \alpha_n^o) = O_P(1)$.

We need not know the structure of $\hat{\alpha}_n$, as it will not affect the limit distribution of the test.

Assumption C.5: For $\{a_n\}$ as in Assumption A.3Γ$J_n^{o-1/2} a_n(W_n(\pi_o) - EW_n(\pi_o)) \to^d N(0,1)$ as $a_n \to \infty$, where J_n^o is a $q \times q$ nonstochastic $O(1)$ uniformly positive definite matrix.

Given the i.n.i.d. assumptionΓwe generally have $var(W_{nts}(Z_t, Z_s)|Z_t) = var(W_{nts}(Z_t, Z_s)|Z_s) = 0$ for $t \neq s$. ThusΓ$W_n(\pi_o) - EW_n(\pi_o)$ is a degenerate U-statistic. HereΓCLT's for non-degenerate U-statistics (e.g.ΓPower, Stock and Stoker [1989]ΓLavergne and Vuong [1996]) do not apply. InsteadΓwe must use CLT's for degenerate generalized quadratic forms (or degenerate U-statistics). For CLT's for quadratic formsΓsee (e.g.) de Jong [1987]ΓHall [1984]ΓMikosch [1991]ΓRotar [1973] and Whittle [1964].

Assumption C.6: (a) $\hat{J}_n : \Omega \to \mathbb{R}^q \times \mathbb{R}^q$ is measurable such that $\hat{J}_n - J_n \to^p 0$, where J_n is a $q \times q$ nonstochastic $O(1)$ uniformly positive definite matrix with $J_n = J_n^o$ under H_{an}, where J_n^o is as in C.5; (b) $\hat{R}_n : \Omega \to \mathbb{R}^q$ is measurable such that $a_n(\hat{R}_n - R_n) \to^p 0$, where $\{a_n\}$ is as in A.3 and R_n is a $q \times 1$ nonstochastic vector with $R_n = EW_n(\pi_o)$ under H_{an} and $\|R_n\| = o(\|\bar{m}_n(\gamma_n^o)\|)$ under H_A.

Our second main result can now be given.

Theorem 2.5: *Suppose $\{m_t\}$ satisfies Definition 2.4 with $\{a_n\}$ as in A.3. Define $M_n = a_n^2(\hat{m}_n - \hat{R}_n)' \hat{J}_n^- (\hat{m}_n - \hat{R}_n)$, where $\hat{m}_n = m_n(\hat{\gamma}_n)$. (i) Suppose A.1-A.3 and C.1-C.6 hold. Then under H_{an},*

$$M_n \to^d \chi_q^2(\zeta_n^o),$$

where $\zeta_n^o = a_n^2 \bar{m}_n'(\gamma_n^o) J_n^{o-1} \bar{m}_n(\gamma_n^o)$. (ii) Suppose A.1, A.2, A.4 and C.6 hold. Then under H_A and for any nonstochastic sequence $\{C_n = o(a_n^2)\}$,

$$P[M_n > C_n] \to 1.$$

To interpret H_{an} for the DMF'sΓwe consider the case in which H_{an} is generated by the sequence $\{\alpha_n^o\}$, where $\alpha_n^o \to \alpha_o$, and α_o is as in H_o. Recall that for DMF'sΓ$\bar{m}_n(\gamma_o) = 0$ and $\nabla_\alpha \bar{m}_n(\gamma_o) = 0$Γso a two term Taylor expansion gives

$$a_n \bar{m}_n(\gamma_n^o) = \tfrac{1}{2} a_n(\alpha_n^o - \alpha_o)' \nabla_\alpha^2 \bar{m}_n(\gamma_o)(\alpha_n^o - \alpha_o) + o(a_n ||\alpha_n^o - \alpha_o||^2).$$

It follows that M_n has nontrivial power against $H_{an}^* : \alpha_n^o - \alpha_o = ca_n^{-1/2}$ for some $c \neq 0$. ObviouslyΓthis local alternative converges to H_o faster than $n^{-1/4}$ because $a_n/n^{1/2} \to \infty$. We thus achieve an efficiency improvement in terms of local powerΓcompared to the various previous approaches that avoid rather than exploit the degeneracy. For these approachesΓ typical local alternatives are $n^{-1/4}$ (see Hong and White [1993] for an example).

Theorem 2.5(i) shows that M_n cannot detect local alternatives H_{an}^* of $O(n^{-1/2})$, because $a_n/n \to 0$. In other wordsΓthe second order tests are less efficient than those that are able to detect local alternatives vanishing at the parametric rate $n^{-1/2}$. HoweverΓthis conclusion is specific to the local power criterion. Using other appropriate efficiency criteriaΓthe conclusion can be different for second order m-tests. SpecificallyΓwe can apply Bahadur's [1960] asymptotic slope criterionΓsuitable for comparing two large sample tests under fixed alternatives. The basic idea is to hold power fixed and compare the resulting test sizes. Bahadur's relative efficiency is the limit of the ratio of the sample sizes required by two tests to achieve the same asymptotic significance level (p-value) under a fixed alternative. This criterion has been used by (e.g.) Geweke [1981aΓ1981b] among others.

For parametric testingΓthe asymptotic slope is the rate at which minus twice the logarithm of the asymptotic significance level of the test statistic tends to infinity as n increases. Because the rate of divergence of second order m-tests is different from that of the parametric testsΓwe cannot use Bahadur's approach directly. InsteadΓwe extend it appropriately. Given $M_n \to^d \chi_q^2$ under H_o, the asymptotic significance level of M_n is $1 - F_q(M_n)$, where F_q is the cdf of χ_q^2. We now define

$$K_n = -2\ln(1 - F_q(M_n)).$$

Because $\ln(1 - F_q(\zeta)) = -\frac{1}{2}\zeta^2(1 + o(1))$ as $\zeta \to +\infty$ (cf. Bahadur [1960ΓSection 5])Γit follows from Theorem 2.5(ii) that

$$K_n/a_n^2 = \bar{m}_n'(\gamma_n^o)J_n^{-1}\bar{m}_n(\gamma_n^o) + o_P(1).$$

Following BahadurΓwe call $\bar{m}_n'(\gamma_n^o)J_n^{-1}\bar{m}_n(\gamma_n^o)$ the "asymptotic slope" of the sequence of tests based on $\{M_n\}$ under H_A. ObviouslyΓa larger asymptotic slope or a faster rate a_n implies a faster rate at which the asymptotic significance level decreases to zero as $n \to \infty$. For parametric tests and first order m-testsΓ$a_n = n^{1/2}$. For second order m-testsΓhoweverΓ $a_n/n^{1/2} \to \infty$. For exampleΓHong and White [1995] have $a_n = n^{1/2+\epsilon}$ for $\epsilon > 0$. ThereforeΓ second order m-tests are more efficient than parametric tests or first order m-tests under fixed alternatives in the sense that Bahadur's relative efficiency is infinite. This conclusion is in sharp contrast to that reached under H_{an}.

3. APPLICATION TO NONPARAMETRIC SERIES ESTIMATION

3.1a First Order M-testing: Results with Fixed Regressors

We first apply Theorem 2.3 to robust nonparametric series regressions. For simplicity and convenienceΓwe assume the following DGP.

Assumption D.1: (a) For each n, $Y_t = \mu_o(X_t) + \sigma_o(X_t)\epsilon_t$, $t = 1, ..., n$, where $\mu_o \in \mathcal{C}^r(\mathbb{X})$ and $\mathbb{X} \subset \mathbb{R}^d$ contains the support of X_t, r, $d \in \mathbb{N}$. Suppose $\{X_t\}$ are nonrandom and $\{\epsilon_t\}$ are independently and identically distributed (i.i.d.) with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = 1$; $(b)\, \sigma_o : \mathbb{X} \to \mathbb{R}^+$ is constant.

Here $\{X_t\}$ (hence $\{Y_t\}$) may implicitly depend on n. The analysis extends to random regressors (see Andrews [1991b] or Gallant Souza [1991]). Homoskedasticity (D.1(b)) can be relaxed; we do so in Section 3.2.

A nonparametric series estimator for μ_o is $\hat{\mu}_n = \arg\min_{\mu \in \Theta_n} Q_n(Z^n, \mu)$, where $Q_n : \Omega \times \Theta_n \to \mathbb{R}$, $Z^n = (Z_1, ..., Z_n)$, $Z_t = (Y_t, X_t')'$, and

$$\Theta_n = \{\theta : \mathbb{X} \to \mathbb{R} \quad | \quad \theta(x) = \textstyle\sum_{j=1}^{p_n} \beta_j \psi_j(x),\ \beta_j \in \mathbb{R}\} \tag{3.1}$$

for given $\{\psi_j : \mathbb{X} \to \mathbb{R}\}$ and $p_n \in \mathbb{N}$. We take $Q_n(Z^n, \mu) = n^{-1} \sum_{t=1}^n \phi(Y_t - \mu(X_t))$, with $\phi : \mathbb{R} \to \mathbb{R}$. If ϕ is convex with derivative φ, then $\hat{\mu}_n(\cdot) = \psi_{np}'\hat{\beta}_n$ solves

$$n^{-1} \textstyle\sum_{t=1}^n \psi_{np}(X_t)\varphi(Y_t - \hat{\mu}_n(X_t)) = 0, \tag{3.2}$$

where $\psi_{np_n}(X_t) = \{\psi_1(X_t), ..., \psi_{p_n}(X_t)\}'$ and $\hat{\beta}_n$ are $p_n \times 1$ vectors.

Andrews [1991b] treats least squares. We complement Andrews by giving new results for robust estimators. We follow Yohai and Maronna [1979]ΓMammen [1989] by restricting φ.

Assumption D.2: $\varphi : \mathbb{R} \to \mathbb{R}$ is a monotonic bounded function with three bounded derivatives such that $E\varphi(\epsilon_t) = 0$ and $E\varphi'(\epsilon_t) > 0$.

Monotonicity ensures a unique solution for $\hat{\beta}_n$. Boundedness ensures robustness to outliers; it rules out least squares. Differentiability is for convenience.

Denote $D^\lambda\theta = (\partial^{\lambda_1}/\partial x_1^{\lambda_1}) \cdots (\partial^{\lambda_d}/\partial x_d^{\lambda_d})\theta$, where $\theta \in \mathcal{C}^r(\mathbb{X})$ and $\lambda = (\lambda_1, ..., \lambda_d)'$ is a $d \times 1$ vector of nonnegative integers. The order of $D^\lambda\theta$ is $|\lambda| = \sum_{i=1}^d \lambda_i \leq r$. When $|\lambda| = 0$, put $D^0\theta = \theta$. We use certain Sobolev spaces.

Definition 3.1 [*Sobolev Space*]: *Let* $\theta \in \mathcal{C}^r(\mathbb{X})$, $r \in \mathbb{N}$. *For* $0 \leq s \leq r$, *define* $\rho_{s,\infty}(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|_{s,\infty}$, *where* $\|\theta\|_{s,\infty} = \max_{|\lambda| \leq s} \sup_{x \in \mathbb{X}} \left|D^\lambda \theta(x)\right|$. *When* $s = 0$, *write* $\rho_\infty = \rho_{0,\infty}$. *Define the Sobolev spaces*

$$\mathcal{W}^s_{\infty,r}(\mathbb{X}) = \left\{\theta \in \mathcal{C}^r(\mathbb{X}) : \|\theta\|_{s,\infty} < \infty\right\},\ 0 \leq s \leq r,\ \mathcal{W}_{\infty,r}(\mathbb{X}) = \mathcal{W}^0_{\infty,r}(\mathbb{X}).$$

Assumption D.3: Let $\lambda_{\min}(A)$ denote the minimum eigenvalue of square matrix A and $\Psi_n = \{\psi'_{np_n}(X_1), ..., \psi'_{np_n}(X_n)\}'$. Suppose (a) $\lambda_{\min}(\Psi'_n\Psi_n) \to \infty$; (b) putting $\xi_{np_n} = \sup_{1 \leq t \leq n}$ $\{\psi'_{np_n}(X_t)(\Psi'_n\Psi_n)^{-}\psi_{np_n}(X_t)\}$, then $\max(n^{1/3}[\ln(n)]^{2/3}, p_n\xi_{np_n}) \to 0$; (c) $\max_{0 \leq j \leq p_n} \|\psi_j\|_{s,\infty} \leq B_s(p_n)$, $0 \leq s \leq r$, where $B_s : \mathbb{N} \to \mathbb{R}^+$ is nondecreasing.

Assumption D.3(a) is key for consistency of $\hat{\mu}_n$ for μ_o. D.3(b) is a strengthened Lindeberg conditionΓimplying $p^2/n \to 0$. We allow $B_0(p_n)$ to increase with p_n. For such series as B-splinesΓGallant's [1981] Fourier Flexible Form (FFF) and the trigonometric seriesΓ$B_0(\cdot)$ is bounded if $\mathbb{X}$ is. For $1 \leq s \leq r$, $B_s(p_n)$ grows with p_n generally.

Assumption D.4: There exists a sequence of $p_n \times 1$ nonstochastic vectors $\{\beta^o_n\}$ such that $\mu^o_n(\cdot) = \psi'_{np_n}(\cdot)\beta^o_n \in \mathcal{W}^s_{\infty,r}(\mathbb{X})$ and (a) $\rho_\infty(\mu^o_n, \mu_o) = o(n^{-1/2})$; or (b) $\rho_{s,\infty}(\mu^o_n, \mu_o) = o(n^{-1/2(1+\lambda)})$, $0 \leq s \leq r$, for λ as in B.2(a).

Given $\{\psi_j\}$, D.4 is ensured by imposing smoothness on μ_o with appropriate choice of p_n. For exampleΓif $\{\psi_j\}$ is the Fourier seriesΓ$\mu_o \in \mathcal{C}^r(\mathbb{X})$, $\mathbb{X} = (0, 2\pi)^d$, then there exists $\mu^o_n(\cdot) = \psi'_{np_n}(\cdot)\beta^o_n \in \mathcal{W}^s_{\infty,r}(\mathbb{X})$, $0 \leq s < r$, such that $\rho_{s,\infty}(\mu^o_n, \mu_o) = o(p_n^{-(r-s)/d+\epsilon})$ for any $\epsilon > 0$ (e.g. Edmunds and Moscalelli [1977]). HenceΓD.4(a) holds if $p_n = n^{d/2r+\epsilon}$ and D.4(b) holds if $p_n = n^{d/2(r-s)(1+\lambda)+\epsilon}$.

We now establish consistency and asymptotic normality for $\hat{\mu}_n$.

Proposition 3.2 [*Consistency*]: *Suppose D.1-D.2, D.3(a) and D.4(a) hold, and* $p_n \to \infty$, $p_n\xi_{np_n} \to 0$. *Let* $\hat{\mu}_n$ *be as in (3.2). Then* $n^{-1}\sum_{t=1}^n(\hat{\mu}_n(X_t) - \mu_o(X_t))^2 = O_P(p_n/n)$ *and* $||\hat{\beta}_n - \beta^o_n|| = O_P(p_n^{1/2}/\lambda^{1/2}_{\min}(\Psi'_n\Psi_n))$.

Proposition 3.3 [*Normality*]: *Suppose D.1-D.3(a,b) and D.4(a) hold. Let* G_{np_n} *be a sequence of* $p_n \times q$ *nonstochastic matrices such that* $I^o_n = G'_{np_n}(\Psi'_n\Psi_n)^{-}G_{np_n}E\varphi^2(\epsilon_t)/E^2\varphi'(\epsilon_t)$ *is a* $q \times q$ *uniformly nonsingular matrix. Let* $v_{nt}(Z_t) = G'_{nt}(\Psi'_n\Psi_n)^{-}\psi_{np_n}(X_t)\varphi(\epsilon_t)/E\varphi'(\epsilon_t)$. *Then*

$I_n^{o-1/2} G'_{np_n}(\hat{\beta}_n - \beta_n^o) = I_n^{o-1/2} \sum_{t=1}^n v_{nt} + o_P(1)$, *and*

$$I_n^{o-1/2} G'_{np_n}(\hat{\beta}_n - \beta_n^o) \to^d N(0, I_q).$$

This complements Andrews [1991bΓTheorem 1(a)]. We will approximate $\delta\bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi_o)$ as $G_{np}^{o\prime}(\hat{\beta}_n - \beta_n^o)$ for some $G_{np_n}^o$. For thisΓthe following is appropriate.

Assumption D.5: (a) For each $\pi \in \Pi$ and $\theta \in \Theta_o$, $\delta\bar{m}_n^o(\theta-\theta_o;\pi) = n^{-1}\sum_{t=1}^n g_{nt}(X_t,\pi)(\theta(X_t)-\theta_o(X_t))$, where $g_{nt} : \mathbb{X}\times\Pi \to \mathbb{R}^q$ and $n^{-1}\sum_{t=1}^n g_{nt}(X_t,\pi_o)g_{nt}(X_t,\pi_o)'$ is a $q\times q$ $O(1)$ uniformly positive definite matrix; (b) there exist some $\eta > 0$ and $D_{nt} : \mathbb{X} \to \mathbb{R}^+$, $n^{-1}\sum_{t=1}^n D_{nt}(X_t) = O(1)$, such that for all $x \in \mathbb{X}$ and $\pi_1, \pi_2 \in \Pi$, $\|g_{nt}(x,\pi_1) - g_{nt}(x,\pi_2)\| \leq D_{nt}(x)^{1/2}\rho_\Pi(\pi_1,\pi_2)^\eta$; and (c) $\rho_\Pi(\hat{\pi}_n,\pi_o) = o_P(p_n^{-1/2\eta})$.

Assumption D.5(a) is a "smoothness" (Goldstein and Messer [1992ΓDefinition 3.2]) or "full mean" (Newey [1994ΓAssumption 3.5]) condition. D.5(bΓc) ensure B.3(b).

The main result of this section follows.

Theorem 3.4: *Suppose Assumptions A.2 (for α,π), A.3, B.1, B.3(a), B.4, B.5 with $v_{nt}(Z_t) = G_{np_n}^{o\prime}(\Psi_n'\Psi_n/n)^-\psi_{np_n}(X_t)\varphi(\epsilon_t)/E\varphi'(\epsilon_t)$, B.6, D.1-D.3(a,b), D.4-D.5 hold. Define $M_n = n\hat{m}_n'\hat{J}_n^-\hat{m}_n$, where $\hat{m}_n = m_n(\hat{\alpha}_n,\hat{\theta}_n,\hat{\pi}_n)$ and $\hat{\theta}_n = \hat{\mu}_n$ as in (3.2). Suppose either (a) $\{m_t\}$ is linear in θ; or (b) B.2(a,b) with $(\Theta,\rho_\Theta) = (\mathcal{W}_{\infty,\mathbb{R}}^s(\mathbb{X}),\rho_{s,\infty})$, and D.3(c) hold, and $n^{1/2(1+\lambda)}p_nB_s(p_n)/\lambda_{\min}^{1/2}(\Psi_n'\Psi_n) \to 0$, where λ is as in B.2(a). Then (i) under H_o,*

$$M_n \to^d N(0,1);$$

(ii) under H_A and for any nonstochastic sequence $\{C_n = o(n)\}$,

$$P[M_n > C_n] \to 1.$$

HenceΓasymptotic $n^{1/2}$-normality is attainableΓwith $\hat{\theta}_n$ a series m-estimator. We omit treatment of local alternatives for the sake of brevity.

3.1b First Order M-testing: Application to Testing Normality

We now apply Theorem 3.4 to construct a new test for normality of the regression error of a nonparametric regression. For this purposeΓwe use the following moment vector:

$$m_t(\alpha,\theta,\pi) = \{(Y_t - \mu(X_t))^3, (Y_t - \mu(X_t))^4 - 3\sigma^4\}',$$

where $(\alpha, \theta) = (\sigma^2, \mu)$, and $\alpha_o = \sigma_o^2$ is the unconditional variance of the regression error.

Here we recognize quantities with expectations proportional to the standard measures of skewness and excess kurtosis. White [1982] shows that the vanishing of these two moments is necessary and sufficient for validity of the information matrix equality when estimating the mean and variance of a normal random variable using maximum likelihood. ThusΓtesting for skewness and excess kurtosis gives an information matrix test for normality. AlternativelyΓ Bera and Jarque [1982] obtain a normality test based on these moments by nesting the normal within the Pearson family. This example clearly demonstrates a typical feature of first order m-testing: the hypothesis H_o^* originally of interest (normality) does not exactly coincide with the null hypothesis H_o tested (absence of skewness and excess kurtosis)Γas there are non-normal distributions with no skewness and excess kurtosis. It is for this reason that first order m-testing often fails to deliver consistent tests against H_o^*.

We have the following new result.

Theorem 3.5 [*Testing for Normality*]**:** *Suppose for each* n, *(a)* $Y_t = \mu_o(X_t) + \epsilon_t$, *where* $X_t = (2t-1)/2n$, $t = 1, ..., n$, *and* ϵ_t *is i.i.d. with* $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma_o^2$; *(b)* $\mu_o \in \mathcal{C}^r(0,1)$ *for some* $r > 2$; *(c)* $\psi_j(x) = \sqrt{2}\cos(j-1)\pi x$, $j = 1, 2, ...$; *(d)* φ *satisfies D.2; and (e)* $p_n^4/n \to 0$, $p_n^{2r-\delta}/n \to \infty$ *for any arbitrarily small* $\delta > 0$.

Put $\hat{\sigma}_n^2 = n^{-1}\sum_{t=1}^n \hat{\epsilon}_t^2$ and $\hat{\eta}_t = \varphi(\hat{\epsilon}_t)/(n^{-1}\sum_{t=1}^n \varphi'(\hat{\epsilon}_t))$, *where* $\hat{\epsilon}_t = Y_t - \hat{\mu}_n(X_t)$, *with* $\hat{\mu}_n$ *as in (3.2). Define* $M_n = n\hat{m}_n'\hat{J}_n^{-1}\hat{m}_n$, $\hat{m}_n = n^{-1}\sum_{t=1}^n(\hat{\epsilon}_t^3, \hat{\epsilon}_t^4 - 3\hat{\sigma}_n^4)'$, *and*

$$\hat{J}_n = \begin{bmatrix} \hat{J}_{11} & \hat{J}_{12} \\ \hat{J}_{12} & \hat{J}_{22} \end{bmatrix},$$

where $\hat{J}_{11} = 15\hat{\sigma}_n^6 - 6\hat{\sigma}_n^2 n^{-1}\sum_{t=1}^n \hat{\epsilon}_t^3\hat{\eta}_t + 9\hat{\sigma}_n^4 n^{-1}\sum_{t=1}^n \hat{\eta}_t^2$, $\hat{J}_{12} = -3\hat{\sigma}_n^2 n^{-1}\sum_{t=1}^n \hat{\epsilon}_t^4\hat{\eta}_t + 18\sigma_n^4 n^{-1}\sum_{t=1}^n \hat{\epsilon}_t^2\hat{\eta}_t$, *and* $\hat{J}_{22} = 24\hat{\sigma}_n^8$. *Then (i) under* H_o,

$$M_n \to^d \chi_2^2;$$

(ii) under H_A, *if* $E(\epsilon_t^6) < \infty$, *then for any nonstochastic sequence* $\{C_n = o(n)\}$,

$$P[M_n > C_n] \to 1.$$

Compared to Bera and Jarque's [1982] testΓour test is insensitive to model misspecification for μ, because we use a nonparametric model rather a parametric model. A similar result holds for more general regression designs; we omit this for brevity.

3.2 Second Order M-testing

NextΓwe apply Theorem 2.5 to give a new consistent specification test for parametric models in the presence of heteroskedasticity of unknown formΓas motivated us at the outset. The following parametric specification applies.

Assumption E.1: Let $\{X_t\}$ be a nonstochastic sequence with $v_n \Rightarrow v$, $v_n(\mathbb{B}) = n^{-1}\sum_{t=1}^n 1[X_t \in \mathbb{B}], \mathbb{B} \subseteq \mathbb{R}^d$. For each $x \in \mathbb{X}$, $f(x,\cdot) : \mathbb{A} \to \mathbb{R}$ is twice continuously differentiable on $\mathbb{A}$, with $|f(x,\cdot)|, \|\nabla_\alpha f(x,\cdot)\|^2$ and $\|\nabla_\alpha^2 f(x,\cdot)\|$ dominated by functions integrable with respect to v.

The null hypothesis originally of interest (correct model specification) is

$$H_o^* : v[f(X,\alpha_o) = \mu_o(X)] = 1 \text{ for some } \alpha_o \in \mathbb{A}$$

for μ_o as in Assumption D.1Γand the global alternative is

$$H_A^* : v[f(X,\alpha) \neq \mu_o(X)] > 0 \text{ for all } \alpha \in \mathbb{A}.$$

Using a nonparametric series estimator for μ_o, Hong and White [1995ΓTheorem A.3] propose a consistent test for H_o^* based on (1.1) achieving robustness to heteroskedasticity through use of a heteroskedasticity consistent covariance matrix estimator. This limitation can be avoided by using

$$m_t(\alpha,\theta,\pi) = (\theta(X_t) - f(X_t,\alpha)/\sigma(X_t))(Y_t - f(X_t,\alpha))/\sigma(X_t), \tag{3.3}$$

where $(\theta,\pi) = (\mu/\sigma,\sigma)$. Observe that at $(\theta_o,\pi_o) = (\mu_o/\sigma_o,\sigma_o)$, $m_n(\alpha,\theta_o,\pi_o) = 0$ if and only if $\alpha = \alpha_o$ under H_o^*. HenceΓtests based on (3.3) are consistent against H_o^*. Because (3.3) is degenerate at γ_oΓTheorem 2.5 applies.

Using Θ_n as in (3.1)Γwe form an adaptive nonparametric least squares estimator for θ_o as

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta_n} n^{-1}\textstyle\sum_{t=1}^n (Y_t/\hat{\sigma}_n(X_t) - \theta(X_t))^2. \tag{3.4}$$

where $\hat{\sigma}_n$ is a nonparametric estimator for σ_o. To verify that using $\hat{\sigma}_n$ in place of σ_o has no asymptotic effectΓwe use the following uniform convergence resultΓextending a method of Hall [1988Γ1989] from a finite dimensional space to an infinite dimensional space. This can also be used to verify stochastic equicontinuity.

Theorem 3.6: *Let $(\Omega,\mathcal{F},P)$ be a complete probability space and $\Pi = \{\pi : \mathbb{X} \subset \mathbb{R}^d \to [c,c^{-1}] \mid |\pi(x_1) - \pi(x_2)| \leq \Delta\|x_1 - x_2\|$ for any $x_1, x_2 \in \mathbb{X}\}$, where $0 < \Delta < \infty$ and $\mathbb{X}$ is* *bounded, $d \in \mathbb{N}$. For each n suppose Π_n is a compact subset of Π, and $Q_n : \Omega \times \Pi_n \to \mathbb{R}$ is a stochastic mapping such that $EQ_n(\pi) = 0$ for each $\pi \in \Pi_n$.*

Suppose (a) for each pair $\epsilon, \lambda > 0, \sup_{\pi \in \Pi_n} P[|Q_n(\pi)| > \epsilon] \leq C_1 n^{-\lambda}$; and (b) for each $\lambda > 0$, there exists $\lambda_1 = \lambda_1(\lambda) > 0$ such that $E \sup_{(\pi_1,\pi_2) \in \mathbb{B}_n(\delta_n)} |Q_n(\pi_1) - Q_n(\pi_2)| \leq Cn^{-\lambda}$, where $\mathbb{B}_n(\delta_n) = \{(\pi_1, \pi_2) \in \Pi_n \times \Pi_n : \rho_\infty(\pi_1, \pi_2) < \delta_n = n^{-\lambda_1}\}$, then for each pair $\epsilon, \lambda > 0$,

$$P\left[\sup_{\pi \in \Pi_n} |Q_n(\pi)| > \epsilon\right] \leq C_2 n^{-\lambda}.$$

Assumption D.1: (b′) $\sigma_o \in \Sigma = \{\sigma : \mathbb{X} \to [c, c^{-1}] \mid |\sigma(x_1) - \sigma_2(x_2)| \leq \Delta \|x_1 - x_2\|$ for any $x_1, x_2 \in \mathbb{X}\}$, where $0 < \Delta < \infty$ and $\mathbb{X} \subset \mathbb{R}^d$ is bounded; (c) all moments of ϵ_t are finite.

Assumption E.2: $\hat{\sigma}_n : \Omega \to \Sigma$ is measurable such that $\rho_\infty(\hat{\sigma}_n, \sigma_o) = o(\min[p_n^{-1/2}, (p_n/n)^{1/2+\delta}])$ for any arbitrarily small $\delta > 0$.

Assumption E.3: For ξ_{np_n} as in D.3Γ$\xi_{np_n} \to 0$.

Assumption E.4: There exists a nonstochastic sequence $\{\beta_n^o\}$ such that $\theta_n^o(\cdot) = \psi'_{np_n}(\cdot)\beta_n^o \in \mathcal{W}_{\infty,r}(\mathbb{X})$ and $\rho_\infty(\theta_n^o, \theta_o) = o(p_n^{1/2}/n^{1/2})$.

We now state a CLT for the quadratic form of $\{\psi_{np_n}(X_t)\epsilon_t\}$ in the presence of σ.

Theorem 3.7: *Suppose D.1(a,b′, c), D.3(a) and E.2-E.3 hold. Let $p_n \to \infty$ as $n \to \infty$. Define $W_n(\sigma) = \sum_{t=1}^n \sum_{s=1}^n W_{nts}(\sigma)$, where*

$$W_{nts}(\sigma) = (\sigma_o(X_t)/\sigma(X_t))\epsilon_t \psi'_{np_n}(X_t)(\Psi'_n \Psi_n)^{-1} \psi_{np_n}(X_s)\epsilon_s(\sigma_o(X_s)/\sigma(X_s)).$$

Then

$$(W_n(\hat{\sigma}_n) - p_n)/(2p_n)^{1/2} \to^d N(0,1).$$

This is obtained by first showing $p_n^{-1/2}(W_n(\hat{\sigma}_n) - W_n(\sigma_o)) \to^p 0$ using Theorem 3.6 and then showing $(W_n(\sigma_o) - p_n)/(2p_n)^{1/2} \to^d N(0,1)$ using de Jong's [1987] CLT for quadratic forms.

A new heteroskedasticity-insensitive test complementing Hong and White [1995] follows.

Theorem 3.8: *Suppose C.4, D.1(a,b′,c), D.3(a) and E.1-E.4 hold. Define $M_n = (n\hat{m}_n - p_n)/(2p_n)^{1/2}$, where $\hat{m}_n = n^{-1} \sum_{t=1}^n (\hat{\theta}_n(X_t) - f(X_t, \hat{\alpha}_n)/\hat{\sigma}_n(X_t))(Y_t - f(X_t, \hat{\alpha}_n))/\hat{\sigma}_n(X_t)$, with $\hat{\theta}_n$ as in (3.4). Let $p_n \to \infty$ as $n \to \infty$. Then (i) under H_o^*,*

$$M_n \to^d N(0,1);$$

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

$$P[M_n > C_n] \to 1.$$

4. APPLICATIONS TO NONPARAMETRIC KERNEL ESTIMATION

4.1a First Order M-testing for the i.i.d. Case

Assumption F.1: (a) For each n the random sample $\{Z_1, ..., Z_n\}$ is i.i.d.Γwhere $Z_t = (X_t', Y_t)' \in \mathbb{R}^{d+1}$, $d \in \mathbb{N}$, and $E|Y_t| < \infty$; (b) the support $\mathbb{X} \subset \mathbb{R}^d$ of X is compact and the distribution of X is absolutely continuous on $\mathbb{X}$ with respect to Lebesgue measureΓwith density p_o bounded above and away from zero on $\mathbb{X}$. FurthermoreΓthe sample $\{X_1, ..., X_n\}$ does not include the boundary points of $\mathbb{X}$; and (c) for some $\delta > 0$, $E|Y_t^{2+\delta}| < \infty$ and $\sup_{x \in \mathbb{X}} E[|Y|^{2+\delta} |X = x] < \infty$.

Boundedness of p_o away from below can be relaxed using moving trimming (e.g. Härdle and Stoker [1989] or Robinson [1988]). To avoid boundary effectsΓwe assume $\{X_1, ..., X_n\}$ does not include boundary points of $\mathbb{X}$. Part (c) gives moment conditions for uniform convergence of kernel estimators (e.g. Mack and Silverman [1982]ΓNewey [1994]).

We use the Nadaraya-Watson kernel estimator for $\mu_o(x) = E(Y|X = x)$:

$$\hat{\mu}_n(x) = \begin{cases} (n\hat{p}_n(x))^{-1} \sum_{t=1}^n Y_t K_n(x - X_t) & \text{if } \hat{p}_n(x) \neq 0 \\ 0 & \text{otherwise}\Gamma \end{cases} \tag{4.1}$$

where $\hat{p}_n(x) = n^{-1} \sum_{t=1}^n K_n(x - X_t)$, $K_n(x - X_t) = b_n^{-d} K[(x - X_t)/b_n]$, and b_n is a bandwidth. We now impose regularity conditions on K, μ_o and p_o.

Assumption F.2: $K : \mathbb{T} \to \mathbb{R}$ is a symmetric bounded kernel of finite order k with compact support $\mathbb{T} = [-\tau, \tau]^d$, $0 < \tau < \infty$, such that K is differentiable of order $s \geq 0$, with Lipschitz s-th derivativeΓ$\int_{\mathbb{T}} K(u)du = 1$, $\int_{\mathbb{T}} u_1^{i_1} u_2^{i_2} \cdots u_d^{i_d} K(u)du = 0$ for $|i| = \sum_{j=1}^d i_j < k$, and $\int_{\mathbb{T}} u_1^{i_1} u_2^{i_2} \cdots u_d^{i_d} K(u)du \neq 0$ for $|i| = k$.

Assumption F.3: There exist extensions of p_o and $p_o \mu_o$ such that these extensions are in $\mathcal{W}_{\infty,r}^r(\mathbb{R}^d)$ for some integer $r > 0$.

We use a uniform convergence result due to Newey [1994].

Lemma 4.1: (*Newey [1994, Theorem B.1]*): *Suppose Assumptions F.1-F.3 hold with* $r \geq s + k$. *Let* $b_n \to 0$, $n^{\delta/(2+\delta)} b_n^d / \ln(n) \to \infty$, *where* δ *is as in F.1. Then (i)* $\rho_{s,\infty}(\hat{\mu}_n, \mu_o) = O_P([nb_n^{d+2s}/\ln(n)]^{-1/2} + b_n^k)$; *and (ii)* $\rho_{s,\infty}(\hat{p}_n, p_o) = O_P([nb_n^{d+2s}/\ln(n)]^{-1/2} + b_n^k)$.

This delivers explicit rates for b_n satisfying certain conditions of Theorem 2.3.

We now impose conditions on $m_t(\alpha, \theta, \pi) = m(Z_t, \alpha, \mu, \pi)$.

Assumption F.4: (a) Let $\theta_o \in \Theta_o \subset \Theta$. For each $\theta \in \Theta_o$ and each $\pi \in \Pi$, $\delta \bar{m}_n^o(\theta - \theta_o; \pi) = E[g(X, \pi)(\theta(X) - \theta_o(X))]$, where $g : \mathbb{X} \times \Pi \to \mathbb{R}^q$ is such that for each $\pi \in \Pi \Gamma g(\cdot, \pi) \in \mathcal{C}^k(\mathbb{X})$, $E[g(X, \pi_o)g(X, \pi_o)']$ is a $q \times q$ finite positive semi-definite matrix and $\sup_{x \in \mathbb{X}} ||D^k g(x, \pi_o)|| \leq \Delta < \infty$; (b) there exist some $\eta > 0$ and $D : \mathbb{R}^d \to \mathbb{R}^+ \Gamma ED(X) < \infty$, such that for all $x \in \mathbb{X}$ and $\pi_1, \pi_2 \in \Pi$, $\|g(x, \pi_2) - g(x, \pi_1)\| \leq D(x)^{1/2} \rho_\Pi(\pi_1, \pi_2)^\eta$; and (c) $\rho_\Pi(\hat{\pi}_n, \pi_o) = o_P((nb_n^d)^{-1/2\eta})$.

F.4(a) is a "smoothness" or "full mean" assumption. F.4(bIc) ensure that replacing $\hat{\pi}_n$ with π_o does not affect the limiting distribution of $\delta \bar{m}_n^o(\theta - \theta_o; \hat{\pi}_n)$. We use the following key result.

Proposition 4.2: *Suppose F.1(a,b) and F.2-F.4(a) hold and* $V_o = E[g(X, \pi_o)g(X, \pi_o)'\epsilon^2]$ *is finite and nonsingular, where* $\epsilon = Y - \mu_o(X)$. *Let* $n^{\delta/(2+\delta)} b_n^d / \ln(n) \to \infty$, $nb_n^{2d}/\ln(n) \to \infty$, $nb_n^{2k} \to 0$, *and* $2k > d$. *Then* $n^{1/2} \delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi_o) = n^{-1/2} \sum_{t=1}^n g(X_t, \pi_o)\epsilon_t + o_P(1)$, *and*

$$n^{1/2} \delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi_o) \to^d N(0, V_o).$$

This is necessary but not sufficient for B.5; the following suffices.

Assumption F.5: For $W(Z_t, \gamma_o) = m(Z_t, \gamma_o) + \nabla'_\alpha \bar{m}(\gamma_o) s(Z_t, \alpha_o) + g(X_t, \pi_o)\epsilon_t$, $J_o = E(W(Z, \gamma_o)W(Z, \gamma_o)')$ is a $q \times q$ finite positive definite matrix.

Theorem 4.3: *Suppose Assumptions A.2 (for* α, π*), A.3, B.1, B.3(a), B.4, B.6 with* $J_n^o = J_o$*, F.1(a,b) and F.2-F.5 hold. Define* $M_n = n \hat{m}_n' \hat{J}_n^- \hat{m}_n$, *where* $\hat{m}_n = m_n(\hat{\alpha}_n, \hat{\theta}_n, \hat{\pi}_n)$ *and* $\hat{\theta}_n = \hat{\mu}_n$ *is as in (4.1). Let* $nb_n^d \to \infty$, $nb_n^{2k} \to 0$, $2k > d$. *Suppose either (a)* $\{m_t\}$ *is linear in* θ *or (b) B.2(a,b) with* $(\Theta, \rho_\Theta) = (\mathcal{W}_{r,\infty}^s(\mathbb{X}), \rho_\infty)$, *and F.1(c) hold,* $n^{\delta/(2+\delta)} b_n^d / \ln(n) \to \infty$, $n^\lambda b_n^{(d+2s)(1+\lambda)} / \ln^{(1+\lambda)}(n) \to \infty$, $nb_n^{2k(1+\lambda)} \to 0$, *where* δ *is as in F.1 and* λ *as in B.2. Then (i) under* H_o

$$M_n \to^d \chi_q^2;$$

(ii) under H_A and for any nonstochastic sequence $\{C_n = o(n)\}$,

$$P[M_n > C_n] \to 1.$$

Asymptotic $n^{1/2}$-normality is thus achievedΓdespite the presence of $\hat{\theta}_n$.

4.1b First Order M-testing: Application to Testing Omitted Variables

Although nonparametric regressions do not require specification of functional formΓthey do require a priori knowledge of relevant explanatory variables. One may be interested in testing the relevance of additional variables. We now give a new test for omitted variables insensitive to model misspecification.

Put $Z_t = (Y_t, X_t')' = (Y_t, (X_{1t}', X_{2t}'))'$, where X_{1t} is a $d_1 \times 1$ random vector with density p_1^o and X_{2t} is a $d_2 \times 1$ random vectorΓ$d_1 + d_2 = d$. Suppose one is interested in testing the relevance of X_{2t} in explaining Y_t. Then the hypotheses originally of interest are

$$H_o^* : P[E(Y_t|X_{1t}) = E(Y_t|X_t)] = 1 \text{ v.s. } \; H_A^* : P[E(Y_t|X_{1t}) = E(Y_t|X_t)] < 1.$$

Put $\mu_1^o(X_{1t}) = E(Y_t|X_{1t})$ and consider the moment function

$$m(Z_t, \alpha, \theta, \pi) = \psi(X_{2t})(Y_t - \mu_1(X_{1t})),$$

where $\theta = \mu_1$, ψ is a given weighting functionΓand α and π are null. Now H_o^* implies $Em(Z_t, \alpha, \theta_o, \pi)$ for $\theta_o = \mu_1^o$. We note that Robinson [1989Γ5.52(e)] suggests a similar approach to testing H_o^* with the choice of $\psi(X_{2t}) = X_{2t}$, but does not construct a test statistic.

To construct our statisticΓwe use a kernel estimator for μ_1^o :

$$\hat{\mu}_{1n}(x_1) = \begin{cases} (n\hat{p}_{1n}(x_1))^{-1} \sum_{t=1}^n Y_t K_n(x_1 - X_{1t}) & \text{if } \hat{p}_{1n}(x_1) \neq 0 \\ 0 & \text{otherwise}Γ \end{cases} \tag{4.2}$$

where $\hat{p}_{1n}(x_1) = n^{-1} \sum_{t=1}^n K_n(x_1 - X_{1t})$, $K_n(x_1 - X_{1t}) = b_n^{-1} K((x_1 - X_{1t})/b_n)$, with $K : \mathbb{R}^{d_1} \to \mathbb{R}$ a kernel and b_n a bandwidth.

We also use the following kernel estimator for $g(X_{1t}) = E(\psi(X_{2t})|X_{1t})$:

$$\hat{g}_n(x_1) = \begin{cases} (n\hat{p}_{1n}(x_1))^{-1} \sum_{t=1}^n \psi(X_{2t}) K_n(x_1 - X_{1t}) & \text{if } \hat{p}_{1n}(x_1) \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 4.4 [*Testing for Omitted Variables*]**:** *Suppose (a) F.1 with $\delta = 2$ hold; (b) $\psi : \mathbb{R}^{d_2} \to \mathbb{R}$ is measurable such that $E\psi^4(X_{2t}) < \infty$ and $sup_{x_1 \in \mathbb{X}_1} E(\psi^4(X_{2t})|X_{1t} = x_1) < \infty$, where $\mathbb{X}_1$ is the compact support of X_1; (c) F.2 with $s = 0$ holds; (d) p_1^o and $p_1^o \mu_1^o$ satisfy F.3; (e) $g(X_{1t}) = E(\psi(X_{2t})|X_{1t}) \in \mathcal{C}^k(\mathbb{X}_1)$ with $\sup_{x_1 \in \mathbb{X}_1} ||D^k g(x_1)|| < \infty$; (f) $nb_n^{2d_1} / \ln^2(n) \to \infty$, $b_n \to 0$.*

Define $M_n = n\hat{m}_n' \hat{J}_n^{-1} \hat{m}_n$, where $\hat{m}_n = n^{-1} \sum_{t=1}^n \psi(X_{2t})(Y_t - \hat{\mu}_{1n}(X_{1t}))$, $\hat{J}_n = n^{-1} \sum_{t=1}^n (\psi(X_{2t}) - \hat{g}_n(X_{1t}))^2 (Y_t - \hat{\mu}_{1n}(X_t))^2$. Then (i) under $H_o : E[\psi(X_{2t})(E(Y_t|X_t) - E(Y_t|X_{1t}))] = 0$,

$$M_n \to^d \chi_1^2;$$

(ii) suppose $H_A : E[\psi(X_{2t})(Y_t - \mu_1^o(X_{1t}))] \neq 0$ holds. Then for any nonstochastic sequence $\{C_n = o(n^{1/2})\}$,

$$P[|M_n| > C_n] \to 1.$$

This test is not necessarily consistent against H_A^* as H_o^* implies H_o but the converse may fail. Power depends on choice of ψ. For consistency we must choose ψ so that H_o^* coincides with H_o. Such choices exist; see Bierens [1990] and Stinchcombe and White [1998].

Lavergne and Vuong [1996] propose a method to determine relevant regressors using kernel estimatorsΓbut this does not apply hereΓdue to the degeneracy of their statistic. Our approach complements theirs. One could also use Theorem 2.5 to construct a consistent test for H_o^*; we leave this for further work.

4.2 Second Order M-testing

In Section 3.2 we gave a new heteroskedasticity-insensitive consistent specification test for the parametric model $f(X_t, \alpha)$ using an estimate of conditional variance. We now give a heteroskedasticity insensitive consistent specification test using kernel regression and a new heteroskedasticity-consistent covariance matrix estimator.

We use a weighted version of (1.1)Γi.e. $m_t(\alpha, \theta, \pi) = p(X_t)(\mu(X_t) - f(X_t, \alpha))(Y_t - f(X_t, \alpha))$. Put $\theta = (\theta_1, \theta_2) = (r, p) = (p\mu, p)$ and let π be null. Then

$$m_t(\alpha, \theta, \pi) = (r(X_t) - p(X_t) f(X_t, \alpha))(Y_t - f(X_t, \alpha)). \tag{4.3}$$

As (4.3) is degenerate at $(\alpha_o, \theta_o) = (\alpha_o, (p_o \mu_o, p_o))$, where $\alpha_o \in \mathbb{A}$ is such that $H_o^* : P[f(X_t, \alpha_o) = \mu_o(X_t)] = 1$ holdsΓTheorem 2.5 applies. A consistent test against $H_A^* : P[f(X_t, \alpha) \neq \mu_o(X_t)] > 1$ for all $\alpha \in \mathbb{A}$ can be based on

$$\hat{m}_n = n^{-1} \textstyle\sum_{t=1}^n (\hat{r}_n(X_t) - \hat{p}_n(X_t) f(X_t, \hat{\alpha}_n))(Y_t - f(X_t, \hat{\alpha}_n)), \tag{4.4}$$

where $\hat{r}_n = \hat{p}_n\hat{\mu}_n$, and $\hat{\mu}_n$ and $\hat{p}_n$ are as in (4.1). We make following additional assumptions.

Assumption F.6: For each $\alpha \in \mathbb{A}$, $f(\cdot, \alpha) : \mathbb{X} \to \mathbb{R}$ is measurable; (b) $f(X, \cdot)$ is twice continuously differentiable *a.s.* on $\mathbb{A}$, with $|f(X,\cdot)|\Gamma||\nabla_\alpha f(X,\cdot)||^2$ and $||\nabla^2_\alpha f(X,\cdot)||$ dominated by $D : \mathbb{X} \to \mathbb{R}^+$, $ED^2(X) < \infty$.

Assumption F.7: $\sigma^2_o(X) = var(Y|X)$ is continuous on $\mathbb{X}$.

The next result is the key to obtaining the distribution of our statistic.

Theorem 4.5: *Suppose F.1 (with $\delta = 2$), F.2-F.3 and F.7 hold. Let $W_n = n^{-2}\sum_{t=1}^n \sum_{s=1}^n W_{nts}$, $W_{nts} = \epsilon_t \epsilon_s K_n(X_t - X_s)$, $\epsilon_t = Y_t - \mu_o(X_t)$. Define $J_o = 2C(K)E(\sigma^4_o(X)p_o(X))$, $C(K) = \int_T K^2(u)du$. Let $nb^d_n \to \infty$, $b_n \to 0$. Then*

$$J_o^{-1/2} nb_n^{d/2}(W_n - EW_n) \to^d N(0,1).$$

This is obtained by applying de Jong's [1987] CLT for generalized quadratic forms. NextΓ we propose a heteroskedasticity-consistent U-statistic estimator for J_o.

Proposition 4.6: *Suppose Assumptions F.1 (with $\delta = 2$), F.2-F.3 and F.7 hold. Define $\hat{J}_n = 4C(K)b^d_n n^{-2}\sum_{t=2}^n \sum_{s=1}^{t-1} \hat{\epsilon}^2_{nt}\hat{\epsilon}^2_{ns}K_n(X_t - X_s)$, where $\hat{\epsilon}_{nt} = Y_t - \hat{\mu}_n(X_t)$ and $\hat{\mu}_n$ is as in (4.1). Let $nb^{3d}_n \to \infty$, $b_n \to 0$ and $2k > d$. Then $\hat{J}_n - J_o \to^p 0$.*

Now the new heteroskedasticity-insensitive consistent test can be given.

Theorem 4.7: *Suppose Assumptions C.4, F.1 (with $\delta = 2$), F.2(with $s = 0$), F.3 and F.6-F.7 hold. Define $M_n = \hat{J}_n^{-1/2} nb_n^{d/2}(\hat{m}_n - \hat{R}_n)$, where $\hat{J}_n$ is as in Proposition 4.6, $\hat{m}_n$ is as in (4.4) and $\hat{R}_n = (nb^d_n)^{-1}K(0)\hat{\sigma}^2_n$, with $\hat{\sigma}^2_n = n^{-1}\sum_{t=1}^n \hat{\epsilon}^2_{nt}$, $\hat{\epsilon}_{nt} = Y_t - \hat{\mu}_n(X_t)$ and $\hat{\mu}_n$ as in (4.1). Let $nb^{3d}_n \to \infty$, $nb^{2k+d}_n \to 0$, $2k > d$. Then (i) under H^*_o,*

$$M_n \to^d N(0,1);$$

(ii) under H^*_A and any nonstochastic sequence $\{C_n = o(nb_n^{d/2})\}$,

$$P[M_n > C_n] \to 1.$$

The growth rate of M_n under H_A is $nb_n^{d/2}$, faster than $n^{1/2}$ because $nb^d_n \to \infty$; howeverΓM_n can only detect local alternatives of $O(n^{-1/2}b_n^{-d/4})$, slightly slower than $O(n^{-1/2})$.

MATHEMATICAL APPENDIX

Proof of Theorem 2.3: (i) Given B.2 and by Hölder's inequalityΓwe have

$$||m_n(\alpha_n^o, \hat{\theta}_n, \hat{\pi}_n) - m_n(\alpha_n^o, \theta_o, \hat{\pi}_n) - \delta m_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n)|| \leq D_n \rho_\Theta(\hat{\theta}_n, \theta_o)^{1+\lambda} = o_P(n^{-1/2}), \quad \text{(A1)}$$

where $D_n = n^{-1}\sum_t D_{nt}(Z_t) = O_P(1)$ by Markov's inequality. By the mean value theoremΓ

$$\hat{m}_n = m_n(\alpha_n^o, \hat{\theta}_n, \hat{\pi}_n) + \nabla'_\alpha m_n(\tilde{\alpha}_n, \hat{\theta}_n, \hat{\pi}_n)(\hat{\alpha}_n - \alpha_n^o)$$

given B.1(a)Γ where a different $\tilde{\alpha}_n$ ($\|\tilde{\alpha}_n - \alpha_n^o\| \leq \|\hat{\alpha}_n - \alpha_n^o\|$) appears in each row of $\nabla_\alpha m_n(\cdot, \hat{\theta}_n, \hat{\pi}_n)$. Substituting $m_n(\alpha_n^o, \hat{\theta}_n, \hat{\pi}_n)$ into (A1) and rearrangingΓwe obtain

$$\hat{m}_n = m_n(\alpha_n^o, \theta_o, \hat{\pi}_n) + \nabla'_\alpha m_n(\tilde{\alpha}_n, \hat{\theta}_n, \hat{\pi}_n)(\hat{\alpha}_n - \alpha_n^o) + \delta m_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) + o_P(n^{-1/2}). \quad \text{(A2)}$$

For the first term in (A2)Γwe have

$$\begin{aligned} m_n(\alpha_n^o, \theta_o, \hat{\pi}_n) &= \bar{m}_n(\alpha_n^o, \theta_o, \hat{\pi}_n) + (m_n(\alpha_n^o, \theta_o, \hat{\pi}_n) - \bar{m}_n(\alpha_n^o, \theta_o, \hat{\pi}_n)) \\ &= \bar{m}_n(\gamma_n^o) + (m_n(\gamma_n^o) - \bar{m}_n(\gamma_n^o)) + o_P(n^{-1/2}) \qquad \text{(A3)} \end{aligned}$$

given A.3 and $a_n = n^{1/2}$. For the second term in (A2)Γwe have

$$\begin{aligned} ||\nabla_\alpha m_n(\tilde{\alpha}_n, \hat{\theta}_n, \hat{\pi}_n) - \nabla_\alpha \bar{m}_n(\gamma_n^o)|| &\leq ||\nabla_\alpha m_n(\tilde{\alpha}_n, \hat{\theta}_n, \hat{\pi}_n) - \nabla_\alpha \bar{m}_n(\tilde{\alpha}_n, \hat{\theta}_n, \hat{\pi}_n)|| \\ &\quad + ||\nabla_\alpha \bar{m}_n(\tilde{\alpha}_n, \hat{\theta}_n, \hat{\pi}_n) - \nabla_\alpha \bar{m}_n(\gamma_n^o)|| \\ &= o_P(1) \end{aligned}$$

by the triangle inequality and B.1. HenceΓwe obtain

$$\nabla'_\alpha m_n(\tilde{\alpha}_n, \hat{\theta}_n, \hat{\pi}_n)(\hat{\alpha}_n - \alpha_n^o) = \nabla'_\alpha \bar{m}_n(\gamma_n^o) S_n + o_P(n^{-1/2}) \quad \text{(A4)}$$

given B.4 and B.5Γwhich implies $\hat{\alpha}_n - \alpha_n^o = O_P(n^{-1/2})$. For the last term in (A2)Γwe have

$$\begin{aligned} \delta m_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) &= \delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi_o) + (\delta m_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) - \delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n)) \\ &\quad + (\delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) - \delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi_o)) \\ &= \delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi_o) + o_P(n^{-1/2}) \\ &= V_n + o_P(n^{-1/2}) \qquad \text{(A5)} \end{aligned}$$

given B.3. Substituting (A3)-(A5) into (A2)Γwe obtain $\hat{m}_n = \bar{m}_n(\gamma_n^o) + W_n + o_P(n^{-1/2})$. It follows that $J_n^{o-1/2} n^{1/2} \hat{m}_n \to^d N(J_n^{o-1/2} n^{1/2} \bar{m}_n(\gamma_n^o), I_q)$ by B.5. By Slutsky's TheoremΓwe have $n^{1/2} \hat{J}_n^{-1/2} \hat{m}_n \to^d N(J_n^{o-1/2} n^{1/2} \bar{m}_n(\gamma_n^o), I_q)$ given $\hat{J}_n^- - J_n^{o-1} = o_P(1)$ from B.6. HenceΓ $M_n \to^d \chi_q^2(\zeta_n^o)$.

(ii) Given A.2ΓA.4 and B.6Γwe have $\hat{m}_n - \bar{m}_n(\gamma_n^o) \to^p 0$ and $\hat{J}_n^- - J_n^{-1} \to^p 0$. HenceΓ $M_n/n = \bar{m}_n'(\gamma_n^o)J_n^{-1}\bar{m}_n(\gamma_n^o) + o_P(1)$ by continuityΓwhere $\bar{m}_n'(\gamma_n^o)J_n^{-1}\bar{m}_n(\gamma_n^o) \geq c > 0$ for all n sufficiently large under H_A. The desired result follows immediately. ■

Proof of Theorem 2.5: (i) Given C.2 and Hölder's inequalityΓwe have

$$m_n(\alpha_n^o, \hat{\theta}_n, \hat{\pi}_n) - m_n(\alpha_n^o, \theta_o, \hat{\pi}_n) - \delta\hat{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) - \delta^2\hat{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) = o_P(a_n^{-1}). \quad \text{(A6)}$$

Under C.1(a)Γa second order Taylor expansion of $\hat{m}_n$ about α_n^o yields

$$\hat{m}_n = m_n(\alpha_n^o, \hat{\theta}_n, \hat{\pi}_n) + \nabla_\alpha' m_n(\alpha_n^o, \hat{\theta}_n, \hat{\pi}_n)(\hat{\alpha}_n - \alpha_n^o) + \frac{1}{2}(\hat{\alpha}_n - \alpha_n^o)'\nabla_\alpha^2 m_n(\tilde{\alpha}_n, \hat{\theta}_n, \hat{\pi}_n)(\hat{\alpha}_n - \alpha_n^o),$$

where a different $\tilde{\alpha}_n$ ($\|\tilde{\alpha}_n - \alpha_n^o\| \leq \|\hat{\alpha}_n - \alpha_n^o\|$) appears in each row of $\nabla_\alpha^2 m_n(\cdot, \hat{\theta}_n, \hat{\pi}_n)$. Substituting $m_n(\alpha_n^o, \hat{\theta}_n, \hat{\pi}_n)$ into (A6) and rearrangingΓwe obtain

$$\begin{aligned} \hat{m}_n &= m_n(\alpha_n^o, \theta_o, \hat{\pi}_n) + \nabla_\alpha' m_n(\alpha_n^o, \hat{\theta}_n, \hat{\pi}_n)(\hat{\alpha}_n - \alpha_n^o) \\ &\quad + \frac{1}{2}(\hat{\alpha}_n - \alpha_n^o)'\nabla_\alpha^2 m_n(\tilde{\alpha}_n, \hat{\theta}_n, \hat{\pi}_n)(\hat{\alpha}_n - \alpha_n^o) \\ &\quad + [\delta\hat{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) + \delta^2\hat{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n)] + o_P(a_n^{-1}). \end{aligned} \quad \text{(A7)}$$

Given C.1 and Definition 2.4(b)Γwe have

$$\nabla_\alpha m_n(\alpha_n^o, \hat{\theta}_n, \hat{\pi}_n) = \nabla_\alpha \bar{m}_n(\gamma_n^o) + (\nabla_\alpha m_n(\alpha_n^o, \hat{\theta}_n, \hat{\pi}_n) - \nabla_\alpha \bar{m}_n(\gamma_n^o)) = o_P(n^{1/2}/a_n), \quad \text{(A8)}$$

and

$$||\nabla_\alpha^2 m_n(\tilde{\alpha}_n, \hat{\theta}_n, \hat{\pi}_n)|| \leq n^{-1}\textstyle\sum_t D_{nt}(Z_t) = O_P(1) \quad \text{(A9)}$$

by Markov's inequality. On the other handΓ

$$\begin{aligned} \delta\hat{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) &= (\delta\hat{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) - \delta\bar{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n)) \\ &\quad + (\delta\bar{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) - \delta\bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi_o)) + \delta\bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi_o) \\ &= (\delta\hat{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) - \delta\bar{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n)) + o_P(a_n^{-1}) \end{aligned}$$

given C.3(d) and Definition 2.4(c)Γand

$$\begin{aligned} \delta^2\hat{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) &= \delta^2\bar{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) + (\delta^2\hat{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) - \delta^2\bar{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n)) \\ &= \delta^2\bar{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) + o_P(a_n^{-1}) \end{aligned}$$

given C.3(a). It follows that

$$\begin{aligned} &\delta\hat{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) + \delta^2\hat{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) \\ =\ &\delta\hat{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) - \delta\bar{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) + \delta^2\bar{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) + o_P(a_n^{-1}) \\ =\ &W_n(\hat{\pi}_n) + o_P(a_n^{-1}) \\ =\ &EW_n(\pi_o) + (EW_n(\hat{\pi}_n) - EW_n(\pi_o)) + (W_n(\hat{\pi}_n) - EW_n(\hat{\pi}_n)) + o_P(a_n^{-1}) \\ =\ &EW_n(\pi_o) + (W_n(\pi_o) - EW_n(\pi_o)) + o_P(a_n^{-1}) \end{aligned} \quad \text{(A10)}$$

given C.3(b,c,d). Substituting (A8-A10) into (A7) and using A.3, C.4, $a_n/n \to 0$, and Definition 2.4(a), we obtain

$$\begin{aligned}\hat{m}_n &= \bar{m}_n(\gamma_n^o) + (m_n(\gamma_n^o) - \bar{m}_n(\gamma_n^o)) + EW_n(\pi_o) + (W_n(\pi_o) - EW_n(\pi_o)) + o_P(a_n^{-1}) \\ &= \bar{m}_n(\gamma_n) + EW_n(\pi_o) + (W_n(\pi_o) - EW_n(\pi_o)) + o_P(a_n^{-1}).\end{aligned}$$

Consequently, given C.5, we have $J_n^{o-1/2} a_n(\hat{m}_n - EW_n(\pi_o)) \to^d N(J_n^{o-1/2} a_n \bar{m}_n(\gamma_n^o), I_q)$. It follows by Slutsky's Theorem that $\hat{J}_n^{-1/2} a_n(\hat{m}_n - \hat{R}_n) \to^d N(J_n^{o-1/2} a_n \bar{m}_n(\gamma_n^o), I_q)$ given $\hat{J}_n^- - J_n^{o-1} \to^p 0$ and $a_n(\hat{R}_n - EW_n(\pi_o)) \to^p 0$ by C.6. Therefore, $M_n \to^d \chi_q^2(\zeta_n^o)$.

(ii) The proof of consistency is similar to that of Theorem 2.3(ii). ■

Proof of Proposition 3.2: Put $\hat{\mu}_{nt} = \hat{\mu}_n(X_t)$, $\mu_{nt}^o = \mu_n^o(X_t)$, and $\mu_t^o = \mu_o(X_t)$. We first apply Yohai and Maronna (YM) [1979, Theorem 2.2] to show $n^{-1}\sum_t(\hat{\mu}_{nt} - \mu_{nt}^o)^2 = O_P(p_n/n)$. YM assume a linear model of the form (see YM (Eq.(1.1)))

$$Y_{nt}^o = \psi_{np_n}'(X_t)\beta_n^o + \epsilon_t = \mu_n^o(X_t) + \epsilon_t,\ t = 1, 2, ..., n,\ n = 1, 2, ...,$$

where $\psi_{np_n}(X_t)$ is a given $p_n \times 1$ vector, and ϵ_t is i.i.d. This is a moving DGP assumption. With this in mind, we must control the bias $\mu_n^o - \mu_o$ properly to apply YM's results to $\hat{\mu}_n$.

Put $\zeta_{nt} = (\Psi_n'\Psi_n)^{-1/2}\psi_{np_n}(X_t)$, $\tilde{\beta}_n = (\Psi_n'\Psi_n)^{1/2}\hat{\beta}_n$, $\beta_n^+ = (\Psi_n'\Psi_n)^{1/2}\beta_n^o$. Then $Y_{nt}^o = \zeta_{nt}'\beta_n^+ + \epsilon_t$. Following the proof of YM, we see that to apply YM's Theorem 2.2 to $\tilde{\beta}_n$, it suffices that

$$||b_n' \textstyle\sum_t \zeta_{nt}\varphi(Y_{nt}^o - \zeta_{nt}'\tilde{\beta}_n)|| \quad \to \quad 0\ a.s. \qquad \text{(A11)}$$

for any $b_n \in \mathbb{R}^{p_n}$ with $||b_n|| = O(1)$. By applying the mean value theorem to the first order condition (3.2) term by term, we obtain

$$0 \;=\; \textstyle\sum_t \zeta_{nt}\varphi(Y_t - \zeta_{nt}'\tilde{\beta}_n) \;=\; \sum_t \zeta_{nt}\varphi(Y_{nt}^o - \zeta_{nt}'\tilde{\beta}_n) + \sum_t \zeta_{nt}\varphi'(\bar{Y}_{nt})(\mu_{nt}^o - \mu_t^o),$$

where $\bar{Y}_{nt}$ lies between Y_{nt} and Y_{nt}^o. Hence, it suffices for (A11) to hold if $||b_n'\sum_t \zeta_{nt}\varphi'(\bar{Y}_{nt})(\mu_{nt}^o - \mu_t^o)|| \to 0\ a.s.$ Given D.2, D.4(a), the identity $\sum_t \zeta_{nt}\zeta_{nt}' = I_{p_n}$ and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}||b_n' \textstyle\sum_t \zeta_{nt}\varphi'(\bar{Y}_t)(\mu_{nt}^o - \mu_t^o)|| &\le c^{-1}\left\{b_n'\left(\textstyle\sum_t \zeta_{nt}\zeta_{nt}'\right)b_n\right\}^{1/2}\left\{\textstyle\sum_t(\mu_{nt}^o - \mu_t^o)^2\right\}^{1/2} \\ &\le c^{-1}n^{1/2}\,\|b_n\|^2\,\rho_\infty(\mu_n^o, \mu_o) \to 0.\end{aligned}$$

Hence, asymptotically, $\tilde{\beta}_n$ can be viewed as a solution to (A11), which is equivalent to Eq. (2.7) of YM. The results of YM then apply to (A11). Given D.1-D.2, D.3(a), D.4(a) and

$p_n\zeta_{np_n} \to 0$, the conditions of YM [1979, Theorem 2.2] are satisfied. Hence, $p_n^{-1/2}||\tilde{\beta}_n - \beta_n^+|| = O_P(1)$, i.e., $(\hat{\beta}_n - \beta_n^o)'(\Psi_n'\Psi_n)(\hat{\beta}_n - \beta_n^o) = O_P(p_n)$. It follows that

$$\begin{aligned}\textstyle\sum_t(\hat{\mu}_{nt} - \mu_t^o)^2 &\leq 2\textstyle\sum_t(\hat{\mu}_{nt} - \mu_{nt}^o)^2 + 2\textstyle\sum_t(\mu_{nt}^o - \mu_t^o)^2 \\ &\leq 2(\hat{\beta}_n - \beta_n^o)'(\Psi_n'\Psi_n)(\hat{\beta}_n - \beta_n^o) + O_P(n\rho_\infty^2(\mu_n^o, \mu_o)) \\ &= O_P(p_n)\end{aligned}$$

given D.4(a). Because $(\hat{\beta}_n - \beta_n^o)'(\Psi_n'\Psi_n)(\hat{\beta}_n - \beta_n^o) \geq \lambda_{\min}(\Psi_n'\Psi_n)||\hat{\beta}_n - \beta_n^o||^2$, we also have $||\hat{\beta}_n - \beta_n^o||^2 = O_P(p_n^{1/2}/\lambda_{\min}^{1/2}(\Psi_n'\Psi_n))$. ■

Proof of Proposition 3.3: We apply Mammen [1989, Theorem 4]. Like YM, Mammen also considers a linear model of the form of (A11). Following the proofs of both his Theorems 1 and 4, we see that Mammen's results can be applied to (A11), which holds given D.2 and D.4(a), as has been shown in the proof of Proposition 3.2.

Let b_n be any sequence of $p_n \times q$ nonstochastic matrices such that $b_n'b_n$ is a $q \times q$ uniformly nonsingular matrix with $\|b_n'b_n\|$ bounded. We first show

$$\tilde{I}_n^{-1/2}b_n'(\tilde{\beta}_n - \beta_n^+) \to^d N(0, I_q), \tag{A12}$$

where $\tilde{\beta}_n = (\Psi_n'\Psi_n)^{1/2}\hat{\beta}_n$, $\beta_n^+ = (\Psi_n'\Psi_n)^{1/2}\beta_n^o$, and $\tilde{I}_n = b_n'b_n\sigma^2(\varphi)$, with $\sigma^2(\varphi) = E\varphi^2(\epsilon_t)/E^2\varphi'(\epsilon_t)$. Since Mammen only considers the univariate case ($q = 1$), we use the Cramer-Wold device (e.g. White [1984, p.108]) to prove (A12).

Let $h \in \mathbb{R}^q$ be an arbitrary constant with $h'h = 1$. Define $c_n = b_nh$, a $p_n \times 1$ vector (thus c_n is equivalent to α_n in Mammen [1989]). Given D.1-D.3(a,b), Condition (2.1) of Mammen [1989] is satisfied. Note that we impose $p_n\zeta_{np_n} \to 0$ to ensure that $b_n'(\tilde{\beta}_n - \beta_n^+)$ is centered at zero asymptotically. It remains to show that $\|c_n\|$ is bounded below and above. Since $b_n'b_n$ is a $q \times q$ symmetric bounded uniformly nonsingular matrix, $0 < c \leq \lambda_{\min}(b_n'b_n) \leq \lambda_{\max}(b_n'b_n) \leq c^{-1} < \infty$. Hence, $\lambda_{\min}(b_n'b_n) \leq c_n'c_n = h'b_n'b_nh \leq \lambda_{\max}(b_n'b_n)$, i.e., $\|c_n\|$ is bounded below and above. Thus, (A12) now follows by Mammen [1989, Theorem 4]. Next we show

$$I_n^{o-1/2}G_{np_n}'(\hat{\beta}_n - \beta_n^o) \to^d N(0, I_q) \tag{A13}$$

for $I_n^o = G_{np_n}'(\Psi_n'\Psi_n)^-G_{np}\sigma^2(\varphi)$. Define the $p_n \times 1$ vector $b_n^o = a_n(\Psi_n'\Psi_n)^{-1/2}G_{np_n}$, where $a_n^{-1} = ||G_{np_n}'(\Psi_n'\Psi_n)^-G_{np_n}||$. It follows immediately that $b_n^{o\prime}b_n^o$ is a $q \times q$ $O(1)$ symmetric uniformly nonsingular matrix and $\mathrm{tr}(b_n^{o\prime}b_n^o) = 1$. Hence, from (A12) we have $\tilde{I}_n^{o-1/2}a_nG_{np_n}'(\Psi_n'\Psi_n)^{-1/2}$ $(\tilde{\beta}_n - \beta_n^+) \to^d N(0, I_q)$, where $\tilde{I}_n^o = b_n^{o\prime}b_n^o\sigma^2(\varphi) = a_n^2I_n^o$. Because $\tilde{\beta}_n - \beta_n^+ = (\Psi_n'\Psi_n)^{1/2}(\hat{\beta}_n - \beta_n^o)$, this is equivalent to (A13). ■

Proof of Theorem 3.4: Put $\hat{\mu}_{nt} = \hat{\mu}_n(X_t)$, $\mu^o_{nt} = \mu^o_n(X_t)$,and $\mu^o_t = \mu_o(X_t)$. (i) We verify the conditions of Theorem 2.3(i). A.1 is ensured by D.1; A.2-A.3 and B.1 are either imposed directly or ensured (in A.2Γwe take $(\Theta, \rho_\Theta) = (\mathcal{W}^s_{\infty,r}, \rho_{s,\infty})$ and $\hat{\theta}_n = \hat{\mu}_n$ as in (3.2)). When $\{m_t\}$ is linear in θ, B.2 holds trivially with $\lambda = \infty$ for any pseudo-metric ρ_Θ; otherwise B.2(aIb) are assumedΓand B.2(c) holds because

$$\begin{aligned}\rho_{s,\infty}(\hat{\mu}_n, \mu_o) &\leq \rho_{s,\infty}(\hat{\mu}_n, \mu^o_n) + \rho_{s,\infty}(\mu^o_n, \mu_o)\\ &\leq p_n^{1/2} \sup_{1\leq j\leq p_n} ||\psi_j||_{\infty,s}||\hat{\beta}_n - \beta^o_n|| + \rho_{s,\infty}(\mu^o_n, \mu_o)\\ &\leq p_n^{1/2} B_s(p_n) O_P(p_n^{1/2}/\lambda^{1/2}_{\min}(\Psi'_n\Psi_n)) + \rho_{s,\infty}(\mu^o_n; \mu_o)\\ &= o(n^{-1/(2(1+\lambda))})\end{aligned}$$

given D.3(c)ΓD.4(b)Γ$n^{1/2(1+\lambda)} p_n B_s(p_n)/\lambda_{\min}(\Psi'_n\Psi_n) \to 0$ and Proposition 3.2.

NextΓwe verify B.3. B.3(a) is imposed directly; given D.5Γwe have

$$\delta\bar{m}^o_n(\hat{\theta}_n - \theta_o; \hat{\pi}_n) = n^{-1}\textstyle\sum_t g_{nt}(X_t, \hat{\pi}_n)(\hat{\mu}_{nt} - \mu^o_t) = n^{-1}\sum_t g_{nt}(X_t, \pi_o)(\hat{\mu}_{nt} - \mu^o_t) + r_n,$$

where $r_n = n^{-1}\sum_t(g_{nt}(X_t, \hat{\pi}_n) - g_{nt}(X_t, \pi_o))(\hat{\mu}_{nt} - \hat{\mu}^o_t) = o_P(n^{-1/2})$ by the Cauchy-Schwarz inequalityΓProposition 3.2 and D.5(bIc). HenceΓB.3(b) holds. B.3(c) also holds because

$$\begin{aligned}n^{-1}\textstyle\sum_t g_{nt}(X_t, \pi_o)(\hat{\mu}_{nt} - \mu^o_t) &= n^{-1}\textstyle\sum_t g_{nt}(X_t, \pi_o)(\hat{\mu}_{nt} - \mu^o_{nt}) + n^{-1}\sum_t g_{nt}(X_t, \pi_o)(\mu^o_{nt} - \mu^o_t)\\ &= [n^{-1}\textstyle\sum_t g_{nt}(X_t, \pi_o)\psi'_{np_n}(X_t)](\hat{\beta}_n - \beta^o_n) + o_P(n^{-1/2})\\ &= n^{-1}\textstyle\sum_t v_{nt}(Z_t) + o_P(n^{-1/2})\end{aligned}$$

by Proposition 3.3 and using the fact that $\|n^{-1}\sum_t g_{nt}(X_t, \pi_o)(\mu^o_{nt} - \mu^o_t)\| = o_P(n^{-1/2})$ given D.4(a) and D.5(a) by the Cauchy-Schwarz inequality. FinallyΓB.4-B.6 are assumed directly. It follows from Theorem 2.3(i) that $M_n \to^d \chi^2_q$.

(ii) Consistency follows immediately from Theorem 2.3(ii). ■

Proof of Theorem 3.5: We apply Theorem 3.4. (i) First we verify D.1-D.5. Both D.1-D.2 hold given (a) and (d). Given (a) and (c)Γ we have $\sum_t \psi_i(X_t)\psi_j(X_t) = n\delta_{ij}$, where $\delta_{ii} = 1$ and $\delta_{ij} = 0$, $i \neq j$. ThereforeΓ $\lambda_{\min}(\Psi'_n\Psi_n) = n$Γso D.3(a) holds. Since $\max_j \sup_{x\in[0,1]} |\psi_j(x)| \leq \sqrt{2}$, D.3(c) (for $s = 0$) holds with $B_0(p_n) = \sqrt{2}$ for all p_n. Because $\xi_{np_n} = \sup_t(\psi'_{np_n}(X_t)(\Psi'_n\Psi_n)^-\psi_{np_n}(X_t)) \leq 2p_n/n$, D.3(b) holds given $p_n^4/n \to 0$.

NextΓ we verify D.4-D.6. Given (b)Γ there exists $\mu^o_n(\cdot) = \psi'_{np_n}(\cdot)\beta^o_n \in \mathcal{C}^r(0,1)$ such that $\rho_\infty(\mu^o_n, \mu_o) = o(p_n^{-r+\delta})$ for all $\delta > 0$ (e.g.ΓEdmunds and Moscatelli [1977]). It follows that D.4 with $s = 0$ hold given (e). AlsoΓ with $B_0(p_n) = \sqrt{2}$ and $\lambda = 1$, the condition $n^{1/2(1+\lambda)} p_n B_0(p_n)/\lambda^{1/2}_{\min}(\Psi'_n\Psi_n) = O(p_n/n^{1/4}) \to 0$ holds given (e). From (A14) belowΓwe see that D.5(a) holds with $g_{nt}(X_t, \pi) = (-3\sigma^2_o, 0)'$; D.5(bIc) are null because π does not appear.

We now verify the remaining conditions. We put $\mu_t = \mu(X_t)$, $\hat{\mu}_{nt} = \hat{\mu}_n(X_t)$, $\mu^o_{nt} = \mu^o_n(X_t)$. A.2 holds with $(\hat{\alpha}_n, \alpha_o) = (\hat{\sigma}^2_n, \sigma^2_o)$, $(\Theta, \rho_\Theta) = (\mathcal{C}^r(0,1), \rho_\infty)$, and $\hat{\theta}_n = \hat{\mu}_n$ as in (3.2). A.3 is null since π does not appear. B.1 holds given $\{m_t\}$. Next we verify B.2(aIb). For $\mu \in \Theta_o = \{\mu \in \mathcal{C}^r(0,1) : \rho_\infty(\mu; \mu_o) \leq \Delta\}$, we have $|(Y_t - \mu_t)^3 - \epsilon^3_t + 3\epsilon^2_t(\mu_t - \mu^o_t)| \leq (3|\epsilon_t| + \Delta)(\mu_t - \mu^o_t)^2$ and $|(Y_t - \mu_t)^4 - \epsilon^4_t + 4\epsilon^3_t(\mu_t - \mu^o_t)| \leq (8\epsilon^2_t + 3\Delta^2)(\mu_t - \mu^o_t)^2$. It follows that

$$\left| m_t(\alpha_o, \theta, \pi) - (\epsilon^3_t, \epsilon^4_t - 3\sigma^4_o)' + (3\epsilon^2_t, 4\epsilon^3_t)'(\mu_t - \mu^o_t) \right| \leq D_{nt}(Z_t)\rho^2_\infty(\mu, \mu_o) \tag{A14}$$

for all $\mu \in \Theta_o$, where $D_{nt}(Z_t) = 8\epsilon^2_t + 3|\epsilon_t| + 3\Delta^2 + \Delta$. ThereforeΓB.2(aIb) with $\lambda = 1$ hold. From (A14) we have $\delta\hat{m}^o_n(\theta - \theta_o; \pi) = -n^{-1}\sum^n_{t=1}(3\epsilon^2_t, 4\epsilon^3_t)'(\mu_t - \mu^o_t)$, and $\delta\bar{m}^o_n(\theta - \theta_o; \pi) = (-3\sigma^2_o, 0)'n^{-1}\sum_t(\mu_t - \mu^o_t)$. HenceΓB.3(a) holds because

$$\begin{aligned} \delta\hat{m}^o_n(\hat{\theta}_n - \theta_o; \pi) - \delta\bar{m}^o_n(\hat{\theta}_n - \theta_o; \pi) &= -n^{-1}\sum_t(3(\epsilon^2_t - \sigma^2_o), 4\epsilon^3_t)'(\hat{\mu}_{nt} - \mu^o_t) \\ &= -n^{-1}\sum_t(3(\epsilon^2_t - \sigma^2_o), 4\epsilon^3_t)'(\hat{\mu}_{nt} - \mu^o_{nt}) \\ &\quad -n^{-1}\sum_t(3(\epsilon^2_t - \sigma^2_o), 4\epsilon^3_t)'(\mu^o_{nt} - \mu^o_t) \\ &= o_P(n^{-1/2}), \end{aligned} \tag{A15}$$

where for the first term $\|n^{-1}\sum_t(3(\epsilon^2_t - \sigma^2_o), 4\epsilon^3_t)'(\hat{\mu}_{nt} - \mu^o_{nt})\| \leq ||n^{-1}\sum_t(3(\epsilon^2_t - \sigma^2_o), 4\epsilon^3_t)'\psi'_{np_n}(X_t)||$ $||\hat{\beta}_n - \beta^o_n|| = O_P(p_n/n^{1/2})O_P(p^{1/2}_n/\lambda^{1/2}_{\min}(\Psi'_n\Psi_n)) = o_P(n^{-1/2})$ by Chebyshev's inequality and Proposition 3.2; and for the second term $n^{-1}\sum_t(3(\epsilon^2_t - \sigma^2_o), 4\epsilon^3_t)(\mu^o_{nt} - \mu^o_t) = O_P(n^{-1/2}\rho_\infty(\mu^o_n; \mu_o)) = o_P(n^{-1/2})$ by Chebyshev's inequality. SimilarlyΓwe can show that $\hat{\sigma}^2_n - \sigma^2_o = n^{-1}\sum^n_{t=1}(\epsilon^2_t - \sigma^2_o) + o_P(n^{-1/2})$, so B.4 with $s_{nt} = \epsilon^2_t - \sigma^2_o$ holds. By definitionΓ$G^o_{np_n} = n^{-1}\sum_t \psi_{np_n}(X_t)(-3\sigma^2_o, 0) = ((-\frac{3}{\sqrt{2}}\sigma^2_o, 0, ..., 0), (0, 0, ..., 0))'$. It follows that $v_{nt}(Z_t) = G^{o\prime}_{np_n}(\Psi'_n\Psi_n/n)^-\psi_n(X_t)\varphi(\epsilon_t)/E\varphi'(\epsilon_t) = (-3\sigma^2_o\eta_t, 0)'$, where $\eta_t = \varphi(\epsilon_t)/E\varphi'(\epsilon_t)$. ThereforeΓwe have $W_{nt} = (\epsilon^3_t - 3\sigma^2_o\eta_t, (\epsilon^4_t - 3\sigma^4_o) - 6\sigma^2_o(\epsilon^2_t - \sigma^2_o))'$. By the Lindeberg-Levy CLTΓ$J_o^{-1/2}n^{1/2}W_n \to^d N(0, I_2)$, where

$$J_o = \begin{bmatrix} J_{11} & J_{12} \\ J_{12} & J_{22} \end{bmatrix},$$

with $J_{11} = 15\sigma^6_o - 6\sigma^2_o E(\epsilon^3_t\eta_t) + 9\sigma^4_o E(\eta^2_t)$, $J_{22} = 24\sigma^8_o$ and $J_{12} = -3\sigma^2_o E(\epsilon^4_t\eta_t) + 18\sigma^4_o E(\epsilon^4_t\eta_t)$. HenceΓB.5 holds. B.6 also holds given $\hat{J}_n$ by straightforward verification using appropriate weak ULLN's and $\rho_\infty(\hat{\mu}_n; \mu_o) \to^p 0$ by Proposition 3.2. The desired result then follows from Theorem 3.4(i). (ii) Consistency follows immediately from Theorem 3.4(ii). ■

Proof of Theorem 3.6: Given $\delta_n = n^{-\lambda_1}$, we choose a subset $\{\pi^1_n, ..., \pi^{\#G_n}_n\}$ from Π_n such that for each $\pi \in \Pi_n$Γthere exists at least one π^j_n such that $\rho_\infty(\pi, \pi^j_n) < \delta_n$, where

$G_n = G_n(\delta_n)$ is a finite open covering of Π_n of cardinality of $\#G_n$. This cardinality is finite as Π (and hence Π_n) has finite metric entropy. For arbitrary $\epsilon > 0$,

$$\begin{aligned}&P\left[\sup_{\pi\in\Pi_n} |Q_n(\pi)| > \epsilon\right]\\ &\le P\left[\max_{1\le i\le \#G_n} \sup_{\pi\in\mathbb{B}_n(\pi_n^i,\delta_n)} |Q_n(\pi)| > \epsilon\right]\\ &\le P\left[\max_{1\le i\le \#G_n} |Q_n(\pi_n^i)| > \epsilon/2\right] + P\left[\max_{1\le i\le \#G_n} \sup_{\pi\in\mathbb{B}_n(\pi_n^i,\delta_n)} |Q_n(\pi) - Q_n(\pi_n^i)| > \epsilon/2\right]\\ &\le P\left[\max_{1\le i\le \#G_n} |Q_n(\pi_n^i)| > \epsilon/2\right] + P\left[\sup_{\pi'\in\Pi_n} \sup_{\pi\in\mathbb{B}_n(\pi',\delta_n)} |Q_n(\pi) - Q_n(\pi')| > \epsilon/2\right],\end{aligned}$$

where $\mathbb{B}_n(\pi_n^i, \delta_n) = \{\pi \in \Pi_n : \rho_\infty(\pi; \pi_n^i) < \delta_n\}$. For the first term

$$\begin{aligned}P\left[\max_{1\le i\le \#G_n} |Q_n(\pi_n^i)| > \epsilon/2\right] &\le \textstyle\sum_{i=1}^{\#G_n} P\left[|Q_n(\pi_n^i)| > \epsilon/2\right]\\ &\le \#G_n \max_{1\le i\le \#G_n} P\left[|Q_n(\pi_n^i)| > \epsilon/2\right]\\ &\le \#G_n \sup_{\pi\in\Pi_n} P\left[|Q_n(\pi)| > \epsilon/2\right]\\ &\le \#G_n C_1 n^{-\lambda}\end{aligned}$$

given (a). This holds in particular for $\lambda_2 = \lambda + d\lambda_1$. NextΓfor the second term

$$P[\sup_{\pi'\in\Pi_n} \sup_{\pi\in\mathbb{B}_n(\pi',\delta_n)} |Q_n(\pi) - Q_n(\pi')|] > \epsilon/2 \le 2\epsilon^{-1} C n^{-\lambda}$$

by Markov's inequalityΓgiven (b). ThereforeΓ

$$P[\sup_{\pi\in\Pi_n} |Q_n(\pi)| > \epsilon] \le \#G_n C_1 n^{-(\lambda+d\lambda_1)} + 2\epsilon^{-1} C n^{-\lambda}.$$

Because $\Pi_n \subseteq \Pi$, $\#G_n \le \#G_n(\delta_n)$ for any $\delta_n > 0$, where $\#G(\delta_n)$ is the metric entropy of Π. Given Π and $\delta_n = n^{-\lambda_1}$, we have from Kolmogorov and Tihomirov [1961Γ Section 2.3] that $\#G(\delta_n) = \Delta n^{d\lambda_1}$. (Kolmogorov and Tihomirov prove this only for $d = 1$, but the proof for $d > 1$ follows analogously.) Substituting this into the above expressionΓwe obtain $P\left[\sup_{\pi\in\Pi_n} |Q_n(\pi)| > \epsilon\right] \le C_2 n^{-\lambda}$ for some C_2. This completes the proof. ■

Proof of Theorem 3.7: Put $\zeta_{nt} = (\Psi_n'\Psi_n)^{-1/2}\psi_{np_n}(X_t)$, $\hat\sigma_{nt} = \hat\sigma_n(X_t)$, $\sigma_t = \sigma(X_t)$ and $\sigma_t^o = \sigma_o(X_t)$. Then $W_{nts} = \epsilon_t\zeta_{nt}'\zeta_{ns}\epsilon_s\sigma_t^o\sigma_s^o/\sigma_t\sigma_s$. The proof consists of showing: (i) $\sum_t(W_{nt}(\hat\sigma_n) - W_{nt}(\sigma_o)) = o_P(p_n^{1/2})$; (ii) $\sum\sum_{t\ne s}(W_{nt}(\hat\sigma_n) - W_{nt}(\sigma_o)) = o_P(p_n^{1/2})$; and (iii) $(\sum_t\sum_s W_{nts}(\sigma_o) - p_n)/(2p_n)^{1/2} \to^d N(0,1)$.

We first consider (i). Given E.2 and the identity $\sum_t \zeta_{nt}'\zeta_{nt} = p_n$, we have $|\sum_t W_{ntt}(\hat\sigma_n) - W_{ntt}(\sigma_o)| \le C\rho_\infty(\hat\sigma_n, \sigma_o)\sum_t \epsilon_t^2\zeta_{nt}'\zeta_{nt} = o_P(p_n^{1/2})$ by Markov's inequality.

NextΓwe apply Theorem 3.6 to show (ii). Choose $\Pi_n = \{\sigma \in \sum : \rho_\infty(\sigma, \sigma_o) \le n^{-\delta}\}$ for some $0 < \delta < 1$. Note that $\hat\sigma_n \in \Pi_n$ in probability. Let $Q_n(\sigma) = \sum\sum_{t\ne s}(W_{nt}(\hat\sigma_n) - W_{nt}(\sigma_o)) = \sum_{t=2}^n Q_{nt}$, where $Q_{nt} = 2\sum_{s=1}^{t-1}(W_{nts}(\sigma) - W_{nts}(\sigma_o))$.

To show (ii)Γit suffices to show $\sup_{\sigma\in\Pi_n} |Q_n(\sigma)| = o_P(p_n^{1/2})$. Given D.1(aΓb′,c)Γ$\sum_t \zeta_{nt}'\zeta_{nt} = p_n$ and $p_n/n \to 0$ (as implied by E.3), we have $E[\sup_{(\sigma_1,\sigma_2)\in\mathbb{B}_n(n^{-\lambda_1})} |Q_n(\sigma_1) - Q_n(\sigma_2)|] \le$

$c^{-1}n^{-\lambda_1}\sum\sum_{t\neq s}\zeta_{nt}'\zeta_{nt} \leq c^{-1}n^{-\lambda_1+1}p_n \leq c^{-1}n^{-\lambda_1+2}$. Hence, condition (b) of Theorem 3.6 holds by choosing $\lambda_1 \geq \lambda+2$. Next, we verify condition (a). Since $E(Q_{nt}|\epsilon_1,\epsilon_2,...,\epsilon_{t-1})=0$ given $\{X_t\}$ nonstochastic, $\{Q_{nt},\mathcal{F}_{t-1}\}$ is a martingale difference sequence, where $\{\mathcal{F}_t\}$ is the sequence of σ-fields consisting of $\epsilon_s, s \leq t$. By Hölder's inequality and Rosenthal's inequality (see Hall and Heyde [1980, p.23], or Hall [1989] for its application), we have that for $k=1,2,...,$

$$E\left(Q_n^{2k}(\sigma)\right) \leq \left\{\sum_{t=2}^{n}\left(EQ_{nt}^{2k}(\sigma)\right)^{1/k}\right\}^k.$$

Conditional on ϵ_t, $Q_{nt}(\sigma)$ is sum of independent random variables, and so given D.1(c) and $\sum_t \zeta_{nt}\zeta_{nt}'=I_n$, $EQ_{nt}^{2k}(\sigma) \leq c(k)^{-1}\rho_\infty^{2k}(\sigma,\sigma_o)\left\{\sum_{s=1}^{t-1}(\zeta_{nt}'\zeta_{ns})^2\right\}^k \leq c(k)^{-1}\rho_\infty^{2k}(\sigma,\sigma_o)(\zeta_{nt}'\zeta_{nt})^k$. It follows that $EQ_n^{2k}(\sigma) \leq c(k)^{-1}\rho_\infty^{2k}(\sigma,\sigma_o)\left\{\sum_{t=2}^{n}\zeta_{nt}'\zeta_{nt}\right\}^k \leq c(k)^{-1}\rho_\infty^{2k}(\sigma,\sigma_o)p_n^k$. Therefore, by Markov's inequality, we have for any $\eta>0$ and for $k=1,2,...,$

$$P\left[|Q_n(\sigma)|>\eta p_n^{1/2}\right] \leq EQ_n^{2k}(\sigma)/(p_n^k\eta^{2k}) = c(k)^{-1}\eta^{-2k}\rho_\infty^{2k}(\sigma,\sigma_o) \leq c(k)^{-1}\eta^{-2k}n^{-\delta k}.$$

It follows from Theorem 3.6 that $\sup_{\sigma\in\Sigma_n}|Q_n(\sigma)|=o(p_n^{-1/2})$ $a.s.$ Thus, (ii) is proved. Finally, the proof of (iii) follows exactly that of Hong and White [1995, Theorem A.1]. ■

Proof of Theorem 3.8: We use the following notations: $f_t^o=f(X_t,\alpha_o)$, $\nabla_\alpha f_t^o=\nabla_\alpha f(X_t,\alpha_o)$, $\mu_t^o=\mu_o(X_t)$, $\theta_t=\theta(X_t)$, $\theta_t^o=\theta_o(X_t)$, $\hat{\theta}_{nt}=\hat{\theta}_n(X_t)$, $\theta_{nt}^o=\theta_n^o(X_t), \sigma_t=\sigma(X_t), \sigma_t^o=\sigma_o(X_t)$, $\hat{\sigma}_{nt}=\hat{\sigma}_n(X_t)$. Under H_o^* we have $m_t(\gamma_o)=(\theta_t^o-f_t^o/\sigma_t^o)(Y_t-f_t^o)/\sigma_t^o=0$, $\nabla_\alpha\bar{m}_n(\gamma_o)=-E[\nabla_\alpha f_t^o(Y_t-f_t^o)/(\sigma_t^o)^2+(\theta_t^o-f_t^o/\sigma_t^o)/\sigma_t^o]=0$ and $\delta\bar{m}_n^o(\theta-\theta_o;\pi_o)=E[(\theta_t-\theta_t^o)(Y_t-f_t^o)/\sigma_t^o]=0$ for $\theta\in\mathcal{W}_{\infty,r}(\mathbb{X})$. It follows that (3.3) is a_n-degenerate at γ_o under H_o^* for any given sequence a_n. Hence, Theorem 2.5 is relevant. (i) We first consider asymptotic normality: A.1 is ensured by D.1(a,b′,c); A.2 holds with $(\Theta,\rho_\Theta)=(\mathcal{W}_{\infty,r}(\mathbb{X}),\rho_\infty)$ and $(\Pi,\rho_\Pi)=(\Sigma,\rho_\infty)$. A.3(a) with $a_n=n/p_n^{1/2}$ holds because

$$\begin{aligned} a_n(m_n(\alpha_o,\theta_o,\pi)-\bar{m}_n(\alpha_o,\theta_o,\pi)) &= a_n n^{-1}\textstyle\sum_{t=1}^{n}(\theta_t^o-f_t^o/\sigma_t^o)(Y_t-f_t^o)/\sigma_t^o \\ &= a_n n^{-1}\textstyle\sum_{t=1}^{n}\mu_t^o\epsilon_t(1/\sigma_t\sigma_t^o-\sigma_t^2) \\ &= o_P(1) \end{aligned}$$

given D.1(a,b′,c) and E.2 by applying Theorem 3.6 (following the analogous reasoning of (ii) in the proof of Theorem 3.7); A.3(b) also holds trivially since $\bar{m}_n^o(\alpha_o,\theta_o,\pi)=0$ for all $\pi=\sigma\in\Sigma$. C.1(a) holds given D.1(a,b′,c) and E.1; C.1(b) holds since

$$\begin{aligned} \nabla_\alpha m_n(\alpha_o,\hat{\theta}_n,\hat{\pi}_n)-\nabla_\alpha\bar{m}_n(\alpha_o,\hat{\theta}_n,\hat{\pi}_n) &= -n^{-1}\textstyle\sum_{t=1}^{n}(\nabla_\alpha f_t^o/\hat{\sigma}_{nt})(Y_t-f_t^o)/\hat{\sigma}_{nt} \\ &= -n^{-1}\textstyle\sum_{t=1}^{n}\nabla_\alpha f_t^o\epsilon_t/\sigma_t^o \\ &\quad -n^{-1}\textstyle\sum_{t=1}^{n}\nabla_\alpha f_t^o\sigma_t^o\epsilon_t(\hat{\sigma}_{nt}^{-2}-\sigma_t^{o-2}) \\ &= O_P(n^{-1/2}) \\ &= o_P(n^{1/2}/a_n). \end{aligned}$$

AboveΓthe first term is $O_P(n^{-1/2})$ by Chebyshev's inequality and the second term is $o_P(n^{-1/2})$ given E.2 by Theorem 3.6 following the analogous reasoning of (ii) of the proof of Theorem 3.7. C.1(c) holds under H_o^* since

$$\begin{aligned}
&\nabla'_\alpha \bar{m}_n^o(\alpha_o, \hat{\theta}_n, \hat{\pi}_n) - \nabla'_\alpha \bar{m}_n^o(\gamma_o)\\
&= -n^{-1}\textstyle\sum_{t=1}^n \nabla'_\alpha f_t^o(\hat{\theta}_{nt} - f_t^o/\hat{\sigma}_{nt})/\hat{\sigma}_{nt}\\
&= n^{-1}\textstyle\sum_t (\nabla'_\alpha f_t^o/\hat{\sigma}_{nt})(\hat{\theta}_{nt} - \theta_t^o) + n^{-1}\sum_{t=1}^n \nabla'_\alpha f_t^o \mu_t^o(1/\hat{\sigma}_{nt}\sigma_t^o - 1/\hat{\sigma}_{nt}^2)\\
&= n^{-1}\textstyle\sum_t (\nabla'_\alpha f_t^o/\sigma_t^o)(\hat{\theta}_{nt} - \theta_t^o) + n^{-1}\sum_{t=1}^n \nabla'_\alpha f_t^o(1/\hat{\sigma}_{nt} - 1/\sigma_t^o)(\hat{\theta}_{nt} - \theta_t^o)\\
&\quad + n^{-1}\textstyle\sum_t \nabla'_\alpha f_t^o \mu_t^o(1/\hat{\sigma}_{nt}\sigma_t^o - 1/\hat{\sigma}_{nt}^2)\\
&= o_P(p_n^{1/2}/n^{1/2}),
\end{aligned}$$

where for the first term (a weighted average of $\hat{\theta}_n - \theta_o$) we have $n^{-1}\sum_t(\nabla'_\alpha f_t^o/\sigma_t^o)(\hat{\theta}_{nt} - \theta_t^o) = o_P(p_n^{1/2}/n^{1/2})$ by straightforward but tedious algebraΓgiven D.1(a)ΓE.2 and E.4. AlsoΓthe last two terms are $o_P(p_n^{1/2}/n^{1/2})$ by the Cauchy-Schwarz inequality given E.2 and E.4.

We now verify the remaining conditions. Since (3.3) is linear in θ, C.2 with $\lambda = \infty$ holds. C.3(a) also holds trivially since $\delta^2 m_{nt}(\theta - \theta_o; \alpha_o, \theta_o, \pi) = 0$; for C.3(b)Γ

$$\begin{aligned}
&\delta \hat{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n)\\
&= n^{-1}\textstyle\sum_t (\hat{\theta}_n - \theta_t^o)\epsilon_t \sigma_t^o/\hat{\sigma}_{nt}\\
&= n^{-1}\textstyle\sum_t (\hat{\theta}_n - \theta_{nt}^o)\epsilon_t \sigma_t^o/\hat{\sigma}_{nt} + n^{-1}\sum_t(\theta_{nt}^o - \theta_t^o)\epsilon_t\sigma_t^o/\hat{\sigma}_{nt}\\
&= n^{-1}\textstyle\sum_t\sum_s (\sigma_t^o/\hat{\sigma}_{nt})\epsilon_t \zeta'_{nt}\zeta_{ns}\epsilon_s(\sigma_s^o/\hat{\sigma}_{ns}) + n^{-1}\sum_t\sum_s(\sigma_t^o/\hat{\sigma}_{nt})\epsilon_t\zeta'_{nt}\zeta_{ns}(\mu_t^o/\hat{\sigma}_{ns} - \theta_{nt}^o)\\
&\quad + n^{-1}\textstyle\sum_t(\theta_t - \theta_t^o)\epsilon_t\sigma_t^o/\hat{\sigma}_{nt}\\
&= n^{-1}\textstyle\sum_t\sum_s \epsilon_t\zeta'_{nt}\zeta_{ns}\epsilon_s + o_P(p_n^{1/2}/n)
\end{aligned}$$

by Theorem 3.6Γgiven E.2 and E.4. AlsoΓC.3(d) holds since $\delta\bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi) = 0$ for all $\pi \in \Pi$; C.4 is given directly; and C.5 with $J_n^o = 2$ holds by Theorem 3.7. FinallyΓC.6 holds with $\hat{J}_n = 2$ and $\hat{R}_n = R_n^o = p_n/n$. The result now follows from Theorem 2.5(i).

(ii) Consistency follows immediately from Theorem 2.5(ii). ■

Proof of Lemma 4.1: See Appendix B of Newey [1994]. Note that the proof for $\rho_\infty(\hat{p}, p_o) = O_P([nb_n^d/\ln(n)]^{-1/2} + b_n^k)$ follows analogously to that of Newey with $y_{ni} = 1$. ■

Proof of Proposition 4.2: Given F.4(a) and $\hat{\theta}_n = \hat{r}_n/\hat{p}_n$Γwe have

$$\begin{aligned}
\delta\bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi_o) &= E[g(X,\pi_o)(\hat{\theta}_n(X) - \theta_o(X))]\\
&= \int_{\mathbb{X}} g(x,\pi_o)[\hat{r}_n(x) - \theta_o(x)\hat{p}_n(x)]dx\\
&\quad + \int_{\mathbb{X}} g(x,\pi_o)[\hat{r}_n(x) - \theta_o(x)\hat{p}_n(x)][p_o(x)/\hat{p}_n(x) - 1]dx \qquad \text{(A16)}
\end{aligned}$$

For the first term, we can write

$$
\begin{aligned}
\int_{\mathbb{X}} g(x,\pi_o)[\hat{r}_n(x)-\theta_o(x)\hat{p}_n(x)]dx &= \int_{\mathbb{X}} g(x,\pi_o)[n^{-1}\sum_{t=1}^{n}(Y_t-\theta_o(x))K_n(X_t-x)]dx \\
&= n^{-1}\sum_t g(X_t,\pi_o)\epsilon_t \\
&\quad +n^{-1}\sum_t \epsilon_t[\int_{\mathbb{X}} g(x,\pi_o)K_n(X_t-x)dx-g(X_t,\pi_o)] \\
&\quad + n^{-1}\sum_t \int_{\mathbb{X}} g(x,\pi)(\theta_o(X_t)-\theta_o(x))K_n(X_t-x)dx \\
&= n^{-1}\sum_t g(X_t,\pi_o)\epsilon_t + o_P(n^{-1/2})+O_P(b_n^k) \\
&= n^{-1}\sum_t g(X_t,\pi_o)\epsilon_t + o_P(n^{-1/2}) \qquad \text{(A17)}
\end{aligned}
$$

given $nb_n^{2k}\to 0,$ where the second term is $o_P(n^{-1/2})$ by Chebyshev's inequality and the fact that $\int_{\mathbb{X}} g(x,\pi_o)K_n(X_t-x)dx - g(X_t,\pi_o)=o(1)$ uniformly in t given F.1(b), F.2 and F.4(a). Also, the second term is $O(b_n^k)$ by Markov's inequality and the fact that $\int_{\mathbb{X}} g(x,\pi_o)(\theta_o(x')-\theta_o(x))K_n(x'-x)dx = O(b_n^k)$ uniformly in $x'\in\mathbb{X}$ given F.1(b) and F.2-F.4(a).

Next, we consider the last term of (A16). By the Cauchy-Schwarz inequality, we obtain

$$
\begin{aligned}
&\left|\int_{\mathbb{X}} g(x,\pi_o)[\hat{r}_n(x)-\theta_o(x)\hat{p}_n(x)][p_o(x)/\hat{p}_n(x)-1]dx\right| \\
\leq\ & \sup_{x\in\mathbb{X}}|p_o(x)/\hat{p}_n(x)-1|\,(\int_{\mathbb{X}}\|g(x,\pi_o)\|^2\,dx)^{1/2}(\int_{\mathbb{X}}[\hat{r}_n(x)-\theta_o(x)\hat{p}_n(x)]^2dx)^{1/2} \\
=\ & O_P(\ln^{1/2}(n)(nb_n^d)^{-1/2}+b_n^k)O_P((nb_n^d)^{-1/2}+b_n^k) \\
=\ & o_P(n^{-1/2}) \qquad \text{(A18)}
\end{aligned}
$$

given $nb_n^{2d}/\ln(n)\to\infty$, $nb_n^{2k}\to 0$, and $2k>d$, where we have made use of the fact that $\sup_{x\in\mathbb{X}}|p_o(x)/\hat{p}_n(x)-1| = O_P((nb_n^d/\ln(n))^{-1/2}+b_n^k)$ by Lemma 4.1, and $\int_{\mathbb{X}}[\hat{r}_n(x)-\theta_o(x)\hat{p}_n(x)]^2dx = O_P((nb_n^d)^{-1}+b_n^{2k})$ by Markov's inequality given F.1-F.3. Combining (A16)-(A18) yields $n^{1/2}\delta\bar{m}_n^o(\hat{\theta}_n-\theta_o;\gamma_o)=n^{-1/2}\sum_t g(X_t,\pi_o)\epsilon_t+o_P(1)$. Because $V_o=E[g(X,\pi_o)g(X,\pi_o)'\epsilon^2]$ is $O(1)$ and nonsingular, $n^{-1/2}\sum_t g(X_t,\pi_o)\epsilon_t \to^d N(0,V_o)$ by the Lindeberg-Levy CLT. ■

Proof of Theorem 4.3: (i) We verify the conditions of Theorem 2.3(i). A.1 is ensured by F.1; A.2 holds with $(\Theta,\rho_\Theta)=(\mathcal{W}^s_{\infty,r}(\mathbb{X}),\rho_\infty)$ and $\hat{\theta}_n=\hat{\mu}_n$ as in (4.1); A.3 and B.1 are imposed directly. When (a) $m(Z_t,\alpha,\theta,\pi)$ is linear in θ, B.2 with $\lambda=\infty$ holds for any norm ρ_Θ; or when (b) B.2(a,b) hold, B.2(c) is ensured by Lemma 4.1 given $n^{\delta/(2+\delta)}b_n^d\ln n(n)\to\infty$, $n^\lambda b_n^{(d+2s)(1+\lambda)}/\ln^{1+\lambda}(n)\to\infty$ and $nb_n^{2k(1+\lambda)}\to 0$. B.3(a) is assumed directly; given F.4, B.3(b) holds because

$$\begin{aligned}
\delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \hat{\pi}_n) &= E[g(X, \hat{\pi}_n)(\hat{\theta}_n(X) - \theta_o(X))] \\
&= E[g(X, \pi_o)(\hat{\theta}_n(X) - \theta_o(X))] + E[(g(X, \hat{\pi}_n) - g(X, \pi_o))(\hat{\theta}_n(X) - \theta_o(X))] \\
&= E[g(X, \pi_o)(\hat{\theta}_n(X) - \theta_o(X))] + o_P(n^{-1/2}),
\end{aligned}$$

where the second term is $o_P(n^{-1/2})$ by the Cauchy-Schwarz inequality, F.4(b,c) and $E(\hat{\theta}_n(X) - \theta_o(X))^2 = O_P((nb_n^d)^{-1} + b_n^{2k})$. By Proposition 4.2, B.3(c) holds with $v_{nt}(Z_t) = g(X_t, \pi_o)\epsilon_t$. B.4 is given directly; B.5 holds by the Lindeberg-Levy CLT given F.5. Finally, B.6 is imposed directly. All conditions of Theorem 2.3(i) are satisfied, so the desired result follows.

(ii) Consistency follows immediately from Theorem 2.3(ii). ■

Proof of Theorem 4.4: Put $\hat{\mu}_{1t} = \hat{\mu}_{1n}(X_{1t})$, $\mu_1^o(X_{1t}) = \mu_{1t}^o$, $\hat{p}_{1t} = \hat{p}_{1n}(X_{1t})$, $p_{1t}^o = p_1^o(X_{1t})$, and $\hat{r}_{1t} = \hat{\mu}_{1t}\hat{p}_{1t}$. (i) We apply Theorem 4.3(i.a) as $\{m_t\}$ is linear in $\theta = \mu_1$. Given that α and π do not appear, A.2 with $(\Theta, \rho_\Theta) = (\mathcal{W}_{\infty,r}(\mathbb{X}), \rho_2)$ holds, where $\rho_2(\theta_1, \theta_2) = (\int_{\mathbb{X}} (\theta_2(x) - \theta_1(x))^2 p_o(x)dx)^{1/2}$. A.3 and B.1 are null since $\mathbb{A}$ and Π are null. The proofs of B.3(a) and B.6 are deferred to the end.

Next, we verify F.1-F.5. F.1 with $\delta = 2$ holds given (a); F.2 with $s = 0$ holds given (b); F.3 (for μ_1^o and p_1^o) holds given (d); F.4(a) holds with $g(X_{1t}, \pi) = g(X_{1t})$ since, with p_o the joint density of X_t,

$$\delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi_o) = E(\psi(X_{2t})(\hat{\mu}_{1t} - \mu_{1t}^o)) = E(g(X_{1t})(\hat{\mu}_{1t} - \mu_{1t}^o)),$$

where g satisfies conditions in F.4(a) given (e); F.4(b,c) are null. F.5 holds with $W(Z_t, \pi_o) = (\psi(X_{2t}) - g(X_{1t}))\epsilon_{1t}$ given (a) and (b), where $\epsilon_{1t} = Y_t - \mu_{1t}^o$. Finally, $nb_n^{2d_1}/\ln^2(n) \to \infty$, $nb_n^{2k} \to 0$, $2k > d_1$ are imposed directly. All conditions of Theorem 4.3(i.a) are satisfied. Hence, $M_n \to^d \chi_1^2$, provided B.3(a) and B.6 hold.

It remains to prove B.3(a) and B.6. For B.3(a), note that

$$\begin{aligned}
\delta m_n^o(\hat{\theta}_n - \theta_o; \pi) &= n^{-1}\textstyle\sum_t \psi(X_{2t})(\hat{\mu}_{1t} - \mu_{1t}^o) \\
&= n^{-1}\textstyle\sum_t \psi(X_{2t})(\hat{r}_{1t} - \mu_{1t}^o\hat{p}_{1t})p_{1t}^{o-1} + \; n^{-1}\textstyle\sum_t \psi(X_{2t})(\hat{r}_{1t} - \mu_{1t}^o\hat{p}_{1t})(\hat{p}_{1t}^{-1} - p_{1t}^{o-1}) \\
&= n^{-1}\textstyle\sum_t \psi(X_{2t})(\hat{r}_{1t} - \mu_{1t}^o\hat{p}_{1t})p_{1t}^{o-1} + o_P(n^{-1/2}),
\end{aligned}$$

where the second term is $o_P(n^{-1/2})$ given the conditions on p_1^o, μ_1^o and ψ, by following reasoning analogous to (A18). Similarly, we can also obtain

$$\delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi) = \int_{\mathbb{X}_1} g(x_1)(\hat{r}_{1n}(x_1) - \mu_1^o(x_1)\hat{p}_{1n}(x_1))dx_1 + o_P(n^{-1/2}).$$

Thus, to show $\delta m_n^o(\hat{\theta}_n - \theta_o; \pi) - \delta \bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi) = o_P(n^{-1/2})$, it suffices to show

$$\begin{aligned} A_n &= n^{-1} \sum_t \psi(X_{2t})(\hat{r}_{1t} - \mu_{1t}^o \hat{p}_{1t}) p_{1t}^{o-1} \\ &= \int_{\mathbb{X}_1} g(x_1)(\hat{r}_n(x_1) - \mu_1^o(x_1)\hat{p}_{1n}(x_1))dx_1 + o_P(n^{-1/2}). \end{aligned} \tag{A19}$$

For this purpose, we write

$$\begin{aligned} A_n &= n^{-2} \textstyle\sum_t \sum_s \psi(X_{2t})(Y_s - \mu_{1t}^o) p_{1t}^{o-1} K_n(X_{1t} - X_{1s}) \\ &= n^{-2} \textstyle\sum\sum_{t \neq s} \psi(X_{2t})(Y_s - \mu_{1t}^o) p_{1t}^{o-1} K_n(X_{1t} - X_{1s}) + h_n \\ &= U_n + o_P(n^{-1/2}), \end{aligned}$$

where $h_n = b_n^{-1} K(0) n^{-2} \sum_t \psi(X_{2t}) \epsilon_{1t} p_{1t}^{o-1} = O_P(n^{-3/2} b_{1n}^{-1}) = o_P(n^{-1/2})$ by Chebyshev's inequality given conditions (a)-(d). We now consider U_n.

Define $U_{nts} = [\psi(X_{2t})(Y_s - \mu_{1t}^o)p_{1t}^{o-1} + \psi(X_{2s})(Y_t - \mu_{1s}^o)p_{1s}^{o-1}]K_n(X_{1t} - X_{1s})$. Then $U_n = n^{-2}\sum\sum_{s<t} U_{nts} = (1-n^{-1})EU_n + n^{-2}\sum\sum_{s<t}\hat{U}_{nts}$, where $\hat{U}_{nts} = U_{nts} - EU_n$, $EU_n = EU_{nts}$. Given $nb_n^{2d_1}\ln^2(n) \to \infty$, we have $E\hat{U}_{nts}^2 = O(b_n^{-d_1}) = o(n)$. Hence, by the extended U-statistic projection theorem of Powell *et al* [1989, Lemma 3.1], we have $U_n = (1-n^{-1})[EU_n + n^{-1}\sum_t E(\hat{U}_{nts}|Z_t)] + o_P(n^{-1/2})$, where

$$\begin{aligned} E(\hat{U}_{nts}|Z_t) &= \textstyle\int_{\mathbb{X}} \psi(x_2)(Y_t - \mu_1^o(x_1))p_1^{o-1}(x_1)K_n(X_{1t} - x_1)p_o(x)dx \\ &\quad + \psi(X_{2t})p_{1t}^{o-1}\textstyle\int_{\mathbb{X}}(\mu_1^o(x_1) - \mu_1^o(X_{1t}))K_n(X_{1t} - x_1)p_o(x)dx - EU_n \\ &= \textstyle\int_{\mathbb{X}_1} g(x_1)(Y_t - \mu_1^o(x_1))p_1^{o-1}(x_1)K_n(X_{1t} - x_1)p_1^o(x_1)dx_1 - EU_n + O(b_n^k) \end{aligned}$$

given the conditions on μ_1^o and K. It follows that

$$U_n = EU_n + n^{-1}\textstyle\sum_t \int_{\mathbb{X}_1} g(x_1)(Y_t - \mu_1^o(x_1))K_n(X_{1t} - x_1)dx_1 + o_P(n^{-1/2})$$

given $nb_n^{2k} \to 0$. Therefore, (A19) holds. Thus, B.3(a) holds. Finally, by Lemma 4.1, we have $\rho_\infty(\hat{\mu}_{1n}, \mu_1^o) \to^p 0$ and $\rho_\infty(\hat{g}_n, g) \to^p 0$ given conditions (a)-(f). Hence, it is straightforward to show $\hat{J}_n \to^p J_o = E((\psi(X_{2t}) - g(X_{1t}))^2\epsilon_{1t}^2)$, so B.6 holds.

(ii) Consistency follows immediately from Theorem 4.3(ii). ■

Proof of Proposition 4.5: Because $E(W_{nts}|X_t) = E(W_{nts}|X_s) = 0, t \neq s$, given F.1(a), we have $EW_n = (nb_n^d)^{-1}K(0)\sigma_o^2$, where $\sigma_o^2 = E(\epsilon_t^2)$. Hence, $W_n - EW_n = n^{-2}\sum\sum_{s<t} 2W_{nts} + (nb_n^d)^{-1}K(0)n^{-1}\sum_t(\epsilon_t^2 - \sigma_o^2) = n^{-2}\sum\sum_{s<t}2W_{nts} + O_P(n^{-3/2}b_n^{-d})$ by Chebyshev's inequality given F.2(a)-(c) with $\delta = 2$. Put $a_n = nb_n^{d/2}$. Then $a_n(W_n - EW_n) = n^{-1}b_n^{d/2}\sum\sum_{s<t}2W_{nts} + o_P(1)$ given $nb_n^d \to \infty$. Therefore, to show $a_n(W_n - EW_n) \to^d N(0,1)$, it suffices to show $U_n \to^d N(0,1)$, where $U_n = \sum\sum_{s<t}U_{nts}$, $U_{nts} = 2n^{-1}b_n^{d/2}W_{nts}$. Because $E(U_{nts}|Z_t) =$

$E(U_{nts}|Z_s) = 0$ for $t \neq s$, U_n is a degenerate second order U-statistic. de Jong's [1987] CLT for generalized forms then applies. By de Jong [1987ΓProposition 3.2]Γit suffices for $J_n^{-1/2}U_n \to^d N(0,1)$ that $G_{ni}/J_n^2 = o(1)$ for $i = 1, 2, 4$, where G_{ni} and J_n are defined as follows. Put $K_n^{ts} = K_n(X_t - X_s)$. ThenΓby change of variableΓit is straightforward to compute that

$$\begin{aligned} J_n = var(U_n) &= \sum\sum_{s<t} EU_{nts}^2 \\ &= (1-n^{-1})2b_n^d E[\epsilon_1^2\epsilon_2^2(K_n^{12})^2] \\ &= 2C(K)E(\sigma_o^4(X)p_o(X))(1+o(1)) \\ &= J_o(K)(1+o(1)); \end{aligned}$$

$$\begin{aligned} G_{n1} = \sum\sum_{s<t} EU_{nts}^4 &\leq 8n^{-2}b_n^{2d}E[\epsilon_1^4\epsilon_2^4(K_n^{12})^4] \\ &\leq \Delta^2 n^{-2}b_n^{2d}E(K_n^{12})^4 \\ &= \Delta^2 n^{-2}b_n^{-d}(\textstyle\int_T K^4(v)dv)E(p_o(X))(1+o(1)); \end{aligned}$$

$$\begin{aligned} G_{n2} &= \sum\sum\sum_{s<t<j} E[U_{nts}^2U_{ntj}^2] \\ &\leq 16n^{-1}b_n^{2d}E[\epsilon_1^2\epsilon_2^2(K_n^{12})^2\epsilon_1^2\epsilon_3^2(K_n^{13})^2] \\ &\leq \Delta^2 n^{-1}b_n^{2d}E[(K_n^{12})^2(K_n^{13})^2] \\ &= \Delta^2 n^{-1}C^2(K)E(p_o^2(X))(1+o(1)); \end{aligned}$$

$$\begin{aligned} G_{n4} &= \sum\sum\sum\sum_{i<j<s<t}[E(U_{nij}U_{nis}U_{ntj}U_{nts}) + E(U_{nij}U_{nit}U_{nsj}U_{nst}) + E(U_{nis}U_{nit}U_{njs}U_{njt})] \\ &\leq b_n^{2d}E(\epsilon_1^2\epsilon_2^2\epsilon_3^2\epsilon_4^2K_n^{12}K_n^{13}K_n^{42}K_n^{43}) \\ &\leq \Delta^2 b_n^{2d}E(K_n^{12}K_n^{13}K_n^{42}K_n^{43}) \\ &= \Delta^2 b_n^d(\textstyle\int_T K(v)K(v+w)K(w)dvdw)E(p_o^3(X))(1+o(1)). \end{aligned}$$

It follows that $G_{ni}/J_n^2 = o(1)$ for $i = 1, 2, 4$ given $nb_n^d \to \infty$, $b_n \to 0$. HenceΓ$J_o(K)^{-1/2}U_n \to^d N(0,1)$, and therefore $J_o(K)^{-1/2}nb_n^{d/2}(W_n - EW_n) \to^d N(0,1)$. ■

Proof of Proposition 4.6: Put $\hat{v}_{nt} = \mu_o(X_t) - \hat{\mu}_n(X_t)$ and $\hat{S}_n = 2n^{-2}\sum\sum_{s<t}\hat{\epsilon}_{nt}^2\hat{\epsilon}_{ns}^2K_n^{ts}$, where $K_n^{ts} = K_n(X_t - X_s)$. Then $\hat{J}_n = 2C(K)\hat{S}_n$. Straightforward but tedious algebra delivers that

$$\begin{aligned} \hat{S}_n &= 2n^{-2}\sum\sum_{s<t}\epsilon_t^2\epsilon_t^2K_n^{ts} + n^{-2}\sum\sum_{s<t}(8\epsilon_t^2\epsilon_s\hat{v}_{ns} + 4\epsilon_t^2\hat{v}_{ns}^2 + 8\epsilon_t\epsilon_s\hat{v}_{nt}\hat{v}_{ns} + 8\epsilon_t\hat{v}_{nt}\hat{v}_{ns}^2 + 2\hat{v}_{nt}^2\hat{v}_{ns}^2)K_n^{ts} \\ &= \tilde{S}_n + 8A_{1n} + 4A_{2n} + 8A_{3n} + 8A_{4n} + 2A_{5n}, \text{ say} \\ &= \tilde{S}_n + o_P(1), \end{aligned}$$

where $A_{jn} = o_P(1)$, $j = 1, ..., 5$, by straightforward but tedious algebra. For exampleΓ $A_{2n} \leq \Delta b_n^{-d}(n^{-1}\sum_t \epsilon_t^2)(n^{-1}\sum_s \hat{v}_{ns}^2) = O_P(n^{-1}b_n^{-2d} + b_n^{2k-d}) = o_P(1)$ given boundedness of K, $nb_n^{3d} \to \infty$, $b_n \to 0$ and $2k > d$, where we have also made use of $n^{-1}\sum_s \hat{v}_{ns}^2 = O_P(n^{-1}b_n^{-d} + b_n^{2k})$.

NextΓwe show $\tilde{S}_n = E(\sigma_o^4(X)p_o(X)) + o_P(1)$. Since $E[\epsilon_1^2\epsilon_2^2(K_n^{12})^2] = O(b_n^{-d}) = o(n)$ given $nb_n^d \to \infty$, it follows by Powell *et al* [1989ΓLemma 3.1] that

$$\tilde{S}_n \ = S_n^o + 2n^{-1}\textstyle\sum_{t=1}^n[\epsilon_t^2 \int_{\mathbb{X}} \sigma_o^2(x)K_n(X_t - x)p_o(x)dx - S_n^o] + o_P(n^{-1/2}),$$

where $S_n^o = b_n^d E[\epsilon_1^2\epsilon_2^2 K_n^{12}]$. FurthermoreΓby Chebyshev's inequalityΓ$\sum_{t=1}^n[\epsilon_t^2 \int_{\mathbb{X}} \sigma_o^2(x)$ $K_n(X_t - x)p_o(x)dx - S_n^o] = O_P(n^{-1/2})$ given F.1 (with $\delta = 2$) and F.2. It follows that $\tilde{S}_n = S_n^o + O_P(n^{-1/2})$; on the other handΓ$S_n^o = E(\sigma_o^4(X_t)p_o(X_t)) + o(1)$ by continuity of σ_o^2 and p_o. Since $\hat{J}_n = 2C(K)\hat{S}_n$, we have $\hat{J}_n = J_o + o_P(1)$. ■

Proof of Theorem 4.7: We first verify that (4.3) is degenerate under H_o^* : noting $f(X_t, \alpha_o) = \mu_o(X_t)$ and $\epsilon_t = Y_t - f(X_t, \alpha_o)$, we have $m_t(\alpha_o, \theta_o, \pi) = (r_o(X_t) - p_o(X)f(X_t, \alpha_o))\epsilon_t = 0$ *a.s*, $\nabla'_\alpha \bar{m}_n(\alpha_o, \theta_o, \pi) = E[-\nabla'_\alpha f(X_t, \alpha_o)(r_o(X_t) - p_o(X_t)f(X_t, \alpha_o) + \epsilon_t)] = 0$, and $\delta\bar{m}_n^o(\theta - \theta_o; \pi) = E\{[(r(X_t) - r_o(X_t)) - (p(X_t) - p_o(X_t))f(X_t, \alpha_o)]\epsilon_t\} = 0$ for all $\theta = (r, p) \in \Theta = \mathcal{W}_{\infty,r}^r(\mathbb{X}) \times \mathcal{W}_{\infty,r}^r(\mathbb{X})$. It follows by Definition 2.4 that (4.3) is a_n-degenerate at γ_o under H_o^* for any given sequence a_n. Theorem 2.5 is applicable. (i) We first show asymptotic normality; A.1 is ensured by F.1; A.2 holds with $(\Theta, \rho_\Theta) = (\mathcal{W}_{\infty,r}^r(\mathbb{X}) \times \mathcal{W}_{\infty,r}^r(\mathbb{X}), \rho)$, where $\rho(\theta, \theta') = \rho_\infty(\mu, \mu') + \rho_\infty(p, p')$. A.3 is null because π does not appear. Given F.6Γ(4.3) is twice differentiable *a.s.* on A, with $\|\nabla_\alpha^2 m_t(\alpha, \theta, \pi)\| = \|[r(X_t) + p(X_t)Y_t - 2p(X_t)f(X_t, \alpha)] - 2p(X_t)\nabla_\alpha f(X_t, \alpha)\nabla'_\alpha f(X_t, \alpha)\|$ dominated by some integrable function. HenceΓC.1(a) holds; C.1(b–c) are null since Π is null. NextΓbecause

$$\begin{aligned} m_t(\alpha_o, \theta, \pi) &= m_t(\alpha_o, \theta, \pi) + \delta m_t(\theta - \theta_o; \alpha_o, \theta_o, \pi) \\ &= m_t(\alpha_o, \theta_o, \pi) + [(r(X_t) - r_o(X_t)) - (p(X_t) - p_o(X_t))f(X_t, \alpha_o)]\epsilon_t, \end{aligned}$$

C.2(a) with $\lambda = \infty$ holds; and C.2(b) holds trivially. C.3(a) is null because $\delta^2 m_t(\alpha, \theta, \pi) = 0$. In additionΓsince $\delta\bar{m}_n^o(\theta - \theta_o; \pi) = 0$ for all $\theta \in \Theta$ under H_o^*, we have

$$\begin{aligned} &\delta\hat{m}_n^o(\hat{\theta}_n - \theta_o; \pi) - \delta\bar{m}_n^o(\hat{\theta}_n - \theta_o; \pi) + \delta^2\bar{m}_n^o(\theta - \theta_o; \pi) \\ &\quad = n^{-1}\textstyle\sum_t[(\hat{r}_n(X_t) - r_o(X_t)) - (\hat{p}_n(X_t) - p_o(X_t))f(X_t, \alpha_o)]\epsilon_t. \end{aligned}$$

Put $\hat{\mu}_{nt} = \hat{\mu}_n(X_t)$, $\mu_t^o = \mu_o(X_t)$ and $K_n^{ts} = K_n(X_t - X_s)$. Substituting expressions for $\hat{r}_n$ and $\hat{p}_n$Γwe obtain

$$\begin{aligned} \delta\hat{m}_n^o(\hat{\theta}_n - \theta_o; \pi) \ &= n^{-2}\textstyle\sum_t\sum_s \epsilon_t(Y_s - \mu_t^o)K_n^{ts} \\ &= n^{-2}\textstyle\sum_t\sum_s \epsilon_t\epsilon_s K_n^{ts} + n^{-2}\sum_t\sum_s \epsilon_t(\mu_s^o - \mu_t^o)K_n^{ts} \\ &= W_n + O_P(n^{-1}b_n^{1-d/2} + n^{-1/2}b_n^k) \\ &= n^{-2}\textstyle\sum_t\sum_s W_{nts} + o_P(n^{-1}b_n^{-d/2}). \end{aligned}$$

given $nb_n^{2k+d} \to 0$, where $W_{nts} = \epsilon_t \epsilon_s K_n^{ts}$, and we have made use of

$$n^{-2} \sum_t \sum_s \epsilon_t (\mu_s^o - \mu_t^o) K_n^{ts} = O_P(n^{-1} b_n^{1-d/2} + n^{-1/2} b_n^k), \tag{A20}$$

as is shown at the end of this proof. ThereforeΓC.3(b) holds for $a_n = nb_n^{d/2}$; C.3(c–d) are null since Π is null. C.4 is given directly. C.5 holds with $J_n^o = J_o = 2C(K)E(\sigma_o^4(X_t)p_o(X_t))$ by Theorem 4.5. FinallyΓC.6(a) is ensured by Proposition 4.6; and C.6(b) with $R_n = n^{-1} b_n^{-d} K(0) \sigma_o^2$ holds because $\hat{R}_n = n^{-1} b_n^{-d} K(0) \hat{\sigma}_n^2$, and

$$\begin{aligned}
\hat{\sigma}_n^2 - \sigma_o^2 &= n^{-1} \textstyle\sum_t (\epsilon_t^2 - \sigma_o^2) + 2n^{-1} \sum_t \epsilon_t (\hat{\mu}_{nt} - \mu_t^o) + n^{-1} \sum_t (\hat{\mu}_{nt} - \mu_t^o)^2 \\
&= O_P(n^{-1/2}) + O_P(n^{-1/2} b_n^{-d/2} + b_n^k) + O_P(n^{-1} b_n^{-d} + b_n^{2k}) \\
&= o_P(b_n^{d/2})
\end{aligned}$$

given $nb_n^{3d} \to \infty$, $nb_n^{2k+d} \to 0$ and $2k > d$. ThereforeΓ$M_n \to^d N(0,1)$ by Theorem 2.5(i).

It remains to show (A20). Define $U_{nts} = (\epsilon_t(\mu_s^o - \mu_t^o) + \epsilon_s(\mu_t^o - \mu_s^o))K_n^{ts}$, and put $\hat{U}_{nts} = (U_{nts} - U_{nt} - U_{ns}), U_{nt} = E(U_{nts}|Z_s)$. Then we can write

$$\begin{aligned}
U_n &= n^{-2} \textstyle\sum_t \sum_s \epsilon_t (\mu_s^o - \mu_t^o) K_n^{ts} \\
&= n^{-2} \textstyle\sum \sum_{t<s} (\epsilon_t(\mu_s^o - \mu_t^o) + \epsilon_s(\mu_t^o - \mu_s^o)) K_n^{ts} \\
&= n^{-2} \textstyle\sum \sum_{t<s} \hat{U}_{nts} + 2(1 - n^{-1}) n^{-1} \sum_t U_{nt}.
\end{aligned}$$

Because $E(\hat{U}_{nts}|Z_t) = E(\hat{U}_{nts}|Z_s) = 0$ and $E\hat{U}_{nts}^2 \leq 2EU_{nts}^2 = O(b_n^{2-d})$, we have $var(\sum \sum_{s<t} \hat{U}_{nts}) = \sum \sum_{s<t} E\hat{U}_{nts}^2 = O(n^2 b_n^{2-d})$. It follows by Chebyshev's inequality that $n^{-2} \sum \sum_{t<s} \hat{U}_{nts} = O_P(n^{-1} b_n^{1-d/2})$. NextΓnoting that $U_{nt} = \epsilon_t \int_{\mathbb{X}} (\mu_o(x) - \mu_o(X_t)) K_n(X_t - x) p_o(x) dx$ and $E(U_{nt}^2) = O(b_n^{2k})$ given F.1-F.3Γwe have $n^{-1} \sum_t U_{nt} = O_P(n^{-1/2} b_n^k)$ by Chebyshev's inequality. it follows that $U_n = O_P(n^{-1} b_n^{1-d/2} + n^{-1/2} b_n^k)$. This completes the proof for asymptotic normality.

(ii) Consistency follows immediately from Theorem 2.5(ii). ■

REFERENCES

Andrews, D.W.K. [1991a]: "An Empirical Process Central Limit Theorem for Dependent Nonidentically Distributed Random Variables"Γ*Journal of Multivariate Analysis* 38Γ187-203.

Andrews, D.W.K. [1991b]: "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models"Γ*Econometrica* 59Γ307-345.

Andrews, D.W.K. [1994]: "Asymptotics for Semiparametric Econometrics Models via Stochastic Continuity"Γ*Econometrica* 62Γ43-72.

Bahadur, R.R. [1960]: "Stochastic Comparison of Tests"Γ*Annals of Mathematical Statistics* 31Γ276-295.

Bera, A.K. and C.M. Jarque [1982]: "Model Misspecification Tests: A Simultaneous Approach"Γ*Journal of Econometrics* 20Γ59-82.

Bierens, H.J. [1990]: "A Consistent Conditional Moment Test for Functional Forms"Γ *Econometrica* 58Γ1443-1458.

Billingsley, P. [1986]: *Probability and Measure, 2nd ed.* New York: John Wiley.

De Jong, P. [1987]: "A Central Limit Theorem for Generalized Quadratic Forms"Γ*Probability Theory and Related Fields* 75Γ261-277.

Durbin, J. [1954]: "Errors in Variables"Γ*Review of the International Statistical Institute* 22Γ23-32.

Edmunds, D.E. and V.B. Moscatelli [1977]: "Fourier Approximation and Embeddings of Sobolev Space"Γ*Dissertationes Mathematicae*ΓWarsaw: Polish Scientific Publishers.

Fan, Y. and R. Gencay [1993]: "Hypothesis Testing Based on Modified Nonparametric Estimation of An Affine Measure Between Two Distributions"Γ*Journal of Nonparametric Statistics* 2Γ389-403.

Gallant, A.R. [1981]: "Unbiased Determination of Production Technologies"Γ*Journal of Econometrics* 30Γ149-169.

Gallant, A. R. and G. Souza [1991]: "On the Asymptotic Normality of Fourier Flexible Form Estimates"Γ*Journal of Econometrics* 50Γ329-353.

Granger, C.W.J. [1990]: "General Introduction"Γin C.W.J. Granger ed.Γ*Modelling Economic Time Series.* Oxford: Oxford University PressΓpp. 1-23.

Geweke, J. [1981a]: "The Approximate Slopes of Econometric Tests"Γ*Econometrica* 49Γ 1427-1442.

Geweke, J. [1981b]: "A Comparison of Tests of Independence of Two Covariance Stationary Time Series"Γ*Journal of the American Statistical Association* 76Γ363-373.

Goldstein, L.G. and K. Messer [1992]: "Optimal Plug-in Estimators for Nonparametric Functional Estimation"Γ*Annals of Statistics* 30Γ1306-1328.

Hall, P. [1984]: "Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators"Γ*Journal of Multivariate Analysis* 14Γ1-16.

Hall, P. [1988]: "Estimating the Direction in Which a Data Set is Most Interesting"Γ*Probability Theory and Related Fields* 80Γ51-77.

Hall, P. [1989]: "On Polynomial-Based Projection Indices for Exploratory Projection Pursuit"Γ*Annals of Statistics* 17Γ589-605.

Hall, P. and C. Heyde [1980]: *Martingale Limit Theory and Its Application.* Academic Press: New York.

Härdle, W. and T. M. Stoker [1989]: "Investigating Smooth Multiple Regression by the Method of Average Derivatives"Γ*Journal of the American Statistical Association* 84Γ 986-995.

Hausman, J.A. [1978]: "Specification Tests in Econometrics"Γ*Econometrica* 46Γ1251-1272.

Hong, Y. and H. White [1991]: "Consistent Specification Testing via Nonparametric Series Regression"ΓUniversity of CaliforniaΓSan DiegoΓDepartment of Economics Discussion Paper.

Hong, Y. and H. White [1993]: "Consistent Nonparametric Entropy-Based Testing"Γ University of CaliforniaΓSan DiegoΓDepartment of Economics Discussion Paper.

Hong, Y. and H. White [1995]: "Consistent Specification Testing via Nonparametric Series Regression"Γ*Econometrica* 63Γ1133-1159.

Kolmogorov, A. N. and V. M. Tihomirov [1961]: "ϵ-Entropy and ϵ-Capacity of Sets in Functional Spaces"Γ*American Mathematical Society Translations* 2Γ17Γ277-364.

Lavergne, P. and Q. Vuong [1996]: "Nonparametric Selection of Regressors: The Nonnested Case"Γ*Econometrica* 64Γ207-219.

Lee, B. J. [1988]: "A Nonparametric Model Specification Test Using a Kernel Regression Method"ΓUniversity of WisconsinΓDepartment of Economics Doctoral Dissertation.

Mack, Y. P. and B.W. Silverman [1982]: "Weak and Strong Uniform Consistency of Kernel Regression Estimates"Γ*Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 61Γ405-415.

Mammen, E. [1989]: "Asymptotics with Increasing Dimension for Robust Regression with Application to the Bootstrap"Γ*Annals of Statistics* 17Γ382-400.

Mikosch, T. [1991]: "Functional Limit Theorems for Random Quadratic Forms"Γ*Stochastic Processes and Their Applications* 37Γ81-98.

Newey, W. K. [1985]: "Maximum Likelihood Specification Testing and Conditional Moment Tests"Γ*Econometrica* 53Γ1047-1070.

Newey, W. K. [1991]: "Uniform Convergence in Probability and Stochastic Equicontinuity"Γ*Econometrica* 59Γ1161-1167.

Newey, W. K. [1994]: "Kernel Estimation of Partial Means and a General Variance Estimator"Γ*Econometric Theory* 10Γ233-253.

Powell, J., J. Stock and T. Stoker [1989]: "Semiparametric Estimation of Index Coefficients"Γ*Econometrica* 57Γ1403-1430.

Robinson, P. M. [1987]: "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form"Γ*Econometrica* 55Γ875-891.

Robinson, P. M. [1988]: "Root-N-Consistent Semiparametric Regression"Γ*Econometrica* 56Γ931-954.

Robinson, P. M. [1989]: "Hypothesis Testing in Nonparametric and Semiparametric Models for Econometric Time Series"Γ*Review of Economic Studies* 56Γ511-534.

Robinson, P. M. [1991]: "Consistent Nonparametric Entropy-Based Testing"Γ*Review of Economic Studies* 58Γ437-453.

Rotar, I. V. [1973]: "Some Limit Theorems for Polynomials of Second Degree"Γ*Theory of Probability and Its Applications* 18Γ499-507.

Stinchcombe, M. and H. White [1998]: "Consistent Specification Testing With Nuisance Parameters Present Only Under the Alternative"Γ*Econometric Theory* 14Γ295-325.

Stoker, T. M. [1989]: "Tests of Additive Derivative Constraints"Γ*Review of Economic Studies* 56Γ535-552.

Tauchen, G. [1985]: "Diagnostic Testing and Evaluation of Maximum Likelihood Models"Γ *Journal of Econometrics* 30Γ415-443.

Whang, Y.-J. and D. W. K. Andrews [1993]: "Tests of Specification for Parametric and Semiparametric Models"Γ*Journal of Econometrics* 57Γ277-318.

White, H. [1982]: "Maximum Likelihood Estimation of Misspecified Models"Γ*Econometrica* 50Γ1-25.

White, H. [1984]: *Asymptotic Theory for Econometricians*ΓSan Diego: Academic Press.

White, H. [1987]: "Specification Testing in Dynamic Models"Γin T.F. Bewley ed.Γ*Advances in Econometrics, Fifth World Congress,* Vol I. New York: Cambridge University PressΓpp.1-58.

White, H. [1994]: *Estimation, Inference and Specification Analysis.* New York: Cambridge University Press.

White, H. and J. Wooldridge [1991]: "Some Results on Sieve Estimation with Dependent Observations"Γin W. BarnettΓJ. PowellΓand G. Tauchen (eds.)Γ*Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics.* New York: Cambridge University

PressΓpp. 459-493.

Whittle, P. [1964]: "On the Convergence to Normality of Quadratic Forms in Independent Variables"Γ*Theory of Probability and Its Applications* 9Γ113-118.

Wooldridge, J. [1992]: "A Test for Functional Form Against Nonparametric Alternatives"Γ *Econometric Theory* 8Γ452-475.

Wu, D.-M. [1973]: "Alternative Tests of Independence Between Stochastic Regressors and Disturbances"Γ*Econometrica* 41Γ733-750.

Yatchew, A. J. [1992]: "Nonparametric Regression Tests Based on an Infinite Dimensional Least Squares Procedure"Γ*Econometric Theory* 8Γ435-451.

Yohai, V. J. and R.A. Maronna [1979]: "Asymptotic Behavior of M-Estimators for the Linear Model"Γ*Annals of Statistics* 7Γ258-268.