COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

Cowles Foundation Discussion Paper No. 968

TESTS OF SPECIFICATION FOR PARAMETRIC AND SEMIPARAMETRIC MODELS

by

Yoon–Jae Whang
and
Donald W. K. Andrews

January 1991

ABSTRACT

This paper provides a general framework for constructing specification tests for parametric and semiparametric models. The paper develops new specification tests using the general framework. In particular, specification tests for semiparametric partially linear regression, sample selection, and censored regression models are introduced. The results apply in time series and cross-sectional contexts. The method of proof exploits results concerning the stochastic equicontinuity or weak convergence of normalized sums of stochastic processes.

JEL Classification Number: 211.

Keywords: Infinite dimensional nuisance parameter, semiparametric model, specification test, stochastic equicontinuity.

1. INTRODUCTION

Semiparametric models and estimation procedures have become increasingly popular in the econometrics literature because they relax some of the parametric assumptions often used in econometric modelling. Semiparametric models, however, still are not robust to a variety of specification errors. Existing specification tests for parametric models are not applicable in such models and there have been few attempts made in the literature to develop new specification tests for semiparametric models.

The purpose of this paper is to provide a general framework for constructing specification tests for parametric and semiparametric models. We use this framework to develop new specification tests for parametric and semiparametric models and to generalize existing specification tests. The general framework considered here covers cases in which the null model is parametric or semiparametric and the alternative model is parametric, semiparametric, or nonparametric. Examples are given below.

Our strategy in constructing the general framework is to extend existing results in the parametric econometrics literature. We note that the work of Newey (1985a, b) and Tauchen (1985) (also see White (1987)), hereafter NT, has provided a single unifying framework for specification tests in parametric models. This framework embeds many available results as special cases. For example, Lagrange Multiplier (LM) tests, Cox's (1961, 1962) test of non–nested hypotheses, White's (1982) information matrix test, and Newey's (1985a) conditional moment tests all fit into the NT framework.

The basic idea of the NT approach is as follows: Consider a vector–valued criterion function that is indexed by a (finite dimensional) parameter. The criterion function is chosen so that its expected value evaluated at the true parameter equals a zero vector if the model is correctly specified and is not necessarily zero if the model is misspecified. Then, a test statistic is constructed by forming a quadratic form in the sample moments of the criterion function evaluated at an estimator of the parameter.

The assumption of finite dimensionality of the parameter space in the NT framework is too restrictive for our purposes. In our general framework, we explicitly introduce infinite dimensional nuisance parameters in the criterion function. This allows us to consider tests of specification of semiparametric models against semiparametric and nonparametric alternatives. It also allows us to consider tests of specification of parametric models against semiparametric alternatives. In addition, we do not make specific distributional assumptions on the underlying random variables (rv's) in the general framework. Therefore, the general framework developed here can be utilized for developing new specification tests for parametric and semiparametric models and for extending the domain of applicability of existing results, e.g., to dependent non–identically distributed (dnid) rv cases.

The asymptotic distribution of the general form of the test statistic is derived under a set of "high–level" assumptions as in Andrews (1990a, b). In particular, we take as basic assumptions certain properties, including consistency, of the infinite dimensional nuisance parameter estimator and the fulfillment of a uniform law of large numbers (ULLN), a central limit theorem (CLT), and a stochastic equicontinuity condition for certain rv's and/or stochastic processes. Advantages and disadvantages of adopting the "high–level" assumptions are discussed in detail in Andrews (1990a). Andrews (1990b) provides primitive conditions under which the ULLN and stochastic equicontinuity conditions hold. It also provides primitive conditions under which kernel estimators of the infinite dimensional nuisance parameters satisfy the requisite properties. In addition, primitive conditions for a CLT are widely available in the literature. Thus, primitive conditions are available to replace each of the high–level assumptions in any given example. We note, however, that the mobilization of these primitive conditions does require work, with the amount of work depending on the context. In some of the examples given below, we present more primitive assumptions than the high–level assumptions referred to above.

We now specify examples that fit into the general framework discussed above. Those marked with an asterisk are discussed in the paper. (For a discussion of those without an asterisk, see Whang and Andrews (1990).)

(1)* Tests of omitted variables, heteroskedasticity, and autocorrelation in partially linear regression models.

(2)* Tests of omitted variables and heteroskedasticity in semiparametric sample selection models.

(3)* A test of the parametric linear model against the semiparametric partially linear regression model.

(4)* Tests of distributional assumptions in parametric censored regression and sample selection models.

(5)* A test of the semiparametric partially linear regression model against a nonparametric regression model.

(6) Tests of omitted variables and heteroskedasticity in semiparametric binary choice models.

(7) A test of distributional assumptions in parametric binary choice models.

(8) Tests of normality (standardized $\sqrt{b_1}$ and b_2 tests) in nonlinear regression models.

(9) Chi–square diagnostic tests.

The remainder of this paper is organized as follows: Section 2 defines the general form of our test statistic, derives its limit distribution under the null hypothesis and local alternatives, and establishes consistency results for it. Section 3 discusses examples in which both the null and alternative models are semiparametric, viz., examples (1)–(2) above. Section 4 discusses examples in which the null model is parametric and the alternative model is semiparametric, viz., examples (3) and (4) above. Section 5 discusses an example in which the null and alternative are semiparametric and nonparametric, respectively, viz., example (5) above. An Appendix contains proofs of the results given in Sections 2–5.

Throughout the paper all limits are taken as the sample size, T, goes to infinity, unless specified otherwise. We let "with probability $\to 1$" abbreviate "with probability that goes to one as $T \to \infty$." We let $\|A\|$ denote the Euclidean norm of a vector or matrix A, i.e., $\|A\| = (\text{trace}(A'A))^{1/2}$. For notational simplicity, we let Σ_a^b denote $\sum_{t=a}^{b}$ and $E\|X\|^a$ denote $E(\|X\|)^a$.

2. GENERAL FRAMEWORK

In this section, we define the general form of the specification test statistic and give sufficient conditions to obtain its limiting distribution under both the null hypothesis of correct specification and local alternative hypotheses of misspecification. Consistency properties of the test are also derived.

2.1. Some Preliminaries

The data are given by a triangular array of rv's $\{W_{Tt}\} = \{W_{Tt} : t = 1, \ldots, T; T \geq 1\}$ defined on some probability space $(\Omega, \mathcal{B}, P)$. In the case where W_{Tt} does not depend on T, we write it as W_t.

We consider a criterion function $r_{Tt}(\cdot,\cdot,\cdot) : R^{k_{Tt}} \times B \times \Pi \to R^m$ that satisfies

$$Er_{Tt}(W_{Tt}, \beta_0, \pi_0) = \underset{\sim}{0} \ \text{ for } \ t = 1, \ldots, T; \ T \geq 1 \tag{2.1.1}$$

when the model is correctly specified, where k_{Tt} is a positive integer $\leq \infty$, $\beta_0 \in B \subset R^q$, $\pi_0 \in \Pi$, and Π is a pseudo–metric space with pseudo–metric defined below. Let $\hat{\beta}$ and $\hat{\pi}$ be estimators of β_0 and π_0 respectively. Note that $\hat{\pi}$ can be an infinite dimensional estimator. It is assumed that $\hat{\pi}$ is a random element of Π with probability $\to 1$.

Our test statistic is constructed as follows. Let

$$\bar{r}_T(\hat{\beta},\hat{\pi}) = \tfrac{1}{T}\Sigma_1^T r_{Tt}(W_{Tt}, \hat{\beta}, \hat{\pi}) \, . \tag{2.1.2}$$

Heuristically, we expect the above sample average (2.1.2) to be close to zero if the model is correctly specified. In order for $\bar{r}_T(\hat{\beta},\hat{\pi})$ to serve as a useful indicator of misspecification of the model, however, it should not be close to zero if the model is misspecified.

Below we provide sufficient conditions under which

$$\sqrt{T}\, \bar{r}_T(\hat{\beta},\hat{\pi}) \xrightarrow{d} N(\underset{\sim}{0}, \Phi) \tag{2.1.3}$$

if the model is correctly specified, where Φ is an $m \times m$ positive definite matrix. Let $\hat{\Phi}$ be an estimator of Φ that is consistent under the null hypothesis of correct specification. We define a general form of specification test statistic as follows:

$$G_T = T\bar{r}_T(\hat{\beta},\hat{\pi})'\hat{\Phi}^{-1}\bar{r}_T(\hat{\beta},\hat{\pi}) \, . \tag{2.1.4}$$

The test statistic G_T is shown to have a limit χ^2 distribution with m degrees of freedom under the null hypothesis of correct specification. We also show that G_T has a noncentral χ^2 distribution with m degrees of freedom asymptotically under a sequence of local alternative hypotheses of misspecification.

Throughout this paper, all functions that are introduced (such as $\hat{\beta}$, $\hat{\pi}$, $r_{Tt}(\cdot,\cdot,\cdot)$) are assumed to be $\mathcal{B}$/Borel or Borel/Borel measurable. The only exceptions are the stochastic processes $\nu_T(\cdot)$ and $\nu_T(\cdot,\cdot)$ defined below, which need not be measurable. Thus, we assume away measurability problems except in those circumstances where measurability may be of real concern.

Below we let $r_t(\beta,\pi)$ denote $r_{Tt}(W_{Tt}, \beta, \pi)$ for notational simplicity.

2.2. Asymptotic Properties of the Test Statistic under Correct Specification

In this section, we give sufficient conditions for the asymptotic normality of $\sqrt{T}\, \bar{r}_T(\hat{\beta},\hat{\pi})$ (see equation (2.1.3)). Consistent estimation of Φ is discussed and the asymptotic distribution of the test statistic G_T under correct specification is derived.

Three alternative sets of sufficient conditions are introduced – Assumptions 1, 1*, and 1**. Each is sufficient for the asymptotic normality of $\sqrt{T}\,\bar{r}_T(\hat{\beta},\hat{\pi})$. Each involves a different tradeoff in the assumptions it imposes. In particular, Assumptions 1 and 1* allow π to be infinite dimensional, whereas Assumption 1** requires it to be finite dimensional. Also, Assumptions 1 and 1** assume $r_t(\beta,\pi)$ is differentiable in β, whereas Assumption 1* only requires $Er_t(\beta,\pi)$ to be differentiable in β. For example, Assumption 1* allows one to consider specification tests using least absolute deviation (LAD), censored LAD, and weighted censored LAD estimators. On the other hand, most tests do satisfy the differentiability condition of Assumptions 1 and 1** and the imposition of this condition allows other conditions in Assumptions 1 and 1** to be weakened.

2.2.1. The Definition of Stochastic Equicontinuity

Before introducing Assumptions 1, 1*, and 1**, we define the concept of stochastic equicontinuity of a sequence of stochastic processes. Below we consider the stochastic equicontinuity of two particular sequences of stochastic processes $\{\nu_T(\cdot) : T \geq 1\}$ and $\{\nu_T(\cdot,\cdot) : T \geq 1\}$, which are indexed by $\pi \in \Pi$ and $(\beta,\pi) \in B \times \Pi$ respectively. The first sequence is used only in Assumption 1 below and the second only in Assumption 1*. Neither is assumed to be measurable. By definition,

$$\begin{aligned} \nu_T(\pi) &= \sqrt{T}(\bar{r}_T(\beta_0, \pi) - \bar{r}_T^*(\beta_0, \pi)) \text{ and} \\ \nu_T(\beta,\pi) &= \sqrt{T}(\bar{r}_T(\beta,\pi) - \bar{r}_T^*(\beta,\pi)) \text{, where} \\ \bar{r}_T^*(\beta,\pi) &= \tfrac{1}{T}\Sigma_1^T Er_t(\beta,\pi) . \end{aligned} \tag{2.2.1}$$

Let $\rho_\Pi(\cdot,\cdot)$ and $\rho_{B\times\Pi}(\cdot,\cdot)$ denote pseudo–metrics on Π and $B \times \Pi$ respectively. The former is used with Assumption 1 and the latter with 1*. Convergence in probability of $\hat{\pi}$ to π_0 and $(\hat{\beta},\hat{\pi})$ to (β_0, π_0) means convergence with respect to ρ_Π and $\rho_{B\times\Pi}$ respectively.

Given the pseudo–metric $\rho_{\Pi}(\cdot,\cdot)$, stochastic equicontinuity of $\{\nu_T(\cdot) : T \geq 1\}$ is defined as follows:

DEFINITION: $\{\nu_T(\cdot) : T \geq 1\}$ is *stochastically equicontinuous at* π_0 if: For all $\epsilon > 0$ and $\eta > 0$ there exists $\delta > 0$ such that

$$\overline{\lim_{T\to\infty}} P^*\Big(\sup_{\pi\in\Pi:\rho_{\Pi}(\pi,\pi_0)\leq\delta} \|\nu_T(\pi) - \nu_T(\pi_0)\| > \eta\Big) < \epsilon , \tag{2.2.2}$$

where **P*** denotes outer probability.

Stochastic equicontinuity of $\{\nu_T(\cdot,\cdot) : T \geq 1\}$ at (β_0, π_0) is defined analogously with π replaced by (β,π) and ρ_{Π} replaced by $\rho_{B\times\Pi}$. For notational simplicity in Sections 3–5 below, we say that a stochastic process $\{X_T(\cdot) : T \geq 1\}$ *minus its mean* is stochastically equicontinuous if $\{X_T(\cdot) - EX_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous.

The stochastic equicontinuity condition is used to establish the asymptotic normality of semiparametric estimators in Andrews (1990a). Primitive sufficient conditions for stochastic equicontinuity are discussed in Andrews (1990b).

2.2.2. Asymptotic Normality of $\sqrt{T}\,\bar{r}_T(\hat{\beta},\hat{\pi})$

We now state Assumptions 1, 1*, and 1** and the asymptotic normality result for $\sqrt{T}\,\bar{r}_T(\hat{\beta},\hat{\pi})$. Throughout this paper, we let B_0 denote a subset of B ($\subset R^q$) that contains a neighborhood of β_0.

ASSUMPTION 1: (a) $\sqrt{T}(\hat{\beta}-\beta_0) = \frac{1}{\sqrt{T}}\Sigma_1^T \psi_{Tt}(W_{Tt}, \beta_0) + o_p(1)$ and $\frac{1}{\sqrt{T}}\Sigma_1^T E\psi_{Tt}(W_{Tt}, \beta_0) \to \underset{\sim}{0}$, where $\psi_{Tt}(\cdot,\cdot) : R^{k_{Tt}} \times B \to R^q$ for $t = 1, \ldots, T$; $T \geq 1$.

(b) $P(\hat{\pi} \in \Pi) \to 1$ and $\hat{\pi} \xrightarrow{p} \pi_0$ for some $\pi_0 \in \Pi$.

(c) $\sqrt{T}\,\bar{r}_T^*(\beta_0, \hat{\pi}) \xrightarrow{p} \underset{\sim}{0}$.

(d) $\sqrt{T}\,\bar{g}_T(\beta_0, \pi_0) = \begin{bmatrix} \sqrt{T}(\bar{r}_T(\beta_0, \pi_0) - \bar{r}_T^*(\beta_0, \pi_0)) \\ \sqrt{T}(\bar{\psi}_T(\beta_0) - \bar{\psi}_T^*(\beta_0)) \end{bmatrix} \xrightarrow{d} N(\underset{\sim}{0},\Sigma)$, where

$\bar{g}_T(\beta_0, \pi_0) = \frac{1}{T}\Sigma_1^T g_t(\beta_0, \pi_0)$, $\bar{\psi}_T(\beta_0) = \frac{1}{T}\Sigma_1^T \psi_{Tt}(W_{Tt}, \beta_0)$, $\bar{\psi}_T^*(\beta_0) = E\bar{\psi}_T(\beta_0)$, and

$\Sigma = \lim_{T\to\infty} \text{Var}(\sqrt{T}\,\bar{g}_T(\beta_0, \pi_0))$.

(e) $\{\nu_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous at π_0 .

(f) $r_t(\beta,\pi)$ is differentiable in β on B_0 $\forall\pi \in \Pi$ $\forall t \geq 1$ $\forall\omega \in \Omega$. $\left\{\frac{\partial}{\partial\beta'} r_t(\beta,\pi) : t \geq 1\right\}$ satisfies a uniform WLLN over $B_0 \times \Pi$. $R(\beta,\pi) = \lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E\frac{\partial}{\partial\beta'} r_t(\beta,\pi)$ exists uniformly over $B_0 \times \Pi$ and is continuous at (β_0, π_0) with respect to some pseudo–metric on $B_0 \times \Pi$ for which $(\hat{\beta},\hat{\pi}) \xrightarrow{p} (\beta_0, \pi_0)$.[2]

ASSUMPTION 1*: (a) Assumption 1(a) holds.

(b) $P(\hat{\pi} \in \Pi) \to 1$ and $(\hat{\beta},\hat{\pi}) \xrightarrow{p} (\beta_0, \pi_0)$ for some $\pi_0 \in \Pi$.

(c) Assumption 1(c) holds.

(d) Assumption 1(d) holds.

(e) $\{\nu_T(\cdot,\cdot) : T \geq 1\}$ is stochastically equicontinuous at (β_0, π_0) .

(f) $Er_t(\beta,\pi)$ is differentiable in β on B_0 $\forall\pi \in \Pi$ $\forall t \geq 1$.
$R(\beta,\pi) = \lim_{T\to\infty} \frac{1}{T}\Sigma_1^T \frac{\partial}{\partial\beta'} Er_t(\beta,\pi)$ exists uniformly over $B_0 \times \Pi$ and is continuous at (β_0, π_0) with respect to some pseudo–metric on $B_0 \times \Pi$ for which $(\hat{\beta},\hat{\pi}) \xrightarrow{p} (\beta_0, \pi_0)$.

ASSUMPTION 1**: Assumption 1 holds with $\Pi \subset R^u$ for some $u < \infty$, with ρ_Π given by the Euclidean metric on Π , and with Assumptions 1(c) and 1(e) replaced by 1**(c) and 1**(e) respectively.

1**(c) $\sqrt{T}\,\bar{r}_T^*(\beta_0, \pi_0) \longrightarrow \underset{\sim}{0}$.

1**(e) $\sqrt{T}(\hat{\pi} - \pi_0) = O_p(1)$, $\frac{\partial}{\partial\pi'} r_t(\beta_0, \pi)$ exists $\forall\pi \in \Pi$ $\forall t \geq 1$ $\forall\omega \in \Omega$, $\left\{\frac{\partial}{\partial\pi'} r_t(\beta_0, \pi) : t \geq 1\right\}$ satisfies a uniform WLLN over $\pi \in \Pi$, $Q(\beta_0, \pi) = \lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E\frac{\partial}{\partial\pi'} r_t(\beta_0, \pi)$ exists uniformly over Π and is continuous at π_0 , $Q(\beta_0, \pi_0) = \underset{\sim}{0}$, and $E \sup_{\pi\in\Pi}\left\|\frac{\partial}{\partial\pi'} r_t(\beta_0, \pi)\right\| < \infty$ $\forall t \geq 1$.

The $(m+q) \times (m+q)$ matrix Σ in part (d) of the above Assumptions can be partitioned as follows:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix}, \text{ where } \Sigma_{11} = \lim_{T\to\infty} \mathrm{Var}(\sqrt{T}\,\bar{r}_T(\beta_0, \pi_0)), \tag{2.2.3}$$

$$\Sigma_{12} = \lim_{T\to\infty} \mathrm{Cov}(\sqrt{T}\bar{r}_T(\beta_0,\pi_0), \sqrt{T}\bar{\psi}_T(\beta_0)), \text{ and } \Sigma_{22} = \lim_{T\to\infty} \mathrm{Var}(\sqrt{T}\bar{\psi}_T(\beta_0)).$$

Let I_m denote an $m \times m$ identity matrix. Using this notation, the asymptotic normality of $\sqrt{T}\,\bar{r}_T(\hat{\beta},\hat{\pi})$ is given in the following lemma.

LEMMA 1: *Suppose Assumption* 1, 1*, *or* 1** *holds under the null hypothesis of correct specification. Then,*

$$\sqrt{T}\,\bar{r}_T(\hat{\beta},\hat{\pi}) \xrightarrow{d} N(\underset{\sim}{0}, \Phi), \text{ where } \Phi = [I_m \vdots R]\Sigma[I_m \vdots R]' = \Sigma_{11} + R\Sigma_{12}' + \Sigma_{12}R' + R\Sigma_{22}R' \text{ and } R = R(\beta_0, \pi_0). \tag{2.2.4}$$

COMMENTS: 1. It is the mean zero property of the asymptotic distribution of $\sqrt{T}\,\bar{r}_T(\hat{\beta},\hat{\pi})$ under the null hypothesis of correct specification that makes $\sqrt{T}\,\bar{r}_T(\hat{\beta},\hat{\pi})$ a useful indicator of misspecification. Under a sequence of local alternatives, the mean generally differs from zero as is shown below.

2. We note that assumptions on $\hat{\pi}$ and on the random criterion function $\bar{r}_T(\beta,\pi)$ are split apart. Assumption 1(b) only involves $\hat{\pi}$ and Assumption 1(c) only involves $\hat{\pi}$ and the *non–random function* $\bar{r}_T^*(\beta_0, \pi)$. Thus, the fact that $\hat{\pi}$ and the random function $\bar{r}_T(\beta_0, \pi)$ are defined using the same underlying rv's does not present a problem.

3. Suppose π_0 is a function of x for $x \in \mathcal{X}$ and $\mathcal{X}$ is an unbounded set. One may wish to trim the criterion function $\bar{r}_T(\beta,\hat{\pi})$ by multiplying each summand $r_{Tt}(W_{Tt},\beta,\pi)$ by $1(X_{Tt} \in \mathcal{X}^*)$, where $\mathcal{X}^*$ $(\subset \mathcal{X})$ is a bounded set and X_{Tt} is a subvector of W_{Tt}. There are two reasons for such trimming. First, trimming can eliminate observations from the sample average $\bar{r}_T(\hat{\beta},\hat{\pi})$ for which the nuisance parameter estimator $\hat{\pi}(X_{Tt})$ is estimated with relatively large error in comparison to non–trimmed observations. This may reduce the discrepancy between the true and nominal size of the test. Second, trimming allows one to obtain uniform consistency of $\hat{\pi}(x)$ for $\pi_0(x)$ over $\mathcal{X}^*$ under suitable

conditions. Uniform consistency simplifies the verification of Assumption 1(b) (see the discussion of Assumption 1(b) below). On the other hand, trimming using a single fixed set $\mathcal{X}^*$ affects the asymptotic distribution of $\sqrt{T}\,\bar{r}_T(\hat{\beta},\hat{\pi})$ under the null and local alternatives and may lead to a reduction in the power of the test. For simplicity, we do not make trimming explicit in the expressions given for $r_t(\beta,\pi)$ and other quantities either in this section or in others below. If trimming is carried out, then indicator functions need to be added in the appropriate places.

We now discuss Assumption 1. Assumption 1(a) can be verified for most parametric and semiparametric estimators that are $\sqrt{T}$–consistent and asymptotically normal using results in the literature. For example, for semiparametric MINPIN estimators, 1(a) can be verified using results of Andrews (1990a, b). Note that for the parametric ML estimator $\psi_{Tt}(W_{Tt}, \beta_0)$ is just the score function evaluated at β_0 premultiplied by the inverse of the information matrix.

Assumptions 1(b), (c), and (e) are key assumptions – they are discussed below.

Assumption 1(d) can be verified using any of a number of CLTs for iid, inid, and dnid contexts. For example, see McLeish (1975, 1977), Hall and Heyde (1980, Chs. 3–5), Herrndorf (1984), Gallant (1987), and Wooldridge and White (1988a, b).

Assumption 1(f) requires $r_t(\beta,\pi)$ to be differentiable in β. This assumption can be avoided, if necessary, by using Assumption 1*. Assumption 1(f) requires a certain uniform WLLN to hold. This can be verified using stochastic equicontinuity results (see Andrews (1990b)), generic uniform WLLN results (see Andrews (1987, 1990c), Pötscher and Prucha (1989), and Newey (1989)), or empirical process or Banach space WLLN results (see Pollard (1984, Theorems II.2, II.24 and II.25)). Assumption 1(f) also requires that $R(\beta,\pi)$ is continuous with respect to some pseudo–metric on $B_0 \times \Pi$ for which $(\hat{\beta},\hat{\pi}) \xrightarrow{p} (\beta_0, \pi_0)$. A convenient choice of the pseudo–metric is

$$\rho^*((\beta_1, \pi_1), (\beta_2, \pi_2)) = \overline{\lim_{N\to\infty}} \frac{1}{N}\Sigma_1^N E\left\| \frac{\partial}{\partial\beta'} r_t(\beta_1, \pi_1) - \frac{\partial}{\partial\beta'} r_t(\beta_2, \pi_2)\right\| . \quad (2.2.5)$$

With this choice, continuity of $R(\beta,\pi)$ at (β_0, π_0) automatically holds (since $\|R(\beta_1, \pi_1) - R(\beta_2, \pi_2)\| \le \rho^*((\beta_1, \pi_1), (\beta_2, \pi_2))$) and it suffices to verify that $\rho^*((\hat{\beta},\hat{\pi}), (\beta_0, \pi_0)) \xrightarrow{p} 0$.

The stochastic equicontinuity assumption, Assumption 1(e), can be verified using results in Andrews (1990b) or other results in the literature. In order to obtain stochastic equicontinuity, the elements of the index set Π need to satisfy some conditions. This creates a tension between Assumption 1(e) and the first part of Assumption 1(b), since the more restricted is Π , the more difficult it is to show that $P(\hat{\pi} \in \Pi) \to 1$. For example, if Π is an infinite dimensional class of functions, the stochastic equicontinuity results of Andrews (1990b) require the functions in Π to satisfy smoothness conditions. When Π is defined as such, one has to show that the nonparametric function estimator $\hat{\pi}$ also satisfies these smoothness conditions with probability $\to 1$ to verify the first part of Assumption 1(b).

For example, if π_0 is a function of x for $x \in \mathcal{X}$, π_0 satisfies the smoothness conditions of Andrews (1990b), and $\hat{\pi}$ and a suitable number of its derivatives converge in probability uniformly over $x \in \mathcal{X}$ to π_0 and its corresponding derivatives, then the first part of Assumption 1(b) will hold. Note that uniform convergence of nonparametric regression estimators and their derivatives generally requires the domain $\mathcal{X}$ of the functions to be bounded and the absolutely continuous components of the distributions of the regressor variables $\{X_t\}$ to be bounded away from zero on $\mathcal{X}$. If $\mathcal{X}$ is unbounded, these properties can often be obtained by restricting $\mathcal{X}$ to a large but bounded set. Alternatively, one can employ a trimming procedure that replaces the uniform convergence requirement with conditions that allow $\mathcal{X}$ to be unbounded and $\{X_t\}$ to have distributions with densities that are not bounded away from zero on its support. In either case, when establishing the first part of Assumption 1(b), one can exploit existing consistency results and proofs for nonparametric estimators of regression and density functions and their derivatives.

Next we discuss the pseudo–metric ρ_Π on Π . As with the choice of Π , there is a tension between Assumptions 1(b) and (e) with regard to the choice of ρ_Π . The stronger is the pseudo–metric, the easier it is to verify Assumption 1(e), but the more difficult it is to verify the condition of Assumption 1(b) that $\rho_\Pi(\hat{\pi}, \pi_0) \xrightarrow{p} 0$. It is this tension and the availability of stochastic equicontinuity results for different pseudo–metrics that determine the most appropriate choice of pseudo–metric. Examples of pseudo-metrics for which stochastic equicontinuity results are available (see Andrews (1990b)) include:

$$\rho_\Pi(\pi_1, \pi_2) = \sup_{N\geq 1}\left[\frac{1}{N}\Sigma_1^N E\|r_t(\beta_0, \pi_1) - r_t(\beta_0, \pi_2)\|^2\right]^{1/2} \text{ and } \tag{2.2.6}$$

$$\rho_\Pi(\pi_1, \pi_2) = \left[\int_{\mathscr{W}}\|r(w, \beta_0, \pi_1) - r(w, \beta_0, \pi_2)\|^2 dw\right]^{1/2} . \tag{2.2.7}$$

The pseudo–metric defined in (2.2.7) can be used when $r_{Tt}(\cdot,\cdot,\cdot)$ does not depend on T or t and W_{Tt} takes values in a bounded set $\mathscr{W}$.

Assumption 1(c) is a key assumption. It is needed to show that preliminary estimation of π_0 does not affect the asymptotic distribution of $\sqrt{T}\, \bar{r}_T(\hat{\beta},\hat{\pi})$. Assumption 1(c) holds if a probability model satisfies

$$\bar{r}_T^*(\beta_0, \pi) = \underset{\sim}{0} \quad \forall\pi \text{ in some neighborhood of } \pi_0 \tag{2.2.8}$$

for all T sufficiently large. In those cases where Assumption 1(c) holds (for suitable $\hat{\pi}$) but (2.2.8) does not hold, Assumption 1(c) often can be verified if $\hat{\pi}$ is L^ξ–consistent for π_0 at rate $T^{1/4}$ for some $\xi \geq 1$, i.e.

$$T^{1/4}\left[\int\|\hat{\pi}(x) - \pi_0(x)\|^\xi dP(x)\right]^{1/\xi} \xrightarrow{p} 0 ,$$

e.g., see the partially linear regression and sample selection model examples of Section 3 below. In some other cases where Assumption 1(c) holds (for suitable $\hat{\pi}$) but (2.2.8) does not hold, one may need to verify L^ξ–consistency of $\hat{\pi}$ *and* some of its derivatives for π_0 and its corresponding derivatives at some rate such as $T^{1/4}$, e.g., see the binary choice example in Whang and Andrews (1990). Note that the criterion function should be chosen such that $\bar{r}^*_T(\beta_0, \pi_0) = \underset{\sim}{0}$ for all T sufficiently large if the model is correctly specified.

Next, we discuss Assumption 1* and compare it with Assumption 1. Assumption 1*(b) requires $\rho_{B\times\Pi}((\hat{\beta},\hat{\pi}), (\beta_0, \pi_0)) \xrightarrow{p} 0$, where $\rho_{B\times\Pi}$ is a pseudo–metric on $B \times \Pi$ that is suitable for establishing the stochastic equicontinuity condition of Assumption 1*(e). For example, one could take

$$\rho_{B\times\Pi}((\beta_1, \pi_1), (\beta_2, \pi_2)) = \sup_{N\geq 1}\left[\frac{1}{N}\Sigma_1^N E\|r_t(\beta_1, \pi_1) - r_t(\beta_2, \pi_2)\|^2\right]^{1/2} . \qquad (2.2.9)$$

Alternatively, one could take

$$\rho_{B\times\Pi}((\beta_1, \pi_1), (\beta_2, \pi_2)) = \left[\int_{\mathcal{W}}\|r(w, \beta_1, \pi_1) - r(w, \beta_2, \pi_2)\|^2 dw\right]^{1/2} \qquad (2.2.10)$$

in the case where $r_{Tt}(\cdot,\cdot,\cdot)$ does not depend on T or t and W_{Tt} takes values in a bounded set $\mathcal{W}$.

Assumption 1*(e) is stronger than Assumption 1(e), because it requires stochastic equicontinuity to hold for a sequence of stochastic processes that is indexed by two parameters rather than just one. Assumption 1*(e) also can be verified using results of Andrews (1990b).

Assumption 1*(f) is weaker than 1(f), because it requires differentiability of $Er_t(\beta,\pi)$ rather than $r_t(\beta,\pi)$ and it does not require certain uniform WLLNs to hold. To verify continuity of $R(\beta,\pi)$ at (β_0, π_0), as required by Assumption 1*(f), the following pseudo–metric can be used:

$$\rho^*((\beta_1, \pi_1), (\beta_2, \pi_2)) = \overline{\lim_{N\to\infty}} \frac{1}{N}\Sigma_1^N\left\|\frac{\partial}{\partial\beta'}Er_t(\beta_1, \pi_1) - \frac{\partial}{\partial\beta'}Er_t(\beta_2, \pi_2)\right\| . \qquad (2.2.11)$$

With this choice of pseudo–metric, continuity of $R(\beta,\pi)$ at (β_0, π_0) automatically holds and it suffices to verify that $\rho^*((\hat{\beta},\hat{\pi}), (\beta_0, \pi_0)) \xrightarrow{p} 0$.

Now we briefly discuss Assumption 1**. This applies in the case where π is finite dimensional and $r_t(\beta,\pi)$ is differentiable in β and π. Assumption 1** is more general than the corresponding assumptions in Newey (1985a) and Tauchen (1985) in that the latter papers consider only iid rv's.

2.2.3. Asymptotic Distribution of the Test Statistic

In this section, we discuss consistent estimation of the covariance matrix Φ (defined in (2.2.4)) of the limit distribution of $\sqrt{T}\,\bar{r}_T(\hat{\beta},\hat{\pi})$. We also derive the asymptotic distribution of the test statistic G_T (defined in (2.1.4)) under the null hypothesis of correct specification.

When Assumption 1 or 1** is used, we define

$$\hat{R} = \frac{1}{T}\Sigma_1^T \frac{\partial}{\partial\beta'} r_t(\hat{\beta},\hat{\pi}) . \tag{2.2.12}$$

$\hat{R}$ is consistent for R under the null hypothesis of correct specification when Assumption 1 or 1** holds. We define $\hat{R}$ differently when Assumption 1* is used, since (2.2.12) does not exist when $r_t(\beta,\pi)$ is not differentiable in β. In this case, we consider the following finite difference estimator of R. For a constant or scalar rv $\epsilon_T > 0$, define the j–th column of $\hat{R}$ by

$$\hat{R}_j = \frac{1}{T}\Sigma_1^T(r_t(\hat{\beta} + \epsilon_T e_j, \hat{\pi}) - r_t(\hat{\beta} - \epsilon_T e_j, \hat{\pi}))/(2\epsilon_T) , \tag{2.2.13}$$

where $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)'$ is the j–th elementary q–vector for $j = 1, \ldots, q$. The estimator $\hat{R}$ of (2.2.13) is consistent for R under Assumption 1* and the following assumption.

ASSUMPTION 2*: (a) $\epsilon_T \xrightarrow{p} 0$ and $\epsilon_T^{-1} = O_p(\sqrt{T})$.

(b) $Er_t(\beta,\pi)$ is differentiable in β uniformly over $\beta \in B_0$, $\pi \in \Pi$, and $t \geq 1$ (i.e., $\lim_{\epsilon\to 0}\sup_{t\geq 1}\sup_{\beta\in B_0, \pi\in\Pi}\left\|(Er_t(\beta + \epsilon e_j, \pi) - Er_t(\beta - \epsilon e_j, \pi))/(2\epsilon) - \frac{\partial}{\partial\beta_j}Em_t(\beta,\pi)\right\| = 0 \;\; \forall j \leq q$).

(c) $(\hat{\beta} + \epsilon_T e_j, \hat{\pi}) \xrightarrow{p} (\beta_0, \pi_0)$ and $(\hat{\beta} - \epsilon_T e_j, \hat{\pi}) \xrightarrow{p} (\beta_0, \pi_0)$ $\forall j \leq q$.

Note that Assumption 2*(c) is very similar to the second part of Assumption 1*(b). In consequence, the verification of the former is usually a trivial extension of the verification of the latter.

Next, we discuss estimation of the matrix Σ. Let $\hat{\Sigma}$ be an estimator of Σ. If $\{g_t(\beta_0, \pi_0)\}$ (defined in Assumption 1(d)) is a sequence of independent or orthogonal rv's, then we can take

$$\hat{\Sigma} = \frac{1}{T}\Sigma_1^T g_t(\hat{\beta},\hat{\pi})g_t(\hat{\beta},\hat{\pi})' . \tag{2.2.14}$$

If $\{g_t(\beta_0, \pi_0)\}$ is m–dependent, then the following estimator can be used:

$$\hat{\Sigma} = \frac{1}{T}\Sigma_1^T \hat{g}_t\hat{g}_t' + \sum_{v=1}^{m} \frac{1}{T}\Sigma_{1+v}^T[\hat{g}_t\hat{g}_{t-v}' + \hat{g}_{t-v}\hat{g}_t'] , \tag{2.2.15}$$

where $\hat{g}_t = g_t(\hat{\beta},\hat{\pi})$. If $\{g_t(\beta_0, \pi_0)\}$ is neither orthogonal nor m–dependent, then a more complicated estimator of Σ is required. In particular, we can apply results for heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimators for parametric and semiparametric estimators, using $\{g_t(\hat{\beta},\hat{\pi})\}$ as the underlying rv's, see White (1984, pp. 147–161), Newey and West (1987), Gallant (1987), Andrews (1990d, 1991b), and Andrews and Monahan (1990).

The estimator $\hat{\Sigma}$ of Σ that is adopted is assumed to satisfy:

ASSUMPTION 2: $\hat{\Sigma} \xrightarrow{p} \Sigma$ (where Σ is as in Assumption 1, 1*, or 1**).

We also assume:

ASSUMPTION 3: Φ is nonsingular.

Let

$$\hat{\Phi} = [I_m \vdots \hat{R}]\hat{\Sigma}[I_m \vdots \hat{R}]' . \tag{2.2.16}$$

We now state the main result of this section.

THEOREM 1: *Suppose Assumptions* 1, 2, *and* 3, *or* 1**, 2, *and* 3, *or* 1*, 2*, 2, *and* 3 *hold under the null hypothesis of correct specification. Then,*

$$G_T = T\bar{r}_T(\hat{\beta},\hat{\pi})'\,\hat{\Phi}^{-1}\bar{r}_T(\hat{\beta},\hat{\pi}) \xrightarrow{d} \chi^2_m\,, \tag{2.2.17}$$

where $\hat{R}$ *is as defined in* (2.2.12), *or* (2.2.12), *or* (2.2.13) *respectively, and* χ^2_m *denotes the chi–square distribution with* m *degrees of freedom.*

(Under the assumptions of the theorem, $\hat{\Phi}$ is nonsingular with probability $\to$ 1, and so $\hat{\Phi}^{-1}$ is well–defined with probability $\to$ 1.)

2.3. Asymptotic Power Properties of the Test under Misspecification

In this section, we establish local power and consistency results for the test discussed above. To obtain asymptotic local power (ℓp) results for the test, we impose one of the three following assumptions.

ASSUMPTION 1–ℓp [1*–ℓp]: Assumption 1 [1*] holds with $\sqrt{T}(\hat{\beta}-\beta_0)$ replaced by $\sqrt{T}(\hat{\beta}-\beta_T)$ in part (a), where $\beta_T = \beta_0 + \eta/\sqrt{T}$ for some $\eta \in R^q$, and with 1(c) [1*(c)] replaced by $\sqrt{T}\,\bar{r}^*_T(\beta_0, \hat{\pi}) \xrightarrow{p} \xi$ for some $\xi \in R^m$.

ASSUMPTION 1–ℓp:** Assumption 1** holds with $\sqrt{T}(\hat{\beta}-\beta_0)$ replaced by $\sqrt{T}(\hat{\beta}-\beta_T)$ in part (a), where $\beta_T = \beta_0 + \eta/\sqrt{T}$ for some $\eta \in R^q$, and with 1**(c) replaced by $\sqrt{T}\,\bar{r}^*_T(\beta_0, \pi_0) \longrightarrow \xi$ for some $\xi \in R^m$.

COMMENT: In the above Assumptions, we consider a sequence of local alternatives under which the mean of the limit distribution of $\sqrt{T}(\hat{\beta}-\beta_0)$ is non–zero and part (c) of Assumptions 1, 1*, and 1** is violated. There are two potential sources of local power of the test since the noncentrality parameter of the limit distribution of the test statistic G_T depends on ξ and η as shown in Theorem 2 below.

In some examples considered below, we show that the values of ξ and η can be determined more specifically. For example, both ξ and η are non–zero vectors under a sequence of local alternatives for a test of omitted variables in partially linear regression models. For a test of autocorrelation, however, η equals zero because we expect $\hat{\beta}$ to be consistent for β_0 even in the presence of autocorrelation of the errors (see Section 3.1 below). On the other hand, in first–order conditions based tests (see Section 4 below) ξ equals a zero vector, but η does not equal zero in general.

THEOREM 2: *Under Assumptions* 1–ℓp, 2, *and* 3, *or* 1**–ℓp, 2, *and* 3, *or* 1*–ℓp, 2*, 2, *and* 3,

$$G_T = T\bar{r}_T(\hat{\beta},\hat{\pi})'\hat{\Phi}^{-1}\bar{r}_T(\hat{\beta},\hat{\pi}) \xrightarrow{d} \chi^2_m(\delta^2)\,, \tag{2.3.1}$$

where $\hat{\Phi}$ *is as defined in Theorem* 1, $\delta^2 = (\xi+R\eta)'\Phi^{-1}(\xi+R\eta)$, *and* $\chi^2_m(\delta^2)$ *denotes the noncentral chi–square distribution with* m *degrees of freedom and noncentrality parameter* δ^2.

Next we consider the consistency properties of the test G_T. We note that the parameters β_0 and π_0 are defined to be the probability limits of $\hat{\beta}$ and $\hat{\pi}$ under the null hypothesis. Under alternative distributions, these probability limits may be different. Thus, we suppose that $\hat{\beta} \xrightarrow{p} \bar{\beta}$ and $\hat{\pi} \xrightarrow{p} \bar{\pi}$ under a given alternative distribution, where $(\bar{\beta},\bar{\pi})$ may differ from (β_0, π_0).

Below, we show that the test G_T is consistent against alternatives for which

$$\|r(\bar{\beta},\bar{\pi})\| > 0\,,\ \text{where}\ r(\bar{\beta},\bar{\pi}) = \lim_{T\to\infty}\tfrac{1}{T}\Sigma_1^T Er_t(\bar{\beta},\bar{\pi})\,, \tag{2.3.2}$$

provided the following assumptions hold. Define $\bar{B}_0$ to be a subset of B $(\subset R^q)$ that contains a neighborhood of $\bar{\beta}$.

ASSUMPTION 4: (a) $\hat{\beta} \xrightarrow{p} \bar{\beta}$ for some $\bar{\beta} \in B$.

(b) $P(\hat{\pi} \in \Pi) \to 1$.

(c) $\{r_t(\beta,\pi) : t \geq 1\}$ satisfies a uniform WLLN over $\bar{B}_0 \times \Pi$. $r(\beta,\pi) = \lim_{T\to\infty}\frac{1}{T}\Sigma_1^T Er_t(\beta,\pi)$ exists uniformly over $\bar{B}_0 \times \Pi$ and is continuous at $(\bar{\beta},\bar{\pi})$ with respect to some pseudo-metric on $\bar{B}_0 \times \Pi$ for which $(\hat{\beta},\hat{\pi}) \xrightarrow{p} (\bar{\beta},\bar{\pi})$ for some $\bar{\pi} \in \Pi$.

(d) $\hat{\Phi} \xrightarrow{p} \bar{\Phi}$ for some nonsingular matrix $\bar{\Phi}$.

THEOREM 3: *Suppose Assumption* 4 *holds and* $\|\mathrm{r}(\bar{\beta},\bar{\pi})\| > 0$. *Then,*

$$G_T = T\bar{r}_T(\hat{\beta},\hat{\pi})'\hat{\Phi}^{-1}\bar{r}_T(\hat{\beta},\hat{\pi}) \xrightarrow{p} \infty . \tag{2.3.3}$$

COMMENTS: 1. If $\hat{\beta}$ is a quasi–MLE in a parametric model, then $\bar{\beta}$ is the parameter vector that minimizes the Kullback–Leibler Information Criterion (KLIC) between the true distribution and the alternative distribution if there exists a unique minimizer (see White (1982)). When π is infinite dimensional, $\bar{\pi}$ is usually equal to a nonparametric regression or density function under the alternative distribution.

2. To verify continuity of $\mathrm{r}(\beta,\pi)$ at $(\bar{\beta},\bar{\pi})$ in part (c) of Assumption 4, the following pseudo–metric can be chosen:

$$\rho^{**}((\beta_1, \pi_1), (\beta_2, \pi_2)) = \overline{\lim_{N\to\infty}} \tfrac{1}{N}\Sigma_1^N E\|r_t(\beta_1, \pi_1) - r_t(\beta_2, \pi_2)\| . \tag{2.3.4}$$

With this choice, continuity of $\mathrm{r}(\beta,\pi)$ at $(\bar{\beta},\bar{\pi})$ automatically holds and it suffices to verify that $\rho^{**}((\hat{\beta},\hat{\pi}), (\bar{\beta},\bar{\pi})) \xrightarrow{p} 0$, where $\rho^{**}(\cdot,\cdot)$ is now defined with the expectation operator E taken under the alternative distribution.

3. SPECIFICATION TESTS FOR SEMIPARAMETRIC MODELS

In this section, we consider some specification tests for partially linear regression (PLR) models (see Robinson (1988)) and semiparametric sample selection models (see Powell (1987) and Newey (1988)). Specifically, tests of omitted variables and heteroskedasticity are considered. In the PLR model, a test of autocorrelation is also considered. We note that to date none of these tests has been considered in the literature.

3.1. Specification Tests for Partially Linear Regression Models

3.1.1. Introduction

The model is given by:

$$Y_t = X_t'\beta_0 + f(Z_t) + U_t \text{ and } E(U_t|X_t, Z_t) = 0 \text{ a.s.} \tag{3.1.1}$$

for $t = 1, \ldots, T$, where the real function f is unknown, $Y_t, U_t \in R$, $X_t, \beta_0 \in R^q$, and $Z_t \in R^p$. Assumptions on $(Y_t, X_t', Z_t', U_t)'$ are given below.

Define

$$\pi_{10}(Z_t) = E(Y_t|Z_t) \text{ and } \pi_{20}(Z_t) = E(X_t|Z_t) . \tag{3.1.2}$$

Let $\hat{\pi}_1(\cdot)$ and $\hat{\pi}_2(\cdot)$ be estimators of $\pi_{10}(\cdot)$ and $\pi_{20}(\cdot)$ respectively. We consider the following semiparametric estimator $\hat{\beta}$ of β_0:

$$\hat{\beta} = \left[\Sigma_1^T(X_t - \hat{\pi}_2(Z_t))(X_t - \hat{\pi}_2(Z_t))'\right]^{-1}\Sigma_1^T(X_t - \hat{\pi}_2(Z_t))(Y_t - \hat{\pi}_1(Z_t)) . \tag{3.1.3}$$

Robinson (1988), Chamberlain (1986), and Andrews (1991a) establish the asymptotic normality of $\hat{\beta}$ defined using different nonparametric estimators of π_{10} and π_{20}.

The semiparametric estimator $\hat{\beta}$ is attractive in the sense that parametric estimators of β_0 based on an incorrect parametrization of $f(\cdot)$ are generally inconsistent. As in parametric models, however, statistical inference based on the model (3.1.1) may result in misleading conclusions about β_0 if some relevant regressor variables have been left out or if the errors exhibit heteroskedasticity or autocorrelation when they are assumed to be iid. Below we consider testing procedures to detect possible misspecifications of the above type using some well–known methods for parametric models.

3.1.2. A Test of Omitted Variables

Suppose the model (3.1.1) is a potentially misspecified version of the model:

$$Y_t = X_t'\beta_0 + f(Z_t) + Q_t'\gamma_0 + \epsilon_t \text{ and } E(\epsilon_t|X_t, Z_t, Q_t) = 0 \text{ a.s.} \tag{3.1.4}$$

for $t = 1, \ldots, T$, where $Q_t, \gamma_0 \in R^m$. The null hypothesis of interest is $H_0 : \gamma_0 = \underset{\sim}{0}$.

Define

$$\pi_{30}(Z_t) = E(Q_t | Z_t) \tag{3.1.5}$$

and let $\hat{\pi}_3(\cdot)$ be an estimator of $\pi_{30}(\cdot)$. The models (3.1.1) and (3.1.4) can be rewritten as follows:

$$Y_t - \pi_{10}(Z_t) = [X_t - \pi_{20}(Z_t)]'\beta_0 + U_t \text{ and} \tag{3.1.1$'$}$$

$$Y_t - \pi_{10}(Z_t) = [X_t - \pi_{20}(Z_t)]'\beta_0 + [Q_t - \pi_{30}(Z_t)]'\gamma_0 + \epsilon_t . \tag{3.1.4$'$}$$

Now a comparison of (3.1.1)′ with (3.1.4)′ shows that

$$U_t = [Q_t - \pi_{30}(Z_t)]'\gamma_0 + \epsilon_t . \tag{3.1.6}$$

To test the null hypothesis $H_0 : \gamma_0 = \underset{\sim}{0}$, we can consider a test of the significance of the regression coefficients when we regress $\hat{U}_t$ against $Q_t - \hat{\pi}_3(Z_t)$, where $\hat{U}_t = Y_t - \hat{\pi}_1(Z_t) - (X_t - \hat{\pi}_2(Z_t))'\hat{\beta}$ is a semiparametric residual and $\hat{\beta}$ is as defined in (3.1.3). That is, we take the sample covariance between $\hat{U}_t$ and $Q_t - \hat{\pi}_3(Z_t)$ as the basis of our test. Specifically, define

$$\bar{r}_T(\hat{\beta},\hat{\pi}) = \tfrac{1}{T}\Sigma_1^T[Y_t - \hat{\pi}_1(Z_t) - (X_t - \hat{\pi}_2(Z_t))'\hat{\beta}][Q_t - \hat{\pi}_3(Z_t)] . \tag{3.1.7}$$

This choice of sample covariance is similar to that of Pagan and Hall (1983) who consider a test of omitted variables in the linear regression model. Essentially, their testing procedure applies to the case where the rv's $Y_t - \pi_{10}(Z_t)$, $X_t - \pi_{20}(Z_t)$, and $Q_t - \pi_{30}(Z_t)$ are all assumed to be observed.

The definition in (3.1.7) corresponds to

$$r_t(W_t, \beta, \pi) \equiv r_t(\beta,\pi) = [Y_t - \pi_1(Z_t) - (X_t - \pi_2(Z_t))'\beta][Q_t - \pi_3(Z_t)] , \tag{3.1.8}$$

where $W_t = (Y_t, X_t', Z_t', Q_t')'$ and $\pi = (\pi_1, \pi_2', \pi_3')'$. Note that under the null hypothesis $H_0 : \gamma_0 = \underset{\sim}{0}$, $Er_t(\beta_0, \pi_0) = E\epsilon_t(Q_t - \pi_{30}(Z_t)) = \underset{\sim}{0}$.

For simplicity, we assume below that $\{W_t\}$ are identically distributed.

Let

$$
\begin{aligned}
&\Phi = [I_m \vdots R]\Sigma[I_m \vdots R]' \text{ , where} \\
&R = -E[Q_t - \pi_{30}(Z_t)][X_t - \pi_{20}(Z_t)]' \text{ , } \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix} \text{ ,} \\
&\Sigma_{11} = \lim_{T\to\infty} \tfrac{1}{T}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T} E\epsilon_t\epsilon_s[Q_t - \pi_{30}(Z_t)][Q_s - \pi_{30}(Z_s)]' \text{ ,} \\
&\Sigma_{12} = \Big[\lim_{T\to\infty} \tfrac{1}{T}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T} E\epsilon_t\epsilon_s[Q_t - \pi_{30}(Z_t)][X_s - \pi_{20}(Z_s)]'\Big]J^{-1} \text{ ,} \\
&\Sigma_{22} = J^{-1}\Big[\lim_{T\to\infty} \tfrac{1}{T}\Sigma_{t=1}^{T}\Sigma_{s=1}^{T} E\epsilon_t\epsilon_s[X_t - \pi_{20}(Z_t)][X_s - \pi_{20}(Z_s)]'\Big]J^{-1} \text{ , and} \\
&J = E[X_t - \pi_{20}(Z_t)][X_t - \pi_{20}(Z_t)]' \text{ .}
\end{aligned}
\tag{3.1.9}
$$

Let $\hat{\Phi} = [I_m \vdots \hat{R}]\hat{\Sigma}[I_m \vdots \hat{R}]'$, where $\hat{R} = -\frac{1}{T}\Sigma_1^T[Q_t - \hat{\pi}_3(Z_t)][X_t - \hat{\pi}_2(Z_t)]'$ and $\hat{\Sigma}$ is an estimator of Σ. (See the comment following Theorem PLO below for an example of $\hat{\Sigma}$.)

Our test statistic for testing $H_0 : \gamma_0 = \underset{\sim}{0}$ is defined as follows:

$$
PLO_T = T\bar{r}_T(\hat{\beta},\hat{\pi})'\hat{\Phi}^{-1}\bar{r}_T(\hat{\beta},\hat{\pi}) \text{ ,} \tag{3.1.10}
$$

where $\bar{r}_T(\hat{\beta},\hat{\pi})$ is as defined in (3.1.7). To obtain the asymptotic null distribution of PLO_T , we impose the following assumption.

ASSUMPTION PLO: (a) $\sqrt{T}(\hat{\beta} - \beta_0) = J^{-1}\frac{1}{\sqrt{T}}\Sigma_1^T(X_t - \pi_{20}(Z_t))\epsilon_t + o_p(1)$.

(b) (i) $P(\hat{\pi} \in \Pi) \to 1$. (ii) $\int\|\hat{\pi}_j(z) - \pi_{j0}(z)\|^4 dP(z) \xrightarrow{p} 0$ for $j = 1, 2, 3$, where $P(\cdot)$ denotes the distribution of Z_t . (iii) $E\epsilon_t^4 < \infty$, $E\|X_t\|^2 < \infty$, and $E\|Q_t\|^4 < \infty$.

(c) $T^{1/4}\Big[\int\|\hat{\pi}_j(z) - \pi_{j0}(z)\|^2 dP(z)\Big]^{1/2} \xrightarrow{p} 0$ for $j = 1, 2, 3$.

(d) $\{(X_t', Z_t', Q_t', \epsilon_t)' : t \geq 1\}$ is a sequence of iid rv's or identically distributed, strong mixing rv's with mixing numbers that satisfy $\Sigma_{s=1}^{\infty}\alpha(s)^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$ such that $E\|\epsilon_t(X_t - \pi_{20}(Z_t))\|^{2+\delta} < \infty$ and $E\|\epsilon_t(Q_t - \pi_{30}(Z_t))\|^{2+\delta} < \infty$. Σ, defined in (3.1.9), exists.

(e) $\Big\{\frac{1}{\sqrt{T}}\Sigma_1^T[\epsilon_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\beta_0][Q_t - \pi_3(Z_t)] : T \geq 1\Big\}$ minus its mean is stochastically equicontinuous at $\pi = \pi_0$ with ρ_Π defined by (2.2.6).

(f) $\{[Q_t - \pi_3(Z_t)][X_t - \pi_2(Z_t)]' : t \geq 1\}$ satisfies a uniform WLLN over $\pi \in \Pi$.

(g) $\hat{\Sigma} \xrightarrow{p} \Sigma$.

(h) Φ has full rank m.

THEOREM PLO: *Suppose Assumption* PLO *holds for model* (3.1.4) *under* $H_0 : \gamma_0 = \underset{\sim}{0}$. *Then,* $PLO_T \xrightarrow{d} \chi^2_m$.

COMMENTS: 1. Assumption PLO (a) can be verified using results of Robinson (1988), Chamberlain (1986), Andrews (1991a), or Andrews (1990a) under suitable assumptions. This assumption combined with Assumption PLO (d) implies asymptotic normality of $\sqrt{T}(\hat{\beta} - \beta_0)$.

2. If the underlying rv's are iid and $E(\epsilon_t^2 | X_t, Q_t, Z_t) = \sigma_0^2 < \infty$ a.s., then the expression for Σ defined in (3.1.9) can be simplified. For example, Σ_{11} can be written as

$$\Sigma_{11} = \sigma_0^2 E[Q_t - \pi_{30}(Z_t)][Q_t - \pi_{30}(Z_t)]' \tag{3.1.11}$$

and Σ_{12} and Σ_{22} can be defined similarly. In this case, the following estimator $\hat{\Sigma}_{11}$ can be shown to be consistent for Σ_{11} under suitable conditions:

$$\hat{\Sigma}_{11} = \hat{\sigma}^2 \tfrac{1}{T}\Sigma_1^T [Q_t - \hat{\pi}_3(Z_t)][Q_t - \hat{\pi}_3(Z_t)]' , \tag{3.1.12}$$

where $\hat{\sigma}^2 = \frac{1}{T}\Sigma_1^T [Y_t - \hat{\pi}_1(Z_t) - (X_t - \hat{\pi}_2(Z_t))'\hat{\beta}]^2$. Consistent estimators of Σ_{12} and Σ_{22} are defined similarly with a consistent estimator of J defined by

$$\hat{J} = \tfrac{1}{T}\Sigma_1^T [X_t - \hat{\pi}_2(Z_t)][X_t - \hat{\pi}_2(Z_t)]' . \tag{3.1.13}$$

Next, we consider the local power properties of the test described above. Specifically, consider a sequence of local alternatives $H_T : \gamma_0 = \mu/\sqrt{T}$ for some $\mu \in R^m$, $\mu \neq \underset{\sim}{0}$. Under H_T, Assumption PLO (a) is violated generally and the following relationship holds under suitable assumptions:

$$\sqrt{T}(\hat{\beta} - \beta_T) = J^{-1} \tfrac{1}{\sqrt{T}}\Sigma_1^T (X_t - \pi_{20}(Z_t))\epsilon_t + o_p(1) , \tag{3.1.14}$$

where $\beta_T = \beta_0 - J^{-1}R'\mu/\sqrt{T}$. Equation (3.1.14) implies that Assumption 1–ℓp(a) holds under H_T with η defined by $-J^{-1}R'\mu$. Assumption 1(c) also fails to hold under H_T. We can show that $\sqrt{T}\,\bar{r}_T^*(\beta_0, \hat{\pi}) \xrightarrow{p} E(Q_t - \pi_{30}(Z_t))(Q_t - \pi_{30}(Z_t))'\mu \equiv \xi$ under H_T and Assumptions PLO (b) and (c). To see this, note that under H_T,

$$\begin{aligned}
&\|\sqrt{T}\,\bar{r}_T^*(\beta_0, \hat{\pi}) - E(Q_t - \pi_{30}(Z_t))(Q_t - \pi_{30}(Z_t))'\mu\| \\
&\leq T^{1/4}\left[\int\|\hat{\pi}_3(z) - \pi_{30}(z)\|^2 dP(z)\right]^{1/2}\left[T^{1/4}\left[\int|\hat{\pi}_1(z) - \pi_{10}(z)|^2 dP(z)\right]^{1/2}\right. \\
&\quad \left. + T^{1/4}\left[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^2 dP(z)\right]^{1/2}\|\beta_0\|\right] + 2\|\mu\|\left[E\|Q_t\|^2\right]^{1/2} \\
&\quad \times \left[\int\|\hat{\pi}_3(z) - \pi_{30}(z)\|^2 dP(z)\right]^{1/2} \xrightarrow{p} 0 .
\end{aligned} \tag{3.1.15}$$

Thus, under the local alternatives $\{H_T\}$ and Assumption PLO with PLO(a) replaced by (3.1.14), we find that Assumption 1–ℓp holds with η and ξ as above. In this case, PLO_T converges in distribution to a noncentral χ^2_m rv with noncentrality parameter given by

$$\begin{aligned}
&\delta^2_{PLO} = (\xi + R\eta)'\Phi^{-1}(\xi + R\eta)\,, \text{ where} \\
&\xi + R\eta = [E(Q_t - \pi_{30}(Z_t))(Q_t - \pi_{30}(Z_t))' - RJ^{-1}R']\mu .
\end{aligned} \tag{3.1.16}$$

Next we discuss the consistency properties of the test PLO_T. Suppose the following assumptions hold under the alternative hypothesis $H_1 : \gamma_0 \neq \underset{\sim}{0}$:

$$\begin{aligned}
&\hat{\beta} \xrightarrow{p} \beta_0 - J^{-1}R'\gamma_0 \equiv \bar{\beta},\ P(\hat{\pi} \in \Pi) \to 1\,,\ E\|X_t\|^2 < \infty\,, E\|Q_t\|^2 < \infty\,, \\
&\{[Y_t - \pi_1(Z_t) - (X_t - \pi_2(Z_t))'\beta][Q_t - \pi_3(Z_t)] : t \geq 1\} \text{ satisfies a uniform} \\
&\text{WLLN over } B_0 \times \Pi\,,\ \int\|\hat{\pi}_j(z) - \pi_{j0}(z)\|^2 dP(z) \xrightarrow{p} 0 \text{ for } j = 1, 2, 3\,, \\
&\text{and } \hat{\Phi} \xrightarrow{p} \bar{\Phi} \text{ for some positive definite matrix } \bar{\Phi} .
\end{aligned} \tag{3.1.17}$$

These assumptions are sufficient for Assumption 4 with $\bar{\pi} = \pi_0$. In addition, we have

$$r(\bar{\beta},\pi_0) = Er_t(\bar{\beta}, \pi_0) = [E(Q_t - \pi_{30}(Z_t))(Q_t - \pi_{30}(Z_t))' - RJ^{-1}R']\gamma_0 . \tag{3.1.18}$$

Hence, by Theorem 3, the test based on PLO_T is consistent whenever the expression inside the square brackets in (3.1.18) is nonsingular.

3.1.3. A Test of Heteroskedasticity

Consider the following PLR model:

$$Y_t = X_t'\theta_0 + f(Z_t) + U_t \text{ for } t = 1, \dots, T, \tag{3.1.19}$$

where $X_t, \theta_0 \in R^{q-1}$. The error U_t is assumed to satisfy the following relationship:

$$U_t = \eta_t \epsilon_t \text{ and } \eta_t = 1 + h(Q_t'\gamma_0), \tag{3.1.20}$$

where $Q_t, \gamma_0 \in R^m$, Q_t does not contain a constant term, $h(\cdot)$ is a known, continuously differentiable function with $h(0) = 0$ and $|h'(0)| < \infty$, and $\{\epsilon_t\}$ satisfy $E(\epsilon_t | X_t, Z_t, Q_t) = 0$ a.s. and $E(\epsilon_t^2 | X_t, Z_t, Q_t) = \sigma_0^2 > 0$ a.s.

For a test of heteroskedasticity of the errors $\{U_t\}$, we take $H_0 : \gamma_0 = \underset{\sim}{0}$ as our null hypothesis. Let $W_t = (Y_t, X_t', Z_t', Q_t')'$ and $\beta = (\theta, \sigma^2) \in \Theta \times R^+ = B \in R^q$. Let $\pi = (\pi_1, \pi_2')' \in \Pi$ be as defined in (3.1.2). Let $\hat{\theta}$ be the semiparametric estimator $\hat{\beta}$ of equation (3.1.3). We consider the following statistic as the basis of our test statistic:

$$\bar{r}_T(\hat{\beta},\hat{\pi}) = \tfrac{1}{T}\Sigma_1^T(\hat{U}_t^2 - \hat{\sigma}^2)Q_t, \tag{3.1.21}$$

where $\hat{\beta} = (\hat{\theta}, \hat{\sigma}^2)$, $\hat{U}_t = Y_t - \hat{\pi}_1(Z_t) - (X_t - \hat{\pi}_2(Z_t))'\hat{\theta}$, and $\hat{\sigma}^2 = \frac{1}{T}\Sigma_1^T \hat{U}_t^2$. The definition in (3.1.21) corresponds to

$$r_t(\beta,\pi) = \left[[Y_t - \pi_1(Z_t) - (X_t - \pi_2(Z_t))'\theta]^2 - \sigma^2\right]Q_t. \tag{3.1.22}$$

Under the null hypothesis $H_0 : \gamma_0 = \underset{\sim}{0}$, $U_t = \epsilon_t$ and hence $Er_t(\beta_0, \pi_0) = E(U_t^2 - \sigma_0^2)Q_t = \underset{\sim}{0}$. If $\{U_t\}$ are heteroskedastic, however, the latter result no longer holds.

The statistic $\bar{r}_T(\hat{\beta},\hat{\pi})$ in (3.1.21) is closely related to the test statistic suggested by Breusch and Pagan (1979) and Koenker (1981) who considered tests of heteroskedasticity in the linear model with errors of the form given by (3.1.20). In fact, if we regard $\hat{\beta}$ and $\hat{U}_t$ as the OLS estimator and the OLS residual, respectively, the statistic defined in (3.1.21) is exactly the same as that used by Koenker (1981). We note that our general framework can be used to generalize the results of Koenker (1981) to the PLR model.[3]

In this section, we assume that $\{W_t\} = \{(Y_t, X_t', Z_t', Q_t')'\}$ are identically distributed. Let

$$\begin{aligned}\Phi &= \Sigma_{11} - (EQ_t)\Sigma_{13}' - \Sigma_{13}(EQ_t') + (EQ_t)\Sigma_{33}(EQ_t'), \text{ where} \\ \Sigma_{11} &= \lim_{T\to\infty} \tfrac{1}{T}\Sigma_{t=1}^T \Sigma_{s=1}^T E(\epsilon_t^2 - \sigma_0^2)(\epsilon_s^2 - \sigma_0^2)Q_t Q_s', \\ \Sigma_{13} &= \lim_{T\to\infty} \tfrac{1}{T}\Sigma_{t=1}^T \Sigma_{s=1}^T E(\epsilon_t^2 - \sigma_0^2)(\epsilon_s^2 - \sigma_0^2)Q_t, \text{ and} \\ \Sigma_{33} &= \lim_{T\to\infty} \tfrac{1}{T}\Sigma_{t=1}^T \Sigma_{s=1}^T E(\epsilon_t^2 - \sigma_0^2)(\epsilon_s^2 - \sigma_0^2).\end{aligned} \tag{3.1.23}$$

In the case where the underlying rv's are iid and ϵ_t is independent of Q_t, the expression for Φ simplifies:

$$\Phi = \phi\cdot[EQ_tQ_t' - EQ_tEQ_t'], \tag{3.1.24}$$

where $\phi = E(\epsilon_t^2 - \sigma_0^2)^2$. Let $\hat{\Phi}$ be defined as Φ is but with EQ_t, Σ_{11}, Σ_{13}, and Σ_{33} replaced by $(1/T)\Sigma_1^T Q_t$, $\hat{\Sigma}_{11}$, $\hat{\Sigma}_{13}$, and $\hat{\Sigma}_{33}$, respectively, where $\hat{\Sigma}_{11}$, $\hat{\Sigma}_{13}$, and $\hat{\Sigma}_{33}$ are some consistent estimators of Σ_{11}, Σ_{13}, and Σ_{33}.

Our test statistic is defined as follows:

$$PLH_T = T\bar{r}_T(\hat{\beta},\hat{\pi})'\hat{\Phi}^{-1}\bar{r}_T(\hat{\beta},\hat{\pi}), \tag{3.1.25}$$

where $\bar{r}_T(\hat{\beta},\hat{\pi})$ is as defined in (3.1.21). To analyze the asymptotic behavior of PLH_T under $H_0 : \gamma_0 = \underset{\sim}{0}$, we impose the following assumption.

ASSUMPTION PLH: (a) $\sqrt{T}\begin{bmatrix}\hat{\theta} - \theta_0 \\ \hat{\sigma}^2 - \sigma_0^2\end{bmatrix} = \begin{bmatrix}J^{-1} & 0 \\ 0 & 1\end{bmatrix}\frac{1}{\sqrt{T}}\Sigma_1^T\begin{bmatrix}(X_t - \pi_{20}(Z_t))\epsilon_t \\ \epsilon_t^2 - \sigma_0^2\end{bmatrix} + o_p(1)$,

where J is as defined in (3.1.9).

(b) (i) $P(\hat{\pi} \in \Pi) \to 1$. (ii) $\int\|\hat{\pi}_j(z) - \pi_{j0}(z)\|^8 dP(z) \xrightarrow{p} 0$ for $j = 1, 2$, where $P(\cdot)$ denotes the distribution of Z_t. (iii) $E\epsilon_t^8 < \infty$, $E\|X_t\|^4 < \infty$, and $E\|Q_t\|^4 < \infty$.

(c) $T^{1/4}\left[\int\|\hat{\pi}(z) - \pi_{j0}(z)\|^4 dP(z)\right]^{1/4} \xrightarrow{p} 0$ for $j = 1, 2$.

(d) $\{(X_t', Z_t', Q_t', \epsilon_t)' : t \geq 1\}$ is a sequence of iid rv's or identically distributed, strong mixing rv's with mixing numbers that satisfy $\Sigma_{s=1}^{\infty}\alpha(s)^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$ such that $E\|\epsilon_t^2 Q_t\|^{2+\delta} < \infty$, $E\|\epsilon_t(X_t - \pi_{20}(Z_t))\|^{2+\delta} < \infty$, and $E|\epsilon_t|^{4+2\delta} < \infty$. $\Sigma = \lim_{T\to\infty} \mathrm{Var}\left[\frac{1}{\sqrt{T}}\Sigma_1^T[(\epsilon_t^2 - \sigma_0^2)Q_t' \vdots J^{-1}(X_t - \pi_{20}(Z_t))'\epsilon_t \vdots \epsilon_t^2 - \sigma_0^2]\right]$ exists.

(e) $\left\{\frac{1}{\sqrt{T}}\Sigma_1^T\left[(\epsilon_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\theta_0)^2 - \sigma_0^2\right]Q_t : T \geq 1\right\}$ minus its mean is stochastically equicontinuous at $\pi = \pi_0$ with ρ_Π given by (2.2.6).

(f) $\left\{[\epsilon_t + \pi_{10}(Z_t) - \pi_1(Z_t) + X_t'(\theta_0 - \theta) + \pi_2(Z_t)'\theta - \pi_{20}(Z_t)'\theta_0][X_t - \pi_2(Z_t)]Q_t' : t \geq 1\right\}$ satisfies a uniform WLLN over $(\theta,\pi) \in \Theta_0 \times \Pi$, where Θ_0 denotes some neighborhood of θ_0.

(g) $\hat{\Sigma}_{11} \xrightarrow{p} \Sigma_{11}$, $\hat{\Sigma}_{13} \xrightarrow{p} \Sigma_{13}$, and $\hat{\Sigma}_{33} \xrightarrow{p} \Sigma_{33}$.

(h) Assumption PLO(h) holds.

THEOREM PLH: *Suppose Assumption* PLH *holds for model* (3.1.19)–(3.1.20) *under* $H_0 : \gamma_0 = \underset{\sim}{0}$. *Then,* $\mathrm{PLH}_T \xrightarrow{d} \chi_m^2$.

COMMENTS: 1. Assumption PLH (a) requires that

$$\sqrt{T}(\hat{\sigma}^2 - \sigma_0^2) = \frac{1}{\sqrt{T}}\Sigma_1^T(\epsilon_t^2 - \sigma_0^2) + o_p(1). \tag{3.1.26}$$

Assumptions PLH (b)–(d) plus the following assumptions are sufficient for (3.1.26) using a similar argument to that given in the proof of Lemma 1:

(i) $\left\{\frac{1}{\sqrt{T}}\Sigma_1^T[Y_t - \pi_1(Z_t) - (X_t - \pi_2(Z_t))'\theta_0]^2 : T \geq 1\right\}$ minus its mean is stochastically equicontinuous at $\pi = \pi_0$ with respect to the pseudo–metric $\rho_*(\cdot,\cdot)$ on Π, where $\rho_*(\pi_a, \pi_b) = \left[E\|m_t(\theta_0, \pi_a) - m_t(\theta_0, \pi_b)\|^2\right]^{1/2}$ and

$$m_t(\theta_0, \pi) = [Y_t - \pi_1(Z_t) - (X_t - \pi_2(Z_t))'\theta]^2, \tag{3.1.27}$$

(ii) $\left\{\frac{1}{T}\Sigma_1^T[Y_t - \pi_1(Z_t) - (X_t - \pi_2(Z_t))'\theta][X_t - \pi_2(Z_t)] : t \geq 1\right\}$ satisfies a uniform WLLN over $\Theta_0 \times \Pi$, and

(iii) $\sqrt{T}(\hat{\theta} - \theta_0) = O_p(1)$.

2. If the underlying rv's are iid and ϵ_t is independent of Q_t, then a consistent estimator of Φ defined in equation (3.1.24) under $H_0 : \gamma_0 = \underset{\sim}{0}$ is as follows:

$$\hat{\Phi} = \left[\frac{1}{T}\Sigma_1^T\left[\hat{U}_t^2 - \hat{\sigma}^2\right]^2\right]\left[\frac{1}{T}\Sigma_1^T Q_t Q_t' - \left[\frac{1}{T}\Sigma_1^T Q_t\right]\left[\frac{1}{T}\Sigma_1^T Q_t'\right]\right] . \qquad (3.1.28)$$

In this case, the test statistic PLH_T is equal to TR^2, where R^2 is the coefficient of determination from a regression of $\hat{U}_t^2$ against unity and Q_t.

Now, to analyze the local power of the test PLH_T, consider a sequence of local alternatives $H_T : \gamma_0 = \mu/\sqrt{T}$ for some $\mu \in R^m$, $\mu \neq \underset{\sim}{0}$. Under H_T, Assumption PLH (a) is violated generally, but Assumption 1–ℓp holds under suitable conditions with $\beta_T = (\underset{\sim}{0}', 2\sigma_0^2 h'(0)EQ_t'\mu)'$ and $R = (\underset{\sim}{0} \vdots -EQ_t)$. Assumption 1(c) of Section 2 also fails to hold under H_T. If Assumption PLH (c) holds and $E\|Q_t\|^2 < \infty$ under H_T, then

$$\sqrt{T}\, \bar{r}_T^*(\beta_0, \hat{\pi}) \xrightarrow{p} 2h'(0)\sigma_0^2[EQ_tQ_t']\mu \equiv \xi \text{ (say)}. \qquad (3.1.29)$$

Therefore, the noncentrality parameter of the limit distribution of PLH_T under H_T is

$$\delta_{PLH}^2 = 4\sigma_0^4 h'(0)^2\mu' E(Q_t - EQ_t)(Q_t - EQ_t)'\Phi^{-1}E(Q_t - EQ_t)(Q_t - EQ_t)'\mu , \qquad (3.1.30)$$

which is positive provided Φ and $E(Q_t - EQ_t)(Q_t - EQ_t)'$ are nonsingular.

Under the fixed alternative hypothesis $H_1 : \gamma_0 \neq \underset{\sim}{0}$, $\hat{\theta}$ and $\hat{\pi}$ are consistent for θ_0 and π_0, respectively, but $\hat{\sigma}^2 \xrightarrow{p} \sigma_0^2 E(1 + h(Q_t'\gamma_0))^2 \equiv \bar{\sigma}^2$ (say). Let $\bar{\beta} = (\theta_0', \bar{\sigma}^2)'$. The test PLH_T is consistent against H_1 under assumptions that imply Assumption 4 whenever

$$r(\bar{\beta}, \pi_0) = Er_t(\bar{\beta}, \pi_0) = \sigma_0^2\left[E(1 + h(Q_t'\gamma_0))^2 Q_t - E(1 + h(Q_t'\gamma_0))^2 EQ_t\right] \qquad (3.1.31)$$

does not equal zero.

3.1.4. A Test of Autocorrelation

Consider the PLR model described in equation (3.1.1) with errors $\{U_t\}$ that satisfy

$$U_t = \gamma_0 U_{t-1} + \epsilon_t \text{ for some } \gamma_0 \in R, \tag{3.1.32}$$

where $|\gamma_0| < 1$ and $\{\epsilon_t\}$ are iid with mean zero, variance σ_0^2, and are independent of $\{X_t, Z_t\}$. The null hypothesis of interest is $H_0 : \gamma_0 = 0$. Let $\hat{U}_t = Y_t - \hat{\pi}_1(Z_t) - (X_t - \hat{\pi}_2(Z_t))'\hat{\beta}$ be the semiparametric residual with $\hat{\beta}$ defined as in equation (3.1.3). To test $H_0 : \gamma_0 = 0$, we can consider the sample covariance between $\hat{U}_t$ and $\hat{U}_{t-1}$ as the basis of our test statistic. Specifically, define $W_t = (Y_t, Z_t', X_t', Y_{t-1}, Z_{t-1}', X_{t-1}')'$, $\pi = (\pi_1, \pi_2')'$ and

$$\bar{r}_T(\hat{\beta},\hat{\pi}) = \tfrac{1}{T}\Sigma_1^T \hat{U}_t \hat{U}_{t-1}. \tag{3.1.33}$$

When $\hat{U}_t$ is the OLS residual in a linear model, the statistic defined in (3.1.33) is the basis of the test statistic of Pagan and Hall (1983).

The definition in (3.1.33) corresponds to

$$r_t(\beta,\pi) = [Y_t - \pi_1(Z_t) - (X_t - \pi_2(Z_t))'\beta][Y_{t-1} - \pi_1(Z_{t-1}) - (X_{t-1} - \pi_2(Z_{t-1}))'\beta]. \tag{3.1.34}$$

This choice of the criterion function fits into our general framework since, under $H_0 : \gamma_0 = 0$, $\epsilon_t = U_t$ and hence $Er_t(\beta_0, \pi_0) = E\epsilon_t\epsilon_{t-1} = 0$.

Let

$$\Phi = \sigma_0^4 \text{ and } \hat{\Phi} = \left[\tfrac{1}{T}\Sigma_1^T \hat{U}_t^2\right]^2. \tag{3.1.35}$$

The test statistic is defined as

$$PLA_T = \left[\tfrac{1}{T}\Sigma_1^T \hat{U}_t^2\right]^{-2}\left[\tfrac{1}{\sqrt{T}}\Sigma_1^T \hat{U}_t \hat{U}_{t-1}\right]^2. \tag{3.1.36}$$

ASSUMPTION PLA: (a) Assumption PLO (a) holds.

(b), (c) Assumptions PLO (b) and (c) hold with $j = 1, 2, 3$ replaced by $j = 1, 2$ and with $E\|Q_t\|^4 < \infty$ deleted.

(d) $\{(X_t', Z_t')' : t \geq 1\}$ is a sequence of iid rv's or identically distributed, strong mixing rv's with mixing numbers that satisfy $\Sigma_{s=1}^{\infty}\alpha(s)^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$ such that $E|\epsilon_t|^{2+\delta} < \infty$ and $E\|X_t\|^{2+\delta} < \infty$. $\{\epsilon_t\}$ are iid with mean zero, variance σ_0^2, and are independent of $\{X_t, Z_t\}$.

(e) $\left\{\frac{1}{\sqrt{T}}\Sigma_1^T[\epsilon_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\beta_0][\epsilon_{t-1} + \pi_{10}(Z_{t-1}) - \pi_1(Z_{t-1}) + (\pi_2(Z_{t-1}) - \pi_{20}(Z_{t-1})'\beta_0] : T \geq 1\right\}$ minus its mean is stochastically equicontinuous at $\pi = \pi_0$ with ρ_{Π} defined by (2.2.6).

(f) $\{[\epsilon_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (X_t - \pi_{20}(Z_t))'\beta_0 - (X_t - \pi_2(Z_t))'\beta][X_{t-1} - \pi_2(Z_{t-1})] + [\epsilon_{t-1} + \pi_{10}(Z_{t-1}) - \pi_1(Z_{t-1}) + (X_{t-1} - \pi_{20}(Z_{t-1}))'\beta_0 - (X_{t-1} - \pi_2(Z_{t-1}))'\beta] \times [X_t - \pi_2(Z_t)] : t \geq 1\}$ satisfies a uniform WLLN over $(\beta,\pi) \in B_0 \times \Pi$.

(g) $\hat{\sigma}^2 \xrightarrow{p} \sigma_0^2$.

THEOREM PLA: *Suppose Assumption* PLA *holds for the model* (3.1.1)–(3.1.32) *under* $H_0 : \gamma_0 = 0$. *Then,* $\text{PLA}_T \xrightarrow{d} \chi_1^2$.

COMMENT: We can easily extend the above results to tests for autocorrelation of the form $U_t = \gamma_0 U_{t-j} + \epsilon_t$ for $j > 1$.

The test PLA_T can be shown using Theorem 3 to be consistent against the fixed alternative hypothesis $H_1 : \gamma_0 \neq 0$ because $\hat{\beta} \xrightarrow{p} \beta_0$ and $\hat{\pi} \xrightarrow{p} \pi_0$ under H_1, so that

$$r(\beta_0, \pi_0) = Er_t(\beta_0, \pi_0) = EU_tU_{t-1} \neq 0 . \qquad (3.1.37)$$

3.2. Specification Tests for Semiparametric Sample Selection Models

3.2.1. Introduction

This section considers tests of omitted variables and heteroskedasticity in semiparametric sample selection models. The model is given by

$$\tilde{Y}_t = \tilde{X}_t'\theta_0 + U_t \text{ and } D_t = 1(h(Z_t, \alpha_0) + \epsilon_t > 0) , \qquad (3.2.1)$$

where $(Y_t, D_t, X_t, Z_t) = (\tilde{Y}_tD_t, D_t, \tilde{X}_tD_t, Z_t)$ are observed for $t = 1, \ldots, T$, $\tilde{Y}_t$ is unobserved when $D_t = 0$, $\tilde{X}_t$ may or may not be observed when $D_t = 0$, the real function $h(\cdot,\cdot)$ is known, $\{(U_t, \epsilon_t, X_t, Z_t) : t \geq 1\}$ are iid, and (U_t, ϵ_t) is independent of $(\tilde{X}_t, Z_t)$ and has unknown distribution. The first equation of model (3.2.1) multiplied by D_t can be re–written as

$$\begin{aligned} & Y_t = X_t' \theta_0 + D_t g(h(Z_t, \alpha_0)) + \mu_t , \text{ where} \\ & g(v) = E(U_t \mid \epsilon_t > -v) , \ \mu_t = D_t(U_t - g(h(Z_t, \alpha_0))) , \text{ and} \\ & E(\mu_t \mid D_t = 1, X_t, Z_t) = 0 \text{ a.s.} \end{aligned} \tag{3.2.2}$$

The function $g(\cdot) : R \to R$ is unknown, since (U_t, ϵ_t) has unknown distribution.

Define

$$\begin{aligned} & \pi_{10}(\alpha,v) = E[Y_t \mid h(Z_t, \alpha) = v, D_t = 1] , \\ & \pi_{20}(\alpha,v) = E[X_t \mid h(Z_t, \alpha) = v, D_t = 1] , \text{ and} \\ & \pi_{30}(\alpha,v) = E[\mu_t^2 \mid h(Z_t, \alpha) = v, D_t = 1] . \end{aligned} \tag{3.2.3}$$

Let $\hat{\pi}_j(\cdot,\cdot)$ denote an estimator of $\pi_{j0}(\cdot,\cdot)$ for $j = 1, 2, 3$. Note that the first equation of the model (3.2.2) can be re–written as

$$Y_t - \pi_{10}(\alpha_0, h(Z_t, \alpha_0)) = [X_t - \pi_{20}(\alpha_0, h(Z_t, \alpha_0))]' \theta_0 + \mu_t . \tag{3.2.4}$$

Define $\beta_0 = (\alpha_0', \theta_0')' \in R^{q_1} \times R^{q_2} \equiv R^q$. Let $\hat{\alpha}$ be some preliminary estimator of α_0 . Below we assume that $\hat{\alpha}$ is Ichimura's (1985) or Klein and Spady's (1987) semiparametric estimator of α_0 . Specifically, $\hat{\alpha}$ is defined to minimize

$$\frac{1}{T}\Sigma_1^T \eta(D_t, \hat{\pi}_5(\alpha, h(Z_t, \alpha))) \tag{3.2.5}$$

over $\alpha \in \mathcal{A} \subset R^{q_1}$. Here, $\hat{\pi}_5(\cdot,\cdot)$ is an estimator of $\pi_{50}(\cdot,\cdot)$, $\pi_{50}(\cdot,\cdot)$ is defined by

$$\pi_{50}(\alpha,v) = E(D_t \mid h(Z_t, \alpha) = v) \text{ for } v \in R , \tag{3.2.6}$$

and $\eta(\cdot,\cdot)$ is some "distance" function. Ichimura's (1985) estimator takes

$$\eta(D_t, \pi_5) = (D_t - \pi_5)^2/2 \tag{3.2.7}$$

and Klein and Spady's (1987) estimator takes

$$\eta(D_t, \pi_5) = -D_t \ln \pi_5 - (1 - D_t)\ln(1 - \pi_5) . \tag{3.2.8}$$

As an estimator of θ_0, we consider the three-step estimator of Andrews (1990a). The three-step estimator $\hat{\theta}$ of θ_0 is defined by

$$\begin{aligned}\hat{\theta} = &\left[\Sigma_1^T[X_t - \hat{\pi}_2(\hat{\alpha}, h(Z_t, \hat{\alpha}))][X_t - \hat{\pi}_2(\hat{\alpha}, h(Z_t, \hat{\alpha}))]'/\hat{\pi}_3(\hat{\alpha}, h(Z_t, \hat{\alpha}))\right]^{-1} \\ &\times \left[\Sigma_1^T[X_t - \hat{\pi}_2(\hat{\alpha}, h(Z_t, \hat{\alpha}))][Y_t - \hat{\pi}_1(\hat{\alpha}, h(Z_t, \hat{\alpha}))]/\hat{\pi}_3(\hat{\alpha}, h(Z_t, \hat{\alpha}))\right] .\end{aligned} \tag{3.2.9}$$

Andrews (1990a) gives high-level conditions under which $\hat{\theta}$ is $\sqrt{T}$-consistent and asymptotically normal. Note that Powell's (1987) and Newey's (1988) two-step estimators equal the three-step estimator $\hat{\theta}$ when the latter is defined with $\hat{\pi}_3(\hat{\alpha}, h(Z_t, \hat{\alpha})) = 1$. Powell uses (higher order bias reducing) kernel estimators to estimate π_{10} and π_{20}, whereas Newey uses series estimators.

3.2.2. A Test of Omitted Variables

Suppose the first equation of model (3.2.1) is a potentially misspecified version of

$$\tilde{Y}_t = \tilde{X}_t'\theta_0 + \tilde{Q}_t'\gamma_0 + U_t^* , \tag{3.2.10}$$

where $Q_t = \tilde{Q}_t D_t \in R^m$ is observed and $\tilde{Q}_t$ may or may not be observed when $D_t = 0$ for $t = 1, \ldots, T$. We assume that (U_t^*, ϵ_t) is independent of $(\tilde{X}_t, Z_t, \tilde{Q}_t)$. Define

$$\pi_{40}(\alpha, v) = E[Q_t | h(Z_t, \alpha) = v, D_t = 1] \tag{3.2.11}$$

and $\hat{\pi}_4(\cdot,\cdot)$ to be an estimator of $\pi_{40}(\cdot,\cdot)$. The above model can be re-written as

$$\begin{aligned}Y_t - \pi_{10}(\alpha_0, h(Z_t, \alpha_0)) = &[X_t - \pi_{20}(\alpha_0, h(Z_t, \alpha_0))]'\theta_0 \\ &+ [Q_t - \pi_{40}(\alpha_0, h(Z_t, \alpha_0))]'\gamma_0 + \mu_t^* ,\end{aligned} \tag{3.2.12}$$

where $\mu_t^* = D_t[U_t^* - E[U_t^* | \epsilon_t > -h(Z_t, \alpha_0)]]$ and $E[\mu_t^* | D_t = 1, X_t, Z_t, Q_t] = 0$ a.s.

To develop a test of $H_0 : \gamma_0 = \underset{\sim}{0}$, we adopt the strategy of Pagan and Hall (1983) as in Section 3.1.2. As the basis of our test statistic, we consider the sample covariance between the normalized "semiparametric residual" $[Y_t - \hat{\pi}_1(\hat{\alpha}, h(Z_t, \hat{\alpha})) - (X_t - \hat{\pi}_2(\hat{\alpha}, h(Z_t, \hat{\alpha})))'\hat{\theta}]/[\hat{\pi}_3(\hat{\alpha}, h(Z_t, \hat{\alpha}))]^{1/2}$ and the normalized "left–out regressor" $[Q_t - \hat{\pi}_4(\hat{\alpha}, h(Z_t, \hat{\alpha}))]/[\hat{\pi}_3(\hat{\alpha}, h(Z_t, \hat{\alpha}))]^{1/2}$. (One could set $\hat{\pi}_3(\hat{\alpha}, h(Z_t, \hat{\alpha})) = 1$, but the normalization by $[\hat{\pi}_3(\hat{\alpha}, h(Z_t, \hat{\alpha}))]^{1/2}$ makes the residuals homoskedastic under the null hypothesis H_0.) Specifically, we define $W_t = (Y_t, D_t, X_t', Z_t', Q_t')'$, $\beta = (\alpha', \theta')'$, $\pi = (\pi_1, \pi_2', \pi_3, \pi_4')'$, and

$$r_t(\beta,\pi) = D_t[Y_t - \pi_1(\alpha, h(Z_t, \alpha)) - (X_t - \pi_2(\alpha, h(Z_t, \alpha))'\theta] \times [Q_t - \pi_4(\alpha, h(Z_t, \alpha))]/\pi_3(\alpha, h(Z_t, \alpha)) . \tag{3.2.13}$$

Note that under $H_0 : \gamma_0 = \underset{\sim}{0}$, $U_t = U_t^*$, $\mu_t = \mu_t^*$, and hence $Er_t(\beta_0, \pi_0) = E\mu_t[Q_t - \pi_{40}(\alpha_0, h(Z_t, \alpha_0))] = \underset{\sim}{0}$.

Our test statistic is defined as

$$\begin{aligned} SSO_T = &\left[\frac{1}{\sqrt{T}}\Sigma_1^T D_t\hat{\mu}_t[Q_t - \hat{\pi}_4(\hat{\alpha}, h(Z_t, \hat{\alpha}))]/\hat{\pi}_3(\hat{\alpha}, h(Z_t, \hat{\alpha}))\right]' \\ &\times \hat{\Phi}^{-1}\left[\frac{1}{\sqrt{T}}\Sigma_1^T D_t\hat{\mu}_t[Q_t - \hat{\pi}_4(\hat{\alpha}, h(Z_t, \hat{\alpha}))]/\hat{\pi}_3(\hat{\alpha}, h(Z_t, \hat{\alpha}))\right] , \end{aligned} \tag{3.2.14}$$

where $\hat{\mu}_t = [Y_t - \hat{\pi}_1(\hat{\alpha}, h(Z_t, \hat{\alpha})) - (X_t - \hat{\pi}_2(\hat{\alpha}, h(Z_t, \hat{\alpha})))'\hat{\theta}]$, $\hat{\Phi} = [I_m \vdots \hat{R}]\hat{\Sigma}[I_m \vdots \hat{R}]'$ is an estimator of Φ, and $\hat{R}$, $\hat{\Sigma}$, and Φ are defined below.

We now define Φ, R, Σ, $\hat{R}$, and $\hat{\Sigma}$. Let $\eta(\cdot,\cdot)$ be as defined in (3.2.7) or (3.2.8) and let $\eta'(\cdot,\cdot)$ and $\eta''(\cdot,\cdot)$ denote the first and second partial derivative of $\eta(\cdot,\cdot)$ with respect to its second argument respectively. Let $\varphi(v) = P(D_t = 1 | h(Z_t, \alpha_0) = v)$ and $\varphi_t = \varphi(h(Z_t, \alpha_0))$. Note that $\pi_{50}(\alpha_0, \cdot) = \varphi(\cdot)$. Let $\hat{\pi}_j^{(i)}(\cdot,\cdot)$ and $\pi_{j0}^{(i)}(\cdot,\cdot)$ denote the partial derivatives of $\hat{\pi}_j(\cdot,\cdot)$ and $\pi_{j0}(\cdot,\cdot)$, respectively, with respect to their i–th argument for $i = 1, 2$ and $j = 1, 2, 3, 4, 5$. Let $\hat{\pi}_j^{(0)}(\cdot,\cdot)$ and $\pi_{j0}^{(0)}(\cdot,\cdot)$ denote $\hat{\pi}_j(\cdot,\cdot)$ and $\pi_{j0}(\cdot,\cdot)$, respectively for $j = 1, 2, 3, 4, 5$. Let

$$H_t = \left[\frac{\partial}{\partial\alpha}h(Z_t, \alpha_0) - E\left[\frac{\partial}{\partial\alpha}h(Z_t, \alpha_0)\middle| h(Z_t, \alpha_0)\right]\right]\frac{\partial}{\partial h}\varphi(h(Z_t, \alpha_0)) , \tag{3.2.15}$$

where $E\left[\frac{\partial}{\partial\alpha}h(Z_t, \alpha_0)\middle| h(Z_t, \alpha_0)\right]$ denotes $E\left[\frac{\partial}{\partial\alpha}h(Z_t, \alpha_0)\middle| h(Z_t, \alpha_0) = v\right]$ evaluated at $v = h(Z_t, \alpha_0)$. The expression for H_t can be shown to equal $\frac{\partial}{\partial\alpha}\pi_{50}(\alpha_0, h(Z_t, \alpha_0))$.

When estimating Φ (defined below), the appropriate sample analogue of H_t to use is $\hat{\pi}_5^{(1)}(\hat{\alpha}, h(Z_t, \hat{\alpha})) + \frac{\partial}{\partial\alpha}h(Z_t, \hat{\alpha})\hat{\pi}_5^{(2)}(\hat{\alpha}, h(Z_t, \hat{\alpha}))$. Let

$$\Phi = [I_m \vdots R]\Sigma[I_m \vdots R]'$$

$$= \Sigma_{11} + R_2M_4^{-1}R_2' + (R_1 - R_2M_4^{-1}M_3)M_1^{-1}S_1M_1^{-1}(R_1 - R_2M_4^{-1}M_3)',$$

where $R = [R_1 \vdots R_2]$, $M = \begin{bmatrix} M_1 & 0 \\ M_3 & M_4 \end{bmatrix}$, $S = \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}$,

$$\Sigma = \begin{bmatrix} \Sigma_{11} & 0 & R_2M_4^{-1} \\ 0 & M_1^{-1}S_1M_1^{-1} & -M_1^{-1}S_1M_1^{-1}M_3'M_4^{-1} \\ M_4^{-1}R_2' & -M_4^{-1}M_3M_1^{-1}S_1M_1^{-1} & M_4^{-1}(S_2 + M_3M_1^{-1}S_1M_1^{-1}M_3')M_4^{-1} \end{bmatrix},$$

$$\Sigma_{11} = ED_t[Q_t - \pi_{40}(\alpha_0, h(Z_t,\alpha_0))][Q_t - \pi_{40}(\alpha_0, h(Z_t,\alpha_0))]'/\pi_{30}(\alpha_0, h(Z_t,\alpha_0)) ,$$

$$R_1 = ED_t[Q_t - \pi_{40}(\alpha_0, h(Z_t,\alpha_0))][\theta_0'\pi_{20}^{(2)}(\alpha_0, h(Z_t,\alpha_0)) - \pi_{10}^{(2)}(\alpha_0, h(Z_t,\alpha_0))]\frac{\partial}{\partial\alpha'}h(Z_t,\alpha_0)/\pi_{30}(\alpha_0, h(Z_t,\alpha_0)) , \quad (3.2.16)$$

$$R_2 = -ED_t[Q_t - \pi_{40}(\alpha_0, h(Z_t,\alpha_0))][X_t - \pi_{20}(\alpha_0, h(Z_t,\alpha_0))]'/\pi_{30}(\alpha_0, h(Z_t,\alpha_0)) ,$$

$$M_1 = E\eta''(D_t, \varphi_t)H_tH_t' , \quad S_1 = E[\eta'(D_t, \varphi_t)]^2H_tH_t' ,$$

$$M_3 = ED_t[X_t - \pi_{20}(\alpha_0, h(Z_t,\alpha_0))][\theta_0'\pi_{20}^{(2)}(\alpha_0, h(Z_t,\alpha_0)) - \pi_{10}^{(2)}(\alpha_0, h(Z_t,\alpha_0))]\frac{\partial}{\partial\alpha'}h(Z_t,\alpha_0)/\pi_{30}(\alpha_0, h(Z_t,\alpha_0)) , \text{ and}$$

$$M_4 = -ED_t[X_t-\pi_{20}(\alpha_0,h(Z_t,\alpha_0))][X_t-\pi_{20}(\alpha_0,h(Z_t,\alpha_0))]'/\pi_{30}(\alpha_0,h(Z_t,\alpha_0)) = -S_2 .$$

The estimators $\hat{R}$ and $\hat{\Sigma}$ are defined by taking sample analogues of the quantities above.

The estimators $\hat{\pi}_1(\hat{\alpha},\cdot)$, $\hat{\pi}_2(\hat{\alpha},\cdot)$, and $\hat{\pi}_4(\hat{\alpha},\cdot)$ can be obtained by nonparametric regressions of Y_t , X_t , and Q_t on $h(Z_t, \hat{\alpha})$ using the observations for which $D_t = 1$. Note that $\pi_{30}(\alpha,v) = E[(Y_t - X_t'\theta_0)^2 | h(Z_t, \alpha) = v, D_t = 1] - (\pi_{10}(\alpha,v) - \pi_{20}(\alpha,v)'\theta_0)^2$. Thus, one can take $\hat{\pi}_3(\hat{\alpha},\cdot) = \hat{\pi}_{3a}(\hat{\alpha},\cdot) - (\hat{\pi}_1(\hat{\alpha},\cdot) - \hat{\pi}_2(\hat{\alpha},\cdot)'\tilde{\theta})^2$, where $\hat{\pi}_{3a}(\hat{\alpha},\cdot)$ is obtained by a nonparametric regression of $(Y_t - X_t'\tilde{\theta})^2$ on $h(Z_t, \hat{\alpha})$ using the observations for which $D_t = 1$ and $\tilde{\theta}$ is some consistent estimator of θ_0 . The most convenient choice for $\tilde{\theta}$ is just the two-step estimator of θ_0, which is the LS estimator of θ_0 from the regression of $Y_t - \hat{\pi}_1(\hat{\alpha}, h(Z_t, \hat{\alpha}))$ on $X_t - \hat{\pi}_2(\hat{\alpha}, h(Z_t, \hat{\alpha}))$.

We impose the following assumption under the null hypothesis $H_0 : \gamma_0 = \underset{\sim}{0}$.

ASSUMPTION SSO: (a) $\sqrt{T}(\hat{\beta} - \beta_0) = \sqrt{T}\begin{bmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\theta} - \theta_0 \end{bmatrix} = \frac{1}{\sqrt{T}} \Sigma_1^T \begin{bmatrix} \psi_{1Tt}(\alpha_0) \\ \psi_{2Tt}(\beta_0) \end{bmatrix} + o_p(1)$,

where $\psi_{1Tt}(\alpha_0) = -M_1^{-1}\eta'(D_t, \varphi_t)H_t$ and $\psi_{2Tt}(\beta_0) = -M_4^{-1}[M_3\psi_{1Tt}(\alpha_0)$ $+ \mu_t[X_t - \pi_{20}(\alpha_0, h(Z_t,\alpha_0))]/\pi_{30}(\alpha_0, h(Z_t,\alpha_0))]$.

(b) (i) $P(\hat{\pi} \in \Pi) \to 1$. (ii) $\int 1(D = 1)\|\hat{\pi}_j^{(i)}(\alpha_0, v) - \pi_{j0}^{(i)}(\alpha_0, v)\|^k dP(D,v) \xrightarrow{p} 0$ for (i,j,k) $= (0,1,4)$, $(0,2,4)$, $(0,3,8)$, $(0,4,4)$, $(1,1,2)$, $(1,2,2)$, $(1,3,4)$, $(1,4,2)$, $(2,1,4)$, $(2,2,4)$, $(2,3,8)$, and $(2,4,4)$, $E\|\pi_{j0}^{(1)}(\alpha_0, h(Z_t,\alpha_0))\|^4 < \infty$ for $j = 1, 2, 3, 4$, and $E\|\pi_{j0}^{(2)}(\alpha_0, h(Z_t,\alpha_0))\|^8$ $< \infty$ for $j = 1, 2, 3, 4$, where $P(\cdot,\cdot)$ denotes the distribution of $(D_t, h(Z_t, \alpha_0))$.

(iii) $E\|X_t\|^4 < \infty$, $E\|Q_t\|^8 < \infty$, $E\left\|\frac{\partial}{\partial\alpha}h(Z_t, \alpha_0)\right\|^8 < \infty$, $ED_tU_t^8 < \infty$, $\inf_{v\in\mathcal{V}}|\pi_{30}(\alpha_0, v)|$ > 0, and $\inf_{v\in\mathcal{V}}|\hat{\pi}_3(\alpha_0, v)| \geq \epsilon$ with probability $\to 1$ for some $\epsilon > 0$, where $\mathcal{V}$ denotes the support of $h(Z_t, \alpha_0)$.

(c) $T^{1/4}\left[\int 1(D = 1)\|\hat{\pi}_j(\alpha_0, v) - \pi_{j0}(\alpha_0, v)\|^2 dP(D,v)\right]^{1/2} \xrightarrow{p} 0$ for $j = 1, 2, 4$.

(d) (i) $\{(U_t, \epsilon_t, \tilde{X}_t, Z_t, \tilde{Q}_t) : t \geq 1\}$ are iid and (U_t, ϵ_t) is independent of $(\tilde{X}_t, Z_t, \tilde{Q}_t)$.

(ii) $\inf_{v\in\mathcal{V}}|\varphi(v)| > 0$ and $\sup_{v\in\mathcal{V}}|\varphi'(v)| < \infty$, where $\varphi'(\cdot)$ denotes the derivative of $\varphi(\cdot)$.

(e) $\Big\{\frac{1}{\sqrt{T}}\Sigma_1^T D_t[\mu_t + \pi_{10}(\alpha_0, h(Z_t, \alpha_0)) - \pi_1(\alpha_0, h(Z_t, \alpha_0)) + (\pi_2(\alpha_0, h(Z_t, \alpha_0))$ $- \pi_{20}(\alpha_0, h(Z_t, \alpha_0)))'\theta_0][Q_t - \pi_4(\alpha_0, h(Z_t, \alpha_0))]/\pi_3(\alpha_0, h(Z_t, \alpha_0)) : T \geq 1\Big\}$ minus its mean is stochastically equicontinuous at $\pi = \pi_0$ with ρ_Π defined by (2.2.6).

(f) $\left\{\frac{\partial}{\partial\beta'}r_t(\beta,\pi) : t \geq 1\right\}$ satisfies a uniform WLLN over $B_0 \times \Pi$.

(g) (i) $h(z,\alpha)$ and $\frac{\partial}{\partial\alpha}h(z,\alpha)$ are continuous at α_0 uniformly over $z \in \mathcal{Z}$, where $\mathcal{Z}$ denotes the support of Z_t. (ii) $\pi_j^{(i)}(\alpha,v)$ is uniformly continuous on $\{(\alpha_0, v) : v \in \mathcal{V}\}$ uniformly over $\pi \in \Pi$ for $i = 0, 1, 2$ and $j = 1, 2, 3, 4$.

(h) $\hat{\Sigma} \xrightarrow{p} \Sigma$ and Φ has full rank m.

Note that in Assumption SSO(f),

$$\frac{\partial}{\partial \beta'} r_t(\beta,\pi) = -[R'_{1t}(\beta,\pi) + R'_{2t}(\beta,\pi) + R'_{3t}(\beta,\pi) \vdots R'_{4t}(\beta,\pi)]' \text{ , where}$$

$$\begin{aligned}
R_{1t}(\beta,\pi) &= D_t[Y_t - \pi_1(\alpha, h(Z_t,\alpha)) - (X_t - \pi_2(\alpha, h(Z_t,\alpha)))'\theta] \\
&\quad \times \frac{\partial}{\partial \alpha'}\pi_4(\alpha, h(Z_t,\alpha))/\pi_3(\alpha, h(Z_t,\alpha)) \text{ ,} \\
R_{2t}(\beta,\pi) &= D_t[Q_t - \pi_4(\alpha, h(Z_t,\alpha))]\Big[\frac{\partial}{\partial \alpha'}\pi_1(\alpha, h(Z_t,\alpha)) \\
&\quad - \theta'\frac{\partial}{\partial \alpha'}\pi_2(\alpha, h(Z_t,\alpha))\Big]/\pi_3(\alpha, h(Z_t,\alpha)) \text{ ,} \\
R_{3t}(\beta,\pi) &= D_t[Y_t - \pi_1(\alpha, h(Z_t,\alpha)) - (X_t - \pi_2(\alpha, h(Z_t,\alpha)))'\theta] \\
&\quad \times [Q_t - \pi_4(\alpha, h(Z_t, \alpha))]\frac{\partial}{\partial \alpha'}\pi_3(\alpha, h(Z_t,\alpha))/\pi_3^2(\alpha, h(Z_t,\alpha)) \text{ , and} \\
R_{4t}(\beta,\pi) &= D_t[Q_t - \pi_4(\alpha, h(Z_t,\alpha))][X_t - \pi_2(\alpha, h(Z_t,\alpha))]' \text{ .}
\end{aligned} \tag{3.2.17}$$

THEOREM SSO: *Suppose Assumption* SSO *holds for model* (3.2.1)–(3.2.10) *under the null hypothesis* $H_0 : \gamma_0 = \underset{\sim}{0}$. *Then,* $SSO_T \xrightarrow{d} \chi^2_m$.

COMMENTS: 1. Assumption SSO (a) can be verified using results of Andrews (1990a).

2. It is straightforward to extend the result of Theorem SSO to the case where the underlying rv's are dependent. This extension, however, requires a more complicated expression for Φ than that given in (3.2.16).

The test SSO_T is consistent against most alternative hypotheses $H_1 : \gamma_0 \neq \underset{\sim}{0}$. Suppose $\hat{\theta} \xrightarrow{p} \bar{\theta}$ under H_1 . Let $\bar{\beta} = (\alpha_0', \bar{\theta}')'$. (Note that $\hat{\alpha}$ and $\hat{\pi}$ are still consistent for α_0 and π_0 respectively under H_1 .) Then,

$$\begin{aligned}
Er_t(\bar{\beta},\pi_0) &= \Big[ED_t[Q_t-\pi_{40}(\alpha_0,h(Z_t,\alpha_0))][X_t-\pi_{20}(\alpha_0,h(Z_t,\alpha_0))]'/\pi_{30}(\alpha_0,h(Z_t,\alpha_0))\Big](\theta_0-\bar{\theta}) \\
&\quad + \Big[ED_t[Q_t-\pi_{40}(\alpha_0,h(Z_t,\alpha_0))][Q_t-\pi_{40}(\alpha_0,h(Z_t,\alpha_0))]'/\pi_{30}(\alpha_0,h(Z_t,\alpha_0))\Big]\gamma_0 \text{ .}
\end{aligned} \tag{3.2.18}$$

Under the conditions of Theorem 3, the test SSO_T is consistent if the expression in (3.2.18) is non–zero. The latter holds provided the second term in square brackets is non-singular, except in the special case where the two summands are exactly offsetting.

3.2.3. A Test of Heteroskedasticity

Consider the sample selection model in (3.2.1). Suppose we want to test whether the variance of the error U_t is heteroskedastic conditional on some observed random variable $Q_t \in R^m$, where elements of Q_t may contain those of $\tilde{X}_t$ and Z_t. (In this section, we assume that ϵ_t is homoskedastic and independent of $(\tilde{X}_t, Z_t, Q_t)$. A test of heteroskedasticity for ϵ_t is discussed in Whang and Andrews (1990, Section 3.2.2).) Specifically, consider a function $f(Q_t, \gamma_0)$ such that $f(Q_t, 0) = 1$. Suppose

$$U_t = f(Q_t, \gamma_0)U_t^*, \tag{3.2.19}$$

where (U_t^*, ϵ_t) is independent of $(\tilde{X}_t, Z_t, Q_t)$. The null hypothesis of interest is $H_0 : \gamma_0 = \underset{\sim}{0}$. Note that if $\gamma_0 \neq \underset{\sim}{0}$, the three–step estimator $\hat{\theta}$ of θ_0 may be inconsistent since $E[\mu_t | D_t = 1, X_t, Z_t]$ is different from zero with positive probability in general.

Since U_t cannot be estimated directly given our estimation strategy for $\hat{\theta}$, it is not possible to use the same testing strategy as in the partially linear regression model (see Section 3.1.3). Here, μ_t is an "estimable" regression error. Note, however, that the conditional variance of μ_t given $D_t = 1$ and $h(Z_t, \alpha_0)$ is heteroskedastic even under the null hypothesis $H_0 : \gamma_0 = \underset{\sim}{0}$. In particular, under H_0, since (U_t, ϵ_t) is independent of Q_t under H_0, we have

$$E[\mu_t^2 | D_t = 1, h(Z_t, \alpha_0)] = E[\mu_t^2 | D_t = 1, h(Z_t, \alpha_0), Q_t] . \tag{3.2.20}$$

We note that the above conditional variance of μ_t depends only on the single index $h(Z_t, \alpha_0)$. If $\gamma_0 \neq \underset{\sim}{0}$, however, $E[\mu_t^2 | D_t = 1, h(Z_t, \alpha_0), Q_t]$ differs from $E[\mu_t^2 | D_t = 1, h(Z_t, \alpha_0)]$ and depends on Q_t. Therefore, we test for the existence of conditional heteroskedasticity of μ_t that is not attributable to $h(Z_t, \alpha_0)$ in the following way.

Let $\hat{\pi}_j(\cdot,\cdot)$, $\pi_{j0}(\cdot,\cdot)$, $\hat{\beta}$, and β_0 be as above. We consider as the basis of our test statistic the sample covariance between μ_t^2 and Q_t after projecting these variables onto the space orthogonal to $h(Z_t, \alpha_0)$. Specifically, we define

$$\begin{aligned} r_t(\beta,\pi) = D_t\Big[&[Y_t - \pi_1(\alpha, h(Z_t,\alpha)) - (X_t - \pi_2(\alpha, h(Z_t,\alpha)))'\theta]^2 \\ &- \pi_3(\alpha, h(Z_t,\alpha))\Big][Q_t - \pi_4(\alpha, h(Z_t,\alpha))]/\pi_3(\alpha, h(Z_t,\alpha))\ . \end{aligned} \tag{3.2.21}$$

(One could set the divisor $\pi_3(\alpha, h(Z_t, \alpha))$ to equal one. But, by dividing by $\pi_3(\alpha, h(Z_t, \alpha))$, we make the residuals homoskedastic under H_0.) Under $H_0 : \gamma_0 = \underset{\sim}{0}$, (U_t, ϵ_t) is independent of Q_t and $Er_t(\beta_0, \pi_0) = ED_t[\mu_t^2 - \pi_{30}(\alpha_0, h(Z_t, \alpha_0))] \times [Q_t - \pi_{40}(\alpha_0, h(Z_t, \alpha_0))]/\pi_{30}(\alpha_0, h(Z_t, \alpha_0)) = \underset{\sim}{0}$ by (3.2.20).

Our test statistic is defined to be

$$\begin{aligned} SSH_T = &\Big[\tfrac{1}{\sqrt{T}}\Sigma_1^T D_t[\hat{\mu}_t^2 - \hat{\pi}_3(\hat{\alpha}, h(Z_t,\hat{\alpha}))][Q_t - \hat{\pi}_4(\hat{\alpha}, h(Z_t,\hat{\alpha}))]/\hat{\pi}_3(\hat{\alpha}, h(Z_t,\hat{\alpha}))\Big]'\hat{\Phi}^{-1} \\ &\times \Big[\tfrac{1}{\sqrt{T}}\Sigma_1^T D_t[\hat{\mu}_t^2 - \hat{\pi}_3(\hat{\alpha}, h(Z_t,\hat{\alpha}))][Q_t - \hat{\pi}_4(\hat{\alpha}, h(Z_t,\hat{\alpha}))]/\hat{\pi}_3(\hat{\alpha}, h(Z_t,\hat{\alpha}))\Big]\ , \end{aligned} \tag{3.2.22}$$

where $\hat{\mu}_t = [Y_t - \hat{\pi}_1(\hat{\alpha}, h(Z_t,\hat{\alpha})) - (X_t - \hat{\pi}_2(\hat{\alpha}, h(Z_t,\hat{\alpha}))'\hat{\theta}]$ and $\hat{\Phi}$ is some consistent estimator of Φ, such as the sample analogue of Φ. The matrix Φ is defined as follows:

$$\begin{aligned} &\Phi = \Sigma_{11} + R_1M_1^{-1}S_1M_1^{-1}R_1'\ , \text{ where} \\ &\Sigma_{11} = ED_t[\mu_t^2 - \pi_{30}(\alpha_0, h(Z_t, \alpha_0))]^2[Q_t - \pi_{40}(\alpha_0, h(Z_t, \alpha_0))] \\ &\qquad \times [Q_t - \pi_{40}(\alpha_0, h(Z_t, \alpha_0))]'/\pi_{30}^2(\alpha_0, h(Z_t, \alpha_0))\ , \\ &R_1 = -ED_t[Q_t - \pi_{40}(\alpha_0, h(Z_t, \alpha_0))]\pi_{30}^{(2)}(\alpha_0, h(Z_t, \alpha_0)) \\ &\qquad \times \frac{\partial}{\partial\alpha'}h(Z_t, \alpha_0)/\pi_{30}(\alpha_0, h(Z_t, \alpha_0))\ , \end{aligned} \tag{3.2.23}$$

and M_1 and S_1 are as defined in (3.2.16).

We impose that the following assumption under the null hypothesis $H_0 : \gamma_0 = \underset{\sim}{0}$.

ASSUMPTION SSH: (a) Assumption SSO (a) holds.

(b) Assumption SSO (b) holds with (0,1,4), (0,2,4), (1,1,2), (1,2,2), (1,4,2), (2,1,4), (2,2,4), and (2,4,4) replaced by (0,1,8), (0,2,8), (1,1,4), (1,2,4), (1,4,4), (2,1,8), (2,2,8), and (2,4,8), respectively, and with $E\|X_t\|^4 < \infty$ replaced by $E\|X_t\|^8 < \infty$.

(c) $T^{1/8}\Big[\int 1(D = 1)\|\hat{\pi}_j(\alpha_0, v) - \pi_{j0}(\alpha_0, v)\|^4 dP(D,v)\Big]^{1/4} \xrightarrow{p} 0$ for $j = 1, 2$ and

$T^{1/4}\left[\int 1(D=1)\|\hat{\pi}_j(\alpha_0, v) - \pi_{j0}(\alpha_0, v)\|^2 dP(D,v)\right]^{1/2} \xrightarrow{p} 0$ for $j = 3, 4$.

(d) Assumption SSO (d) holds with $\tilde{Q}_t$ replaced by Q_t.

(e) $\Big\{\frac{1}{\sqrt{T}}\Sigma_1^T D_t\Big[[\mu_t + \pi_{10}(\alpha_0, h(Z_t,\alpha_0)) - \pi_1(\alpha_0, h(Z_t,\alpha_0)) + (\pi_2(\alpha_0, h(Z_t,\alpha_0)) - \pi_{20}(\alpha_0, h(Z_t,\alpha_0)))'\theta_0]^2 - \pi_3(\alpha_0, h(Z_t,\alpha_0))\Big][Q_t - \pi_{40}(\alpha_0, h(Z_t,\alpha_0))]/\pi_3(\alpha_0, h(Z_t,\alpha_0)) : T \geq 1\Big\}$ minus its mean is stochastically equicontinuous at $\pi = \pi_0$ with ρ_Π defined by (2.2.6).

(f) $\left\{\frac{\partial}{\partial\beta'}r_t(\beta,\pi) : t \geq 1\right\}$ satisfies a uniform WLLN over $B_0 \times \Pi$.

(g), (h) Assumptions SSO (g) and (h) hold.

Note that in Assumption SSH(f),

$$\frac{\partial}{\partial\beta'}r_t(\beta,\pi) = -[R_{1t}'(\beta,\pi) + R_{2t}'(\beta,\pi) + R_{3t}'(\beta,\pi) \vdots 2R_{4t}'(\beta,\pi)]', \text{ where}$$

$$\begin{aligned}
R_{1t}(\beta,\pi) &= D_t[Q_t - \pi_4(\alpha, h(Z_t,\alpha))]\Big[2[Y_t - \pi_1(\alpha, h(Z_t,\alpha)) - (X_t - \pi_2(\alpha, h(Z_t,\alpha)))'\theta] \\
&\quad \times \left[\frac{\partial}{\partial\alpha'}\pi_1(\alpha,h(Z_t,\alpha)) - \theta'\frac{\partial}{\partial\alpha'}\pi_2(\alpha,h(Z_t,\alpha))\right] - \frac{\partial}{\partial\alpha'}\pi_3(\alpha,h(Z_t,\alpha))\Big]/\pi_3(\alpha,h(Z_t,\alpha)), \\
R_{2t}(\beta,\pi) &= D_t\Big[[Y_t - \pi_1(\alpha, h(Z_t,\alpha)) - (X_t - \pi_2(\alpha, h(Z_t,\alpha)))'\theta]^2 - \pi_3(\alpha, h(Z_t,\alpha))\Big] \\
&\quad \times \frac{\partial}{\partial\alpha'}\pi_4(\alpha, h(Z_t,\alpha))/\pi_3(\alpha, h(Z_t,\alpha)), \qquad (3.2.24) \\
R_{3t}(\beta,\pi) &= D_t\Big[[Y_t - \pi_1(\alpha, h(Z_t,\alpha)) - (X_t - \pi_1(\alpha, h(Z_t,\alpha)))'\theta]^2 - \pi_3(\alpha, h(Z_t,\alpha))\Big] \\
&\quad \times [Q_t - \pi_4(\alpha, h(Z_t,\alpha))]\frac{\partial}{\partial\alpha'}\pi_3(\alpha, h(Z_t,\alpha))/\pi_3^2(\alpha, h(Z_t,\alpha)), \text{ and} \\
R_{4t}(\beta,\pi) &= D_t[Y_t - \pi_1(\alpha, h(Z_t,\alpha)) - (X_t - \pi_2(\alpha, h(Z_t,\alpha)))'\theta][Q_t - \pi_4(\alpha, h(Z_t,\alpha))] \\
&\quad \times [X_t - \pi_2(\alpha, h(Z_t,\alpha))]'/\pi_3(\alpha, h(Z_t,\alpha)).
\end{aligned}$$

THEOREM SSH: *Suppose Assumption* SSH *holds for the model* (3.2.1)–(3.2.19) *under the null hypothesis* $H_0 : \gamma_0 = \underset{\sim}{0}$. *Then,* $SSH_T \xrightarrow{d} \chi_m^2$.

Next we consider the consistency properties of the test SSH_T against the fixed alternative hypothesis $H_1 : \gamma_0 \neq \underset{\sim}{0}$. If $\hat{\beta} \xrightarrow{p} \bar{\beta} = (\alpha_0', \bar{\theta}')'$ and $\hat{\pi} \xrightarrow{p} \bar{\pi} = (\pi_{10}, \pi_{20}', \bar{\pi}_3, \pi_{40}')'$ under H_1, then the test SSH_T generally is consistent against H_1. This follows because

$$\begin{aligned}
Er_t(\bar{\beta},\bar{\pi}) = & ED_tK_{1t}[Q_t - \pi_{40}(\alpha_0, h(Z_t, \alpha_0))]/\bar{\pi}_3(\alpha_0, h(Z_t, \alpha_0)) \\
& + 2\Big[ED_tK_{2t}[Q_t - \pi_{40}(\alpha_0, h(Z_t, \alpha_0))] \\
& \times [X_t - \pi_{20}(\alpha_0, h(Z_t, \alpha_0))]'/\bar{\pi}_3(\alpha_0, h(Z_t, \alpha_0))\Big](\theta_0 - \bar{\theta}) \\
& + ED_t[X_t - \pi_{20}(\alpha_0, h(Z_t, \alpha_0))'(\theta_0 - \bar{\theta})]^2 \\
& \times [Q_t - \pi_{40}(\alpha_0, h(Z_t, \alpha_0))]/\bar{\pi}_3(\alpha_0, h(Z_t, \alpha_0))
\end{aligned} \tag{3.2.25}$$

is not equal to zero generally, where

$$\begin{aligned}
K_{1t} = & f^2(Q_t,\gamma_0)E[U_t^{*2}|D_t = 1, h(Z_t,\alpha_0)] - E[f^2(Q_t,\gamma_0)U_t^{*2}|D_t = 1, h(Z_t,\alpha_0)] \\
& + \Big[E[f(Q_t,\gamma_0)U_t^*|D_t = 1, h(Z_t,\alpha_0)]\Big]^2 - \Big[f(Q_t,\gamma_0)E[U_t^*|D_t = 1, h(Z_t,\alpha_0)]\Big]^2
\end{aligned} \tag{3.2.26}$$

$$K_{2t} = f(Q_t,\gamma_0)E[U_t^*|D_t = 1, h(Z_t,\alpha_0)] - E[f(Q_t,\gamma_0)U_t^*|D_t = 1, h(Z_t,\alpha_0)] . \tag{3.2.27}$$

Note that K_{1t} and K_{2t} are zero a.s. if $f(Q_t, \gamma_0)$ is a function of $h(Z_t, \alpha_0)$.

4. TESTS OF PARAMETRIC VERSUS SEMIPARAMETRIC MODELS

4.1. First–Order Conditions Based Tests

In this section, we consider a class of specification tests that test parametric assumptions in parametric models against semiparametric alternatives. The method is as follows: Suppose we have a parametric model that is nested in a semiparametric model. We base a test of the parametric assumptions on the vector of first–order conditions (FOC) that define a semiparametric estimator $\tilde{\beta}$ evaluated at a parametric estimator $\hat{\beta}$. That is, we take $r_t(\beta,\pi)$ to equal the defining equation of the semiparametric estimator $\tilde{\beta}$ and evaluate β at the parametric estimator $\hat{\beta}$. If the parametric assumptions used to obtain the consistency of $\hat{\beta}$ are true, then the FOC evaluated at $\hat{\beta}$ will be close to zero and the test will reject with probability α in large samples. If these parametric assumptions are violated, however, the FOC evaluated at $\hat{\beta}$ will not be close to zero and the test will reject with high probability in large samples.

The idea of an FOC test is similar to that of the LM test because we can regard $\hat{\beta}$ and $\tilde{\beta}$ as "restricted" and "unrestricted" estimators of β_0 respectively. Ruud (1984), Newey (1985b), and Pagan and Vella (1989) discuss FOC tests when both $\hat{\beta}$ and $\tilde{\beta}$ are parametric estimators. Pagan and Vella (1989) also discuss FOC tests when $\tilde{\beta}$ is a semiparametric estimator. They do not provide asymptotic distribution theory for such tests, however, when infinite dimensional nuisance parameters are present in the first–order conditions and/or the first–order conditions are not differentiable with respect to β. Our results cover these cases.

The FOC test is also similar to the Hausman test (see Hausman (1978)) that is based on the difference between $\tilde{\beta}$ and $\hat{\beta}$. Whang and Andrews (1990) discuss conditions under which the latter two tests are asymptotically equivalent both under the null and local alternative hypotheses.

The semiparametric estimators $\tilde{\beta}$ that we consider are called MINPIN estimators (see Andrews (1990a)). These estimators MINimize a criterion function that may depend on a Preliminary Infinite dimensional Nuisance parameter estimator.

DEFINITION: A sequence of *MINPIN estimators* $\{\tilde{\beta}\} = \{\tilde{\beta} : T \geq 1\}$ is any sequence of rv's such that

$$d(\bar{r}_T(\tilde{\beta},\hat{\pi}),\ \hat{\gamma}) = \inf_{\beta \in B} d(\bar{r}_T(\beta,\hat{\pi}),\ \hat{\gamma}) \tag{4.1.1}$$

with probability $\to 1$, where $\bar{r}_T(\beta,\hat{\pi}) = \frac{1}{T}\Sigma_1^T r_t(\beta,\hat{\pi})$, $\hat{\gamma}$ is a random element of Γ (and $\hat{\gamma}$ depends on T in general), Γ is a pseudo–metric space, and $d(\cdot,\cdot)$ is a non–random, real valued function (which does not depend on T).

Let $\hat{\beta}$ be a $\sqrt{T}$–consistent and asymptotically normal estimator of β_0 under the parametric assumptions we want to test. We define the FOC test statistic FOC_T to be the statistic G_T of equation (2.1.4) with $r_t(\cdot,\cdot)$ as in the definition of a MINPIN estimator, $\hat{\Phi}$ defined by (2.2.16), and $\hat{\beta}$ defined as above.

In the examples given below, we consider FOC tests based on semiparametric estimators for censored regression (Section 4.2), partially linear regression (Section 4.3), and sample selection (Section 4.4) models.

4.2. A Test of Parametric Assumptions in a Censored Regression Model

In this section, we consider a test of normality and homoskedasticity of the errors in a censored regression model. The model is

$$Y_t = \max\{0, X_t'\beta_0 + U_t\} \text{ for } t = 1, \ldots, T, \tag{4.2.1}$$

where Y_t is an observed dependent variable, X_t is an observed q–vector of regressors, and U_t is an unobserved error.

Under the normality of U_t and some additional regularity conditions, Amemiya (1973) shows that the tobit MLE $\hat{\beta}$ of β_0 is $\sqrt{T}$–consistent and asymptotically normal. It is well known, however, that the tobit MLE usually is inconsistent if the assumption of normality or homoskedasticity of the errors is violated (see Goldberger (1983) and Arabmazar and Schmidt (1982)).

Powell (1984)'s censored least absolute deviations (CLAD) estimator $\tilde{\beta}$ of β_0 is a semiparametric estimator that is $\sqrt{T}$–consistent and asymptotically normal for a wide class of error distributions with a zero median (conditional on X_t) and is robust to heteroskedasticity of the errors. The latter fact suggests that we can use the FOC of the CLAD estimator $\tilde{\beta}$ evaluated at the tobit MLE $\hat{\beta}$ as the basis of our test statistic. We expect the resulting test to exhibit power when the errors are non–normal or heteroskedastic.

The CLAD estimator $\tilde{\beta}$ of β_0 is defined to minimize

$$\Sigma_1^T |Y_t - \max\{0, X_t'\beta\}| \tag{4.2.2}$$

over $\beta \in B \subset R^q$. Under suitable assumptions, it also solves the FOC:

$$\underset{\sim}{0} = \Sigma_1^T \left[1/2 - 1(Y_t - X_t'\beta < 0)\right] 1(X_t'\beta > 0) X_t \tag{4.2.3}$$

with probability $\to 1$.

Define $W_t = (Y_t, X_t')'$. Then we can write our criterion function as follows:

$$r_t(W_t, \beta) \equiv r_t(\beta) = \left[1/2 - 1(Y_t - X_t'\beta < 0)\right] 1(X_t'\beta > 0) X_t . \tag{4.2.4}$$

Note that under the null hypothesis $H_0 : U_t \sim N(0, \sigma_0^2)$, we have $Er_t(\beta_0) = \underset{\sim}{0}$.

Our test statistic is defined to be:

$$\text{CLAD}_T = T\bar{r}_T(\hat{\beta})'\hat{\Phi}^{-1}\bar{r}_T(\hat{\beta}) , \tag{4.2.5}$$

where $\bar{r}_T(\beta) = \frac{1}{T}\Sigma_1^T r_t(\beta)$, $\hat{\beta}$ is the tobit MLE defined below, and $\hat{\Phi}$ is defined below.

To define $\hat{\beta}$ and $\hat{\Phi}$, we re–write the model (4.2.1) as

$$\tilde{Y}_t = X_t'\beta_0 + U_t \text{ and } D_t = 1(X_t'\beta_0 + U_t > 0) , \tag{4.2.6}$$

where $(Y_t, D_t, X_t) = (\tilde{Y}_t D_t, D_t, X_t)$ are observed for $t = 1, \ldots, T$. Let

$$\begin{aligned} &f(u, \sigma^2) = \left[2\pi\sigma^2\right]^{-1/2} \exp[-(u/\sigma)^2/2] , F(\lambda, \sigma^2) = \int_{-\infty}^{\lambda} f(u, \sigma^2)du , \\ &f_{0t} = f(X_t'\beta_0, \sigma_0^2) , \text{ and } F_{0t} = F(X_t'\beta_0, \sigma_0^2) . \end{aligned} \tag{4.2.7}$$

The tobit MLE $(\hat{\beta}, \hat{\sigma}^2)$ of (β_0, σ_0^2) is defined to minimize the log–likelihood function:

$$\Sigma_1^T(1-D_t)\log[1 - F(X_t'\beta,\sigma^2)] - (\Sigma_1^T D_t/2)\ln 2\pi\sigma^2 - \Sigma_1^T D_t(Y_t - X_t'\beta)^2/(2\sigma^2) . \tag{4.2.8}$$

The asymptotic covariance matrix Φ of $\sqrt{T}\,\bar{r}_T(\hat{\beta})$ is defined as follows:

$$\begin{aligned} &\Phi = M/4 - f_0^2 M[I_q \vdots 0]J^{-1}[I_q \vdots 0]' , \text{ where } f_0 = f(0, \sigma_0^2) = 1/\sqrt{2\pi\sigma_0^2} , \\ &M = \lim_{T\to\infty} \tfrac{1}{T}\Sigma_1^T E1(X_t'\beta_0 > 0)X_t X_t' , \\ &J = \lim_{T\to\infty} \frac{1}{T}\begin{bmatrix} \Sigma_1^T E a_t X_t X_t' & \Sigma_1^T E b_t X_t \\ \Sigma_1^T E b_t X_t' & \Sigma_1^T E c_t \end{bmatrix} , \\ &a_t = -\left[X_t'\beta_0 f_{0t} - \sigma_0^2 f_{0t}^2/(1 - F_{0t}) - F_{0t}\right]/\sigma_0^2 , \\ &b_t = \left[(X_t'\beta_0)^2 f_{0t}/\sigma_0 + \sigma_0 f_{0t} - \sigma_0 X_t'\beta_0 f_{0t}^2/(1 - F_{0t})\right]/(2\sigma_0^3) , \text{ and} \\ &c_t = -\left[(X_t'\beta_0)^3 f_{0t}/\sigma_0^2 + X_t'\beta_0 f_{0t} - (X_t'\beta_0 f_{0t})^2/(1 - F_{0t}) - 2F_{0t}\right]/(4\sigma_0^4) . \end{aligned} \tag{4.2.9}$$

We define the estimator $\hat{\Phi}$ of Φ as follows:

$$\hat{\Phi} = \hat{M}/4 - \hat{M}[I_q \vdots 0]\hat{J}^{-1}[I_q \vdots 0]'/\left[2\pi\hat{\sigma}^2\right]^{1/2}, \tag{4.2.10}$$

where $\hat{M} = \frac{1}{T}\Sigma_1^T 1(X_t'\hat{\beta} > 0)X_tX_t'$ and $\hat{J}$ is defined to equal the sample analogue of J with (β_0, σ_0^2) replaced by the tobit MLE $(\hat{\beta}, \hat{\sigma}^2)$ everywhere they appear.

We now introduce conditions under which $CLAD_T$ has the desired asymptotic null distribution. These conditions are sufficient for Assumption 1*.

ASSUMPTION CLAD: (a) $\{U_t\}$ are iid normal rv's with mean zero and variance σ_0^2. $\{U_t\}$ and $\{X_t\}$ are independent. $\{X_t\}$ are rv's with $\sup_{t\geq 1} E\|X_t\|^3 < \infty$. M and J, defined in (4.2.9), exist.

(b) $\sqrt{T}(\hat{\beta} - \beta_0) = [I_q \vdots 0]J^{-1}\frac{1}{\sqrt{T}}\Sigma_1^T \begin{bmatrix} \psi_{1Tt}(\beta_0) \\ \psi_{2Tt}(\beta_0) \end{bmatrix} + o_p(1)$, where $\psi_{1Tt}(\beta_0) = X_t[D_tU_t/\sigma_0^2 - (1-D_t)f_{0t}/(1-F_{0t})]$ and $\psi_{2Tt}(\beta_0) = [(1-D_t)X_t'\beta_0 f_{0t}/(1-F_{0t}) - D_t + D_tU_t^2/\sigma_0^2]/2\sigma_0^2$.

(c) $\sup_{t\geq 1} E[1(|X_t'\beta| \leq \|X_t\|\cdot z)\|X_t\|^r]$ is $o(z)$ as $z \to 0$ for all $\beta \in B_0$ and $r = 1, 2$.

(d) $\lim_{T\to\infty} \frac{1}{T}\Sigma_1^T Ef(X_t'(\beta - \beta_0), \sigma_0^2)1(X_t'\beta > 0)X_tX_t'$ exists uniformly over $\beta \in B_0$.

(e) $\hat{J} \xrightarrow{p} J$ and Φ has full rank q.

THEOREM CLAD: *Suppose* H_0 *is true and Assumption* CLAD *holds for the model* (4.2.1). *Then,* $CLAD_T \xrightarrow{d} \chi_q^2$.

COMMENTS: 1. The expression $[\psi_{1Tt}(\beta_0)', \psi_{2Tt}(\beta_0)]'$ in Assumption CLAD (b) equals the score function evaluated at (β_0, σ_0^2) corresponding to the log-likelihood function (4.2.8). J equals the information matrix. Under maximum likelihood regularity conditions, Assumption CLAD (b) can be verified. (For example, Amemiya (1973) gives regularity conditions for Assumption CLAD (b) when $\{X_t\}$ are fixed constants.)

2. Assumption CLAD (c) is the same as Assumption R.2. of Powell (1984, p. 310). For a discussion of this assumption, see the latter paper.

3. It is straightforward to extend the result of Theorem CLAD to the case where $\{X_t\}$ are m–dependent. The corresponding result where $\{X_t\}$ are more generally dependent requires different arguments from those used in the proof of Theorem CLAD to verify the stochastic equicontinuity condition of Assumption 1*.

Consider a sequence of local alternatives under which $\sqrt{T}(\hat{\beta}-\beta_T)$ has a mean zero limit distribution, where $\beta_T = \beta_0 + \eta/\sqrt{T}$ for some $\eta \in R^q$. Then, the noncentrality parameter of the limit distribution of $CLAD_T$ is

$$\delta^2_{CLAD} = f_0^2 \eta' M \Phi^{-1} M \eta , \tag{4.2.11}$$

where f_0, M, and Φ are as defined in (4.2.9).

Next, suppose $\hat{\beta} \xrightarrow{p} \bar{\beta} \neq \beta_0$ when $\{U_t\}$ are not normally distributed or are heteroskedastic. The test $CLAD_T$ usually is consistent against such alternatives because

$$\lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E r_t(\bar{\beta}) = \lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E\left[\frac{1}{2} - 1(u_t < X_t'(\bar{\beta}-\beta_0))\right] 1(X_t'\bar{\beta} > 0) X_t \tag{4.2.12}$$

is usually non–zero for $\bar{\beta} \neq \beta_0$.

4.3. A Test of Linearity in the Linear Regression Model

In this section, we consider a test of the linearity assumption in the linear regression model. The test has power directed towards semiparametric alternatives. Specifically, we consider the partially linear regression (PLR) model of Section 3.1 as the alternative model.

Consider the PLR model given in equation (3.1.1). We suppose the underlying rv's $\{W_t\} = \{(Y_t, X_t', Z_t')'\}$ are identically distributed. The null hypothesis of interest is:

$$H_0 : P(f(Z_t) = Z_t'\gamma_0) = 1 \text{ for some } \gamma_0 \in R^p . \tag{4.3.1}$$

Define $\pi_{10}(\cdot)$ and $\pi_{20}(\cdot)$ as in (3.1.2). Let

$$\pi_{30}(Z_t) = E(U_t^2 | Z_t) . \tag{4.3.2}$$

Let $\hat{\pi}_j(\cdot)$ denote an estimator of $\pi_{j0}(\cdot)$ for $j = 1, 2, 3$. Let $\pi = (\pi_1, \pi_2', \pi_3)'$.

We consider the defining equation for the semiparametric weighted least squares (WLS) estimator for the PLR model (see Andrews (1990a)) as our criterion function, i.e.,

$$r_t(\beta,\pi) = [Y_t - \pi_1(Z_t) - (X_t - \pi_2(Z_t))'\beta][X_t - \pi_2(Z_t)]/\pi_3(Z_t) . \tag{4.3.3}$$

Under H_0 , $Er_t(\beta_0, \pi_0) = EU_t[X_t - \pi_{20}(Z_t)]/\pi_{30}(Z_t) = \underset{\sim}{0}$ since $E(U_t | X_t, Z_t) = 0$ a.s. Let $\hat{\beta}$ be the OLS estimator of β_0 for the null model, i.e.

$$\hat{\beta} = [I_q \vdots 0](Q'Q)^{-1}Q'Y , \tag{4.3.4}$$

where $Q = (Q_1', ..., Q_T')'$, $Q_t = (X_t', Z_t')'$, and $Y = (Y_1, \ldots, Y_T)'$. Our test statistic is based on:

$$\bar{r}_T(\hat{\beta},\hat{\pi}) = \tfrac{1}{T}\Sigma_1^T[Y_t - \hat{\pi}_1(Z_t) - (X_t - \hat{\pi}_2(Z_t))'\hat{\beta}][X_t - \hat{\pi}_2(Z_t)]/\hat{\pi}_3(Z_t) . \tag{4.3.5}$$

The test statistic is defined to be

$$LPL_T = T\bar{r}_T(\hat{\beta},\hat{\pi})'\hat{\Phi}^{-1}\bar{r}_T(\hat{\beta},\hat{\pi}) , \tag{4.3.6}$$

where $\hat{\Phi}$ is defined below.

We now define Φ and $\hat{\Phi}$. Let

$$\Phi = [I_q \vdots R]\Sigma[I_q \vdots R]' , \text{ where}$$

$$R = -E[X_t - \pi_{20}(Z_t)][X_t - \pi_{20}(Z_t)]'/\pi_{30}(Z_t) ,$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix} , \ J = \begin{bmatrix} EX_tX_t' & EX_tZ_t' \\ EZ_tX_t' & EZ_tZ_t' \end{bmatrix} = EQ_tQ_t' , \tag{4.3.7}$$

$$\Sigma_{11} = \lim_{T\to\infty}\tfrac{1}{T}\Sigma_{t=1}^T\Sigma_{s=1}^T EU_tU_s[X_t - \pi_{20}(Z_t)][X_s - \pi_{20}(Z_s)]'/(\pi_{30}(Z_t)\pi_{30}(Z_s)) ,$$

$$\Sigma_{12} = \left[\lim_{T\to\infty}\tfrac{1}{T}\Sigma_{t=1}^T\Sigma_{s=1}^T EU_tU_s[X_t - \pi_{20}(Z_t)]Q_s'/\pi_{30}(Z_t)\right]J^{-1}[I_q \vdots 0]' , \text{ and}$$

$$\Sigma_{22} = [I_q \vdots 0]J^{-1}\left[\lim_{T\to\infty}\tfrac{1}{T}\Sigma_{t=1}^T\Sigma_{s=1}^T EU_tU_sQ_tQ_s'\right]J^{-1}[I_q \vdots 0]' .$$

Let $\hat{\Sigma}$ be some estimator of Σ, as in Section 2.2.3, and let

$$\hat{\Phi} = [I_q \vdots \hat{R}]\hat{\Sigma}[I_q \vdots \hat{R}]' , \text{ where } \hat{R} = -\tfrac{1}{T}\Sigma_1^T[X_t - \hat{\pi}_2(Z_t)][X_t - \hat{\pi}_2(Z_t)]'/\hat{\pi}_3(Z_t) . \quad (4.3.8)$$

If the errors U_t and U_s are uncorrelated conditional on (X_t, Z_t, X_s, Z_s) $\forall t \neq s$ and $E(U_t^2 | X_t, Z_t) = E(U_t^2 | Z_t) = \sigma_0^2$ a.s., then the above expression for Φ simplifies to

$$\Phi = [M - MJ^{11}M]/\sigma_0^2 , \text{ where}$$
$$M = E[X_t - \pi_{20}(Z_t)][X_t - \pi_{20}(Z_t)]' , \text{ and} \quad (4.3.9)$$
$$J^{11} = \left[EX_tX_t' - (EX_tZ_t')(EZ_tZ_t')^{-1}(EZ_tX_t')\right]^{-1} .$$

In this case, the estimators $\hat{\Sigma}$ and $\hat{\Phi}$ can be simplified correspondingly.

To establish the asymptotic null distribution of LPL_T we assume:

ASSUMPTION LPL: (a) $\{(U_t, X_t, Z_t) : t \geq 1\}$ is a sequence of iid rv's or identically distributed, strong mixing rv's with mixing numbers that satisfy $\Sigma_{s=1}^{\infty} \alpha(s)^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$ such that $E|U_t|^{4+2\delta} < \infty$, $E\|X_t\|^{4+2\delta} < \infty$, and $E\|Z_t\|^{4+2\delta} < \infty$.

(b) (i) $\Pi \subset \{\pi : \pi = (\pi_1, \pi_2', \pi_3)', \inf_{z \in \mathcal{Z}} |\pi_3(z)| \geq \epsilon\}$ for some $\epsilon > 0$ and $P(\hat{\pi} \in \Pi) \to 1$, where $\mathcal{Z}$ denotes the support of Z_t. (ii) $EU_t^8 < \infty$, $E\|X_t\|^8 < \infty$.

(iii) $\int \|\hat{\pi}_j(z) - \pi_{j0}(z)\|^4 dP(z) \xrightarrow{p} 0$ for $j = 1, 2, 3$, where $P(\cdot)$ denotes the distribution of Z_t.

(c) $T^{1/4}\left[\int \|\hat{\pi}_j(z) - \pi_{j0}(z)\|^2 dP(z)\right]^{1/2} \xrightarrow{p} 0$ for $j = 1, 2$.

(d) Σ, defined in (4.3.7), exists.

(e) $\left\{\frac{1}{\sqrt{T}}\Sigma_1^T[U_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\beta_0][X_t - \pi_2(Z_t)]/\pi_3(Z_t) : T \geq 1\right\}$ minus its mean is stochastically equicontinuous at $\pi = \pi_0$ with ρ_Π defined by (2.2.6).

(f) $\{[X_t - \pi_2(Z_t)][X_t - \pi_2(Z_t)]'/\pi_3(Z_t) : t \geq 1\}$ satisfies a uniform WLLN over $\pi \in \Pi$.

(g) $\hat{\Sigma} \xrightarrow{p} \Sigma$ and Φ has full rank q.

THEOREM LPL: *Suppose Assumption* LPL *holds for the model* (3.1.1)–(4.3.1) *under* H_0. *Then,* $LPL_T \xrightarrow{d} \chi_q^2$.

Now suppose $\hat{\beta} \xrightarrow{p} \bar{\beta} \neq \beta_0$ under the fixed alternative hypothesis

$$H_1 : P(f(Z_t) \neq Z_t'\gamma) > 0 \quad \forall \gamma \in R^p . \tag{4.3.10}$$

If R is nonsingular, the test LPL_T is consistent against such alternatives, since

$$r(\bar{\beta}, \pi_0) = Er_t(\bar{\beta}, \pi_0) = R(\beta_0 - \bar{\beta}) . \tag{4.3.11}$$

4.4. A Test of Parametric Assumptions in a Sample Selection Model

Consider the sample selection model given in (3.2.1) of Section 3.2. Here, we consider a test of bivariate normality of the errors (U_t, ϵ_t) :

$$H_0 : \begin{bmatrix} U_t \\ \epsilon_t \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right] , \text{ where wlog } \sigma_{22} = 1 . \tag{4.4.1}$$

The test is based on the FOC of the semiparametric three–step estimator of Andrews (1990a) evaluated at the parametric two-step estimator of Heckman (1979). Since the latter estimator usually is inconsistent if the bivariate normality assumption is violated, the test has power against such violations.

Let $\hat{\alpha}$ be the probit estimator of $\alpha_0 \in R^{q_1}$ and let $\hat{\theta}$ be Heckman's two–step estimator of $\theta_0 \in R^{q_2}$. Let $\hat{\beta} = (\hat{\alpha}', \hat{\theta}')'$ and $\beta_0 = (\alpha_0', \theta_0')' \in R^{q_1} \times R^{q_2} = R^q$. By definition, $\hat{\alpha}$ maximizes the log–likelihood function

$$\log L(\alpha) = \Sigma_1^T \Big[D_t \log[1 - \Phi(h(Z_t, \alpha_0))] + (1 - D_t)\log \Phi(h(Z_t, \alpha_0))\Big] , \tag{4.4.2}$$

where $\Phi(\cdot)$ denotes the standard normal distribution function.

Heckman's two–step estimator $\hat{\theta}$ is defined by

$$\hat{\theta} = [I_{q_2} \vdots 0] \begin{bmatrix} \Sigma_1^T X_t X_t' & \Sigma_1^T \hat{\lambda}_t X_t \\ \Sigma_1^T \hat{\lambda}_t X_t' & \Sigma_1^T D_t \hat{\lambda}_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma_1^T X_t Y_t \\ \Sigma_1^T \hat{\lambda}_t Y_t \end{bmatrix} , \hat{\lambda}_t = \frac{\hat{\phi}_t}{1-\hat{\Phi}_t} , \text{ and } \lambda_t = \frac{\phi_t}{1-\Phi_t} , \tag{4.4.3}$$

where I_{q_2} denotes the $q_2 \times q_2$ identity matrix, $\phi(\cdot)$ denotes the standard normal density function, $\hat{\phi}_t = \phi(h(Z_t, \hat{\alpha}))$, $\hat{\Phi}_t = \Phi(-h(Z_t, \hat{\alpha})) = 1 - \Phi(h(Z_t, \hat{\alpha}))$, $\phi_t = \phi(h(Z_t, \alpha_0))$, and $\Phi_t = \Phi(-h(Z_t, \alpha_0))$.

Let $\pi_{j0}(\alpha,v)$ be as defined in (3.2.3) for $j = 1, 2, 3$, and let

$$\pi_{40}(\alpha_0, v) = P(D_t = 1 | h(Z_t, \alpha_0) = v) \text{ for } v \in R . \tag{4.4.4}$$

Let $\varphi(\cdot) = \pi_{40}(\alpha_0, \cdot)$ and $\varphi_t = \varphi(h(Z_t, \alpha_0))$. Let $\eta(\cdot,\cdot)$ be as defined in (3.2.7) or (3.2.8). The FOC of Andrews' (1990a) three–step estimator of β_0 yield the following criterion function:

$$r_t(\beta,\pi) = \begin{bmatrix} r_{1t}(\alpha, \pi_4) \\ r_{2t}(\beta, \pi) \end{bmatrix}$$
$$= \begin{bmatrix} \eta'(D_t, \pi_4(\alpha, h(Z_t, \alpha)))\frac{\partial}{\partial\alpha}\pi_4(\alpha, h(Z_t, \alpha)) \\ D_t[Y_t - \pi_1(\alpha, h(Z_t, \alpha)) - (X_t - \pi_2(\alpha, h(Z_t, \alpha)))'\theta][X_t - \pi_2(\alpha, h(Z_t, \alpha))]/\pi_3(\alpha, h(Z_t, \alpha)) \end{bmatrix}, \tag{4.4.5}$$

where $W_t = (Y_t, D_t, X_t', Z_t')'$, $\pi = (\pi_1, \pi_2', \pi_3, \pi_4)'$, and η' denotes the derivative of η with respect to its second argument. Note that $Er_t(\beta_0, \pi_0) = \underset{\sim}{0}$ since $E[\eta'(D_t, \varphi_t) | Z_t] = 0$ a.s. and $E[\mu_t | D_t = 1, X_t, Z_t] = 0$ a.s.

We define our test statistic as follows:

$$SSD_T = T\bar{r}_T(\hat{\beta},\hat{\pi})'\hat{\Phi}^{-1}\bar{r}_T(\hat{\beta},\hat{\pi}) , \tag{4.4.6}$$

where $\bar{r}_T(\hat{\beta},\hat{\pi}) = \frac{1}{T}\Sigma_1^T r_t(\hat{\beta},\hat{\pi})$, $\hat{\Phi} = [I_q \vdots \hat{R}]\hat{\Sigma}[I_q \vdots \hat{R}]'$ is an estimator of Φ (defined below), and $\hat{R}$ and $\hat{\Sigma}$ are sample analogues of R and Σ (defined below). Under sufficient conditions for Theorem 1 for the model (3.2.1)–(4.4.1) (see Whang and Andrews (1990)), the test statistic SSD_T converges in distribution to a χ^2_q rv under H_0 .

We now define Φ . Let

$\Phi = [I_q \vdots R]\Sigma[I_q \vdots R]'$, where

$$R = \begin{bmatrix} R_1 & 0 \\ R_3 & R_4 \end{bmatrix}, \ \Sigma = \begin{bmatrix} \Sigma_{11} & 0 & \Sigma_{13} & \Sigma_{14} \\ 0 & \Sigma_{22} & 0 & \Sigma_{24} \\ \Sigma_{13}' & 0 & \Sigma_{33} & \Sigma_{34} \\ \Sigma_{14}' & \Sigma_{24}' & \Sigma_{34}' & \Sigma_{44} \end{bmatrix},$$

$$R_1 = E\eta''(D_t, \Phi_t)\phi_t^2 \frac{\partial}{\partial\alpha}h(Z_t, \alpha_0)\frac{\partial}{\partial\alpha'}h(Z_t, \alpha_0) ,$$

$$R_3 = ED_t[X_t - \pi_{20}(\alpha_0, h(Z_t,\alpha_0))][\theta_0'\pi_{20}^{(2)}(\alpha_0, h(Z_t,\alpha_0)) - \pi_{10}^{(2)}(\alpha_0, h(Z_t,\alpha_0))] \times \frac{\partial}{\partial\alpha'}h(Z_t,\alpha_0)/\pi_3(\alpha_0, h(Z_t,\alpha_0)) ,$$

$$R_4 = -ED_t[X_t - \pi_{20}(\alpha_0,h(Z_t,\alpha_0))][X_t - \pi_{20}(\alpha_0,h(Z_t,\alpha_0))]'/\pi_{30}(\alpha_0,h(Z_t,\alpha_0)) = -\Sigma_{22} ,$$

$$\Sigma_{11} = E[\eta'(D_t, \Phi_t)\phi_t]^2 \frac{\partial}{\partial\alpha}h(Z_t, \alpha_0)\frac{\partial}{\partial\alpha'}h(Z_t, \alpha_0) , \qquad (4.4.7)$$

$$\Sigma_{13} = \left[E\eta'(D_t, \Phi_t)(D_t - \Phi_t)\phi_t^2 \frac{\partial}{\partial\alpha}h(Z_t, \alpha_0)\frac{\partial}{\partial\alpha'}h(Z_t, \alpha_0)/\Phi_t(1 - \Phi_t)\right]J^{-1} ,$$

$$\Sigma_{24} = \left[ED_t[X_t - \pi_{20}(\alpha_0,h(Z_t,\alpha_0))][X_t - \pi_{20}(\alpha_0,h(Z_t,\alpha_0))]'\right]A^{11} ,$$

$$\Sigma_{34} = J^{-1}B'A^{-1}[I_{q_2} \vdots 0]' , \ \Sigma_{14} = \Sigma_{13}J\Sigma_{34} , \ \Sigma_{33} = J^{-1} ,$$

$$\Sigma_{44} = [I_{q_2} \vdots 0]A^{-1}\left[BJ^{-1}B' + E\pi_{30}(\alpha_0,h(Z_t,\alpha_0))\begin{bmatrix} X_tX_t' & X_t\lambda_t \\ X_t'\lambda_t & D_t\lambda_t^2 \end{bmatrix}\right]A^{-1}\begin{bmatrix} I_{q_2} \\ 0 \end{bmatrix} ,$$

$$A^{11} = [I_{q_2} \vdots 0]A^{-1}[I_{q_2} \vdots 0]' ,$$

$$J = E\phi_t^2 \frac{\partial}{\partial\alpha}h(Z_t, \alpha_0)\frac{\partial}{\partial\alpha'}h(Z_t, \alpha_0)/\Phi_t(1 - \Phi_t) ,$$

$$A = \begin{bmatrix} EX_tX_t' & EX_t\lambda_t \\ EX_t'\lambda_t & ED_t\lambda_t^2 \end{bmatrix}, \text{ and}$$

$$B = \sigma_{12}\begin{bmatrix} ED_t\lambda_t(\lambda_t + h(Z_t, \alpha_0))X_t \frac{\partial}{\partial\alpha'}h(Z_t, \alpha_0) \\ ED_t\lambda_t^2(\lambda_t + h(Z_t, \alpha_0))\frac{\partial}{\partial\alpha'}h(Z_t, \alpha_0) \end{bmatrix} .$$

Here, $\pi_{j0}^{(2)}(\cdot,\cdot)$ $[\eta''(\cdot,\cdot)]$ denotes the derivative of $\pi_{j0}(\cdot,\cdot)$ $[\eta'(\cdot,\cdot)]$ with respect to its second argument for $j = 1, 2$. The above expression for Φ uses the fact that under the null hypothesis $H_0 : (U_t, \epsilon_t)$ is bivariate normal, $\varphi_t = \Phi_t$ and $\frac{\partial}{\partial\alpha}\pi_{40}(\alpha_0, h(Z_t, \alpha_0)) = \frac{\partial}{\partial\alpha}\Phi(-h(Z_t, \alpha_0)) = -\phi_t \frac{\partial}{\partial\alpha}h(Z_t, \alpha_0)$.

If the bivariate normality assumption of (U_t, ϵ_t) is violated and $\hat{\beta} = (\hat{\alpha}', \hat{\theta}')' \xrightarrow{p} \bar{\beta} = (\bar{\alpha}', \bar{\theta}')' \neq \beta_0$, then the test usually is consistent, because

$$r(\bar{\beta}, \pi_0) = Er_t(\bar{\beta}, \pi_0) = \begin{bmatrix} Er_{1t}(\bar{\alpha}, \pi_{40}) \\ Er_{2t}(\bar{\beta}, \pi_0) \end{bmatrix} \tag{4.4.8}$$

usually is not equal to zero, where

$$\begin{aligned} Er_{1t}(\bar{\alpha},\pi_{40}) &= E\eta'(\pi_{40}(\alpha_0, h(Z_t,\alpha_0)), \pi_{40}(\bar{\alpha}, h(Z_t,\bar{\alpha})))\frac{\partial}{\partial\alpha}\pi_{40}(\bar{\alpha},h(Z_t,\bar{\alpha})) \text{ and} \\ Er_{2t}(\bar{\beta},\pi_0) &= E[\pi_{10}(\alpha_0,h(Z_t,\alpha_0)) - \pi_{10}(\bar{\alpha},h(Z_t,\bar{\alpha})) + (X_t-\pi_{20}(\alpha_0,h(Z_t,\alpha_0)))' \\ &\quad \times (\theta_0 - \bar{\theta}) + (\pi_{20}(\bar{\alpha}, h(Z_t,\bar{\alpha})) - \pi_{20}(\alpha_0, h(Z_t,\alpha_0)))'\bar{\theta}] \\ &\quad \times [X_t - \pi_{20}(\bar{\alpha}, h(Z_t,\bar{\alpha}))]/\pi_{30}(\bar{\alpha}, h(Z_t,\bar{\alpha})) . \end{aligned} \tag{4.4.9}$$

5. A TEST OF SEMIPARAMETRIC VERSUS NONPARAMETRIC MODELS

In this section, we consider a specification test for the functional form of the regression function. The null model of interest is the partially linear regression (PLR) model and the alternative model is the nonparametric regression (NR) model. The PLR model and the NR model can be written as

$$Y_t = \pi_{10}(Z_t) + (X_t - \pi_{20}(Z_t))'\beta_0 + U_t \text{ and } E(U_t|X_t, Z_t) = 0 \text{ a.s. and} \tag{5.1}$$

$$Y_t = \pi_{30}(X_t, Z_t) + \epsilon_t \text{ and } E(\epsilon_t|X_t, Z_t) = 0 \text{ a.s.,} \tag{5.2}$$

respectively, for $t = 1, \ldots, 2T$, where $Y_t, U_t, \epsilon_t \in R$, $X_t, \beta_0 \in R^q$, $Z_t \in R^p$,

$$\pi_{10}(Z_t) = E(Y_t|Z_t), \ \pi_{20}(Z_t) = E(X_t|Z_t), \text{ and } \pi_{30}(X_t,Z_t) = E(Y_t|X_t,Z_t). \tag{5.3}$$

We note that Yatchew (1988) has recently proposed a specification test of a linear regression model against a nonparametric regression model. Consistency of Yatchew's (1988) test is proved by Wooldridge (1989). Below we extend the results of Yatchew (1988) and Wooldridge (1989) to the case where the null model is semiparametric rather than parametric. As in Yatchew (1988), our test statistic is based on the difference between the sums of squared residuals from the "restricted" and "unrestricted" models. Our testing procedure also requires a sample splitting (described below) to prevent degeneracy of the limit distribution of the test statistic.

Suppose we have an iid sample $\{(Y_t, X_t', Z_t')' : t = 1, \ldots, 2T\}$ and we split the sample into two independent sub–samples of size T. Let $\{(Y_t, X_t', Z_t')' : t = 1, \ldots, T\}$ and $\{(Y_t^*, X_t^{*\prime}, Z_t^{*\prime})' : t = 1, \ldots, T\}$ denote the first and the second sub–samples. We estimate the null model (5.1) using the first sub–sample and the alternative model (5.2) using the second sub–sample.

Let $W_t = (Y_t, X_t', Z_t', Y_t^*, X_t^{*\prime}, Z_t^{*\prime})'$ and $\pi = (\pi_1, \pi_2', \pi_3)'$. Define

$$\bar{r}_T(\hat{\beta},\hat{\pi}) = \tfrac{1}{T}\Sigma_1^T\Big[[Y_t - \hat{\pi}_1(Z_t) - (X_t - \hat{\pi}_2(Z_t))'\hat{\beta}]^2 - [Y_t^* - \hat{\pi}_3(X_t^*, Z_t^*)]^2\Big], \qquad (5.4)$$

where $\hat{\beta}$ is as defined in equation (3.1.3) and $\hat{\pi}_j$ denotes a nonparametric estimator of π_{j0} for $j = 1, 2, 3$. Note that the expression in (5.4) is the difference (scaled by $1/T$) between the sums of square residuals from the null and alternative models. Note also that if the PLR model is correctly specified, then $Er_t(\beta_0, \pi_0) = EU_t^2 - EU_t^{*2} = 0$, where $U_t^* = Y_t^* - \pi_{30}(X_t^*, Z_t^*)$.

Let $\hat{\eta}$ be an estimator of $\eta_0 = \mathrm{Var}\, U_t^2$. For example, $\hat{\eta}$ could be $\frac{1}{T}\Sigma_1^T\hat{U}_t^4 - \left[\frac{1}{T}\Sigma_1^T\hat{U}_t^2\right]^2$ or $\frac{1}{T}\Sigma_1^T\hat{U}_t^{*4} - \left[\frac{1}{T}\Sigma_1^T\hat{U}_t^{*2}\right]^2$, where $\hat{U}_t = Y_t - \hat{\pi}_1(Z_t) - (X_t - \hat{\pi}_2(Z_t))'\hat{\beta}$ and $\hat{U}_t^* = Y_t^* - \hat{\pi}_3(X_t^*, Z_t^*)$. Our test statistic is defined as follows:

$$PLN_T = \sqrt{T}\, \bar{r}_T(\hat{\beta},\hat{\pi})/(2\hat{\eta})^{1/2}. \qquad (5.5)$$

Under assumptions given below PLN_T is asymptotically standard normal.

ASSUMPTION PLN: (a) $\sqrt{T}(\hat{\beta}-\beta_0) = \left[E[X_t - \pi_{20}(Z_t)][X_t - \pi_{20}(Z_t)]'\right]^{-1}$
$\times \frac{1}{\sqrt{T}}\Sigma_1^T(X_t - \pi_{20}(Z_t))U_t + o_p(1)$.

(b) (i) $P(\hat{\pi} \in \Pi) \to 1$. (ii) $\int\|\hat{\pi}_j(z) - \pi_{j0}(z)\|^4 dP(z) \xrightarrow{p} 0$ for $j = 1, 2$ and $\int\|\hat{\pi}_3(x,z) - \pi_{30}(x,z)\|^4 dP(x,z) \xrightarrow{p} 0$, where $P(\cdot)$ and $P(\cdot,\cdot)$ denote the distributions of Z_t and (X_t, Z_t) respectively.

(c) $T^{1/4}\left[\int\|\hat{\pi}_j(z) - \pi_{j0}(z)\|^2 dP(z)\right]^{1/2} \xrightarrow{p} 0$ for $j = 1, 2$ and
$T^{1/4}\left[\int\|\hat{\pi}_3(x,z) - \pi_{30}(x,z)\|^2 dP(x,z)\right]^{1/2} \xrightarrow{p} 0$.

(d) $\{(U_t, X_t', Z_t', U_t^*, X_t^{*\prime}, Z_t^{*\prime})' : t \geq 1\}$ are iid. $(U_t, X_t', Z_t')'$ and $(U_t^*, X_t^{*\prime}, Z_t^{*\prime})'$ are independent and have the same distribution.

(e) $\left\{\frac{1}{\sqrt{T}}\Sigma_1^T\left[(U_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\beta_0)^2 - (U_t^* + \pi_{30}(X_t^*,Z_t^*) - \pi_3(X_t^*,Z_t^*))^2\right] : T \geq 1\right\}$ minus its mean is stochastically equicontinuous at $\pi = \pi_0$ with ρ_Π as in (2.2.6).

(f) $\{[U_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (X_t - \pi_{20}(Z_t))'\beta_0 - (X_t - \pi_2(Z_t))'\beta][X_t - \pi_2(Z_t)] : t \geq 1\}$ satisfies a uniform WLLN over $B_0 \times \Pi$.

(g) $EU_t^4 < \infty$ and $E\|X_t\|^2 < \infty$.

(h) $\hat{\eta} \xrightarrow{p} \eta_0$.

THEOREM PLN: *Suppose Assumption* PLN *holds. If the* PLR *model* (5.1) *is correctly specified, then* $PLN_T \xrightarrow{d} N(0,1)$.

COMMENT: If the null model is the linear regression model, then Yatchew's (1988) test fits into our general framework by taking $(X_t', Z_t') = Q_t'$, $(X_t^{*\prime}, Z_t^{*\prime}) = Q_t^{*\prime}$, $\pi_0(Q_t^*) = E(Y_t^*|Q_t^*)$, $\hat{\beta}$ equals to the OLS estimator of β_0 in the model $Y_t = Q_t'\beta_0 + U_t$ and $E(U_t|Q_t) = 0$ a.s., and $r_t(\beta,\pi) = (Y_t - Q_t'\beta)^2 - (Y_t^* - \pi(Q_t^*))^2$.

If the regression function of the PLR model is not correctly specified, the test PLN_T is consistent, since under suitable assumptions

$$\hat{\beta} \xrightarrow{p} \bar{\beta} = \underset{\beta \in B}{\operatorname{argmin}} \, E[\pi_{10}(Z_t) - (X_t - \pi_{20}(Z_t))'\beta - \pi_{30}(X_t, Z_t)]^2 \text{ and} \tag{5.6}$$

$$\bar{r}_T(\hat{\beta}, \hat{\pi}) \xrightarrow{p} Er_t(\bar{\beta}, \pi_0) = E[\pi_{10}(Z_t) - (X_t - \pi_{20}(Z_t))'\bar{\beta} - \pi_{30}(X_t, Z_t)]^2 > 0. \tag{5.7}$$

APPENDIX

Throughout the Appendix, we let A_j denote the j–th row of A for any matrix A .

PROOF OF LEMMA 1: The proof consists of three parts corresponding to Assumptions 1, 1*, and 1** respectively.

(1) Suppose Assumption 1 holds. Element by element mean value expansions of $\sqrt{T}\,\bar{r}_T(\hat{\beta},\hat{\pi})$ about β_0 give: $\forall j = 1, \ldots, m$,

$$\begin{aligned}\sqrt{T}\,\bar{r}_{Tj}(\hat{\beta},\hat{\pi}) &= \sqrt{T}[\bar{r}_{Tj}(\beta_0, \hat{\pi}) - \bar{r}^*_{Tj}(\beta_0, \hat{\pi})] + \sqrt{T}\,\bar{r}^*_{Tj}(\beta_0, \hat{\pi}) \\ &\quad + \left[\frac{1}{T}\Sigma_1^T \frac{\partial}{\partial\beta'} r_{tj}(\beta^*, \hat{\pi})\right]\sqrt{T}(\hat{\beta} - \beta_0) ,\end{aligned} \tag{A.1}$$

where β^* is a rv that depends on j and lies on the line segment joining $\hat{\beta}$ and β_0 , and hence, $\beta^* \xrightarrow{p} \beta_0$ by Assumptions 1(a) and (d). (See Jennrich (1969) Lemma 3 for the mean value theorem for random functions.)

Below we show that

$$\nu_{Tj}(\hat{\pi}) \equiv \sqrt{T}[\bar{r}_{Tj}(\beta_0, \hat{\pi}) - \bar{r}^*_{Tj}(\beta_0, \hat{\pi})] = \nu_{Tj}(\pi_0) + o_p(1) \tag{A.2}$$

and

$$\bar{R}_{Tj}(\beta^*, \hat{\pi}) \equiv \frac{1}{T}\Sigma_1^T \frac{\partial}{\partial\beta'} r_{tj}(\beta^*, \hat{\pi}) = R_j(\beta_0, \pi_0) + o_p(1) . \tag{A.3}$$

Since the second term on the right–hand side (rhs) of (A.1) is $o_p(1)$ by Assumption 1(c), equations (A.1)–(A.3) and Assumptions 1(a) and (d) give the following result by stacking terms for $j = 1, \ldots, m$:

$$\begin{aligned}\sqrt{T}\bar{r}_T(\hat{\beta},\hat{\pi}) &= \sqrt{T}[\bar{r}_T(\beta_0,\pi_0) - \bar{r}^*_T(\beta_0,\pi_0)] + R(\beta_0,\pi_0)\sqrt{T}[\bar{\psi}_T(\beta_0) - \bar{\psi}^*_T(\beta_0)] + o_p(1) \\ &= [I_m \vdots R(\beta_0, \pi_0)]\sqrt{T}\,\bar{g}_T(\beta_0, \pi_0) + o_p(1) ,\end{aligned} \tag{A.4}$$

where the first equality uses Assumptions 1(a) and (d) and $\bar{g}_T(\cdot,\cdot)$ is as defined in Assumption 1(d). Now from equation (A.4) and Assumption 1(d), the result of Lemma 1 follows directly.

Equation (A.2) follows from Assumptions 1(b) and (e) because: $\forall \epsilon > 0$ and $\eta > 0$ $\exists \delta > 0$ such that

$$\begin{aligned}
&\overline{\lim_{T\to\infty}} P(|\nu_{Tj}(\hat{\pi}) - \nu_{Tj}(\pi_0)| > \eta) \\
&\leq \overline{\lim_{T\to\infty}} P(|\nu_{Tj}(\hat{\pi}) - \nu_{Tj}(\pi_0)| > \eta,\ \hat{\pi} \in \Pi,\ \rho_{\Pi}(\hat{\pi}, \pi_0) \leq \delta) \\
&\quad + \overline{\lim_{T\to\infty}} P(\hat{\pi} \notin \Pi \text{ or } \rho_{\Pi}(\hat{\pi}, \pi_0) > \delta) \qquad (A.5) \\
&\leq \overline{\lim_{T\to\infty}} P^*\Big(\sup_{\pi\in\Pi:\rho_{\Pi}(\pi,\pi_0)\leq\delta} |\nu_{Tj}(\pi) - \nu_{Tj}(\pi_0)| > \eta\Big) \\
&< \epsilon .
\end{aligned}$$

Equation (A.3) follows because

$$\begin{aligned}
&\|\bar{R}_{Tj}(\beta^*, \hat{\pi}) - R_j(\beta_0, \pi_0)\| \\
&\leq \|\bar{R}_{Tj}(\beta^*, \hat{\pi}) - \bar{R}^*_{Tj}(\beta_0, \hat{\pi})\| + \|\bar{R}^*_{Tj}(\beta^*, \hat{\pi}) - R_j(\beta^*, \hat{\pi})\| \qquad (A.6) \\
&\quad + \|R_j(\beta^*, \hat{\pi}) - R_j(\beta_0, \pi_0)\| \xrightarrow{p} 0 ,
\end{aligned}$$

where $\bar{R}^*_T(\beta,\pi) = \frac{1}{T}\Sigma_1^T E\frac{\partial}{\partial\beta'}r_t(\beta,\pi)$ and the convergence to zero uses Assumptions 1(a), (b), (d), and (f).

(2) Suppose Assumption 1* holds. By adding and subtracting terms, we have

$$\sqrt{T}\,\bar{r}_T(\hat{\beta},\hat{\pi}) = \nu_T(\hat{\beta},\hat{\pi}) + \sqrt{T}\,\bar{r}^*_T(\hat{\beta},\hat{\pi}) . \qquad (A.7)$$

Below we show that

$$\nu_T(\hat{\beta},\hat{\pi}) = \nu_T(\beta_0, \pi_0) + o_p(1) \qquad (A.8)$$

and

$$\sqrt{T}\,\bar{r}^*_T(\hat{\beta},\hat{\pi}) = R(\beta_0, \pi_0)\sqrt{T}(\hat{\beta} - \beta_0) + o_p(1) . \qquad (A.9)$$

Now equations (A.7)–(A.9) and Assumptions 1*(a) and (d) yield the following result by a similar argument to that of (A.4):

$$\sqrt{T}\,\bar{r}_T(\hat{\beta},\hat{\pi}) = [I_m \vdots R(\beta_0, \pi_0)]\sqrt{T}\,\bar{g}_T(\beta_0, \pi_0) + o_p(1)\,. \tag{A.10}$$

The result of Lemma 1 follows directly from equation (A.10) and Assumption 1*(d).

Note that (A.8) follows from Assumptions 1*(b) and (e) by a similar argument to that of (A.5). To show (A.9), consider element by element mean value expansions of $\sqrt{T}\,\bar{r}_T^*(\hat{\beta},\hat{\pi})$ about β_0 : $\forall j = 1, \ldots, m$,

$$\sqrt{T}\,\bar{r}_{Tj}^*(\hat{\beta},\hat{\pi}) = \sqrt{T}\,\bar{r}_{Tj}^*(\beta_0, \hat{\pi}) + \left[\frac{1}{T}\Sigma_1^T \frac{\partial}{\partial\beta'} Er_{tj}(\beta,\hat{\pi})\right]_{\beta=\beta^*} \sqrt{T}(\hat{\beta} - \beta_0)\,, \tag{A.11}$$

where β^* is a rv that depends on j and lies on the line segment joining $\hat{\beta}$ and β_0, and hence, $\beta^* \xrightarrow{p} \beta_0$ by Assumptions 1*(a) and (d). Note that the first term on the RHS of (A.11) is $o_p(1)$ by Assumption 1*(c). Define $\bar{R}_T^*(\beta,\pi) = \frac{1}{T}\Sigma_1^T \frac{\partial}{\partial\beta'} Er_t(\beta,\pi)$. Then, by Assumptions 1*(a), (b), (d), and (f),

$$\begin{aligned} &\|\bar{R}_{Tj}^*(\beta^*,\hat{\pi}) - R_j(\beta_0,\pi_0)\| \\ &\leq \|\bar{R}_{Tj}^*(\beta^*, \hat{\pi}) - R_j(\beta^*,\hat{\pi})\| + \|R_j(\beta^*,\hat{\pi}) - R_j(\beta_0,\pi_0)\| \xrightarrow{p} 0\,. \end{aligned} \tag{A.12}$$

This establishes (A.9) since $\sqrt{T}(\hat{\beta} - \beta_0) = O_p(1)$ by Assumptions 1*(a) and (d).

(3) Suppose Assumption 1** holds. The difference between Assumptions 1 and 1** is that the latter specifies Π to be finite dimensional and it replaces Assumption 1(c) by Assumption 1**(c) and stochastic equicontinuity of $\{\nu_T(\cdot)\}$ at π_0 (Assumption 1(e)) by Assumption 1**(e). Stochastic equicontinuity of $\{\nu_T(\cdot)\}$ is used in the proof only to show that

$$\nu_T(\hat{\pi}) - \nu_T(\pi_0) \xrightarrow{p} \underset{\sim}{0}\,. \tag{A.13}$$

Thus, it suffices to show that Assumptions 1**(c) and (e) imply Assumption 1(c) and (A.13).

To show that Assumption 1(c) holds, consider element by element mean value expansions of $\sqrt{T}\,\bar{r}_T^*(\beta_0, \hat{\pi})$ about π_0 : $\forall j = 1, \ldots, m$,

$$\sqrt{T}\,\bar{r}_{Tj}^*(\beta_0, \hat{\pi}) = \sqrt{T}\,\bar{r}_{Tj}^*(\beta_0, \pi_0) + \left[\frac{1}{T}\Sigma_1^T \frac{\partial}{\partial\pi'} Er_{tj}(\beta_0, \pi^*)\right]\sqrt{T}(\hat{\pi} - \pi_0)\,, \tag{A.14}$$

where π^* lies on the line segment joining $\hat{\pi}$ and π_0, and hence, $\pi^* \xrightarrow{p} \pi_0$ by Assumption 1**(b). Define

$$Q_T(\beta,\pi) = \frac{1}{T}\Sigma_1^T E\frac{\partial}{\partial\pi'}r_t(\beta,\pi) . \tag{A.15}$$

Note that by the moment condition in Assumption 1**(e),

$$\frac{\partial}{\partial\pi'}Er_{tj}(\beta_0, \pi)\Big|_{\pi=\pi^*} = E\frac{\partial}{\partial\pi'}r_{tj}(\beta_0, \pi)\Big|_{\pi=\pi^*} \quad \forall t \geq 1 . \tag{A.16}$$

Thus, the term in square brackets in (A.14) is equal to $Q_{Tj}(\beta_0, \pi^*)$. Since

$$\|Q_{Tj}(\beta_0, \pi^*) - Q_j(\beta_0, \pi_0)\| \xrightarrow{p} 0 \tag{A.17}$$

by a similar argument to that of (A.12) under Assumptions 1**(b) and (e) and $Q_j(\beta_0, \pi_0) = \underset{\sim}{0}$, we have $Q_{Tj}(\beta_0, \pi^*) \xrightarrow{p} \underset{\sim}{0}$. This result, $\sqrt{T}(\hat{\pi} - \pi_0) = O_p(1)$ (Assumption 1**(e)), Assumption 1**(c), and equation (A.14) establish Assumption 1(c).

To show (A.13), consider element by element mean value expansions of $\nu_T(\hat{\pi})$ about π_0 : $\forall j = 1, \ldots, m$,

$$\nu_{Tj}(\hat{\pi}) - \nu_{Tj}(\pi_0) = \left[\frac{1}{\sqrt{T}}\frac{\partial}{\partial\pi'}\nu_{Tj}(\pi^*)\right]\sqrt{T}(\hat{\pi} - \pi_0) , \tag{A.18}$$

where π^* lies on the line segment joining $\hat{\pi}$ and π_0. The rhs of (A.18) is $o_p(1)$ under Assumption 1**(e), since $\sqrt{T}(\hat{\pi} - \pi_0) = O_p(1)$ and

$$\begin{aligned}\left\|\frac{1}{\sqrt{T}}\frac{\partial}{\partial\pi}\nu_{Tj}(\pi^*)\right\| &= \left\|\frac{1}{T}\Sigma_1^T\left[\frac{\partial}{\partial\pi}r_{tj}(\beta_0, \pi^*) - E\frac{\partial}{\partial\pi}r_{tj}(\beta_0, \pi)\Big|_{\pi=\pi^*}\right]\right\| \\ &\leq \sup_{\pi\in\Pi}\left\|\frac{1}{T}\Sigma_1^T\left[\frac{\partial}{\partial\pi}r_{tj}(\beta_0, \pi) - E\frac{\partial}{\partial\pi}r_{tj}(\beta_0, \pi)\right]\right\| \xrightarrow{p} 0 ,\end{aligned} \tag{A.19}$$

where the first equality in (A.19) holds by (A.16). □

PROOF OF THEOREM 1: The proof of Theorem 1 follows by the results of Lemma 1, Assumption 3, and the continuous mapping theorem if we show that $\hat{\Phi} \xrightarrow{p} \Phi$. Since $\hat{\Sigma} \xrightarrow{p} \Sigma$ under Assumption 2, it suffices to prove that $\hat{R} \xrightarrow{p} R(\beta_0, \pi_0)$.

First, suppose either Assumption 1 or 1** holds. Then $\hat{R}$ (defined in (2.2.12)) is consistent for $R(\beta_0, \pi_0)$ (defined in part (f) of Assumptions 1 and 1**) by a similar argument to that of (A.6) under either Assumptions 1(a), (b), (d), and (f) or Assumptions 1**(a), (b), (d), and (f).

Next, suppose Assumptions 1* and 2* hold. Below we show that $\hat{R} \xrightarrow{p} R(\beta_0, \pi_0) = R$, where $\hat{R}$ and $R(\beta_0, \pi_0)$ are defined in (2.2.13) and Assumption 1*(f) respectively. Let $Er_t(\hat{\beta},\hat{\pi})$ denote $Er_t(\beta,\pi)$ evaluated at $(\beta,\pi) = (\hat{\beta},\hat{\pi})$. We have

$$\begin{aligned}\|\hat{R}_j - R_j\| \leq & \left\|\hat{R}_j - \tfrac{1}{T}\Sigma_1^T(Er_t(\hat{\beta} + \epsilon_T e_j, \hat{\pi}) - Er_t(\hat{\beta} - \epsilon_T e_j, \hat{\pi}))/(2\epsilon_T)\right\| \\ & + \left\|\tfrac{1}{T}\Sigma_1^T(Er_t(\hat{\beta} + \epsilon_T e_j, \hat{\pi}) - Er_t(\hat{\beta} - \epsilon_T e_j, \hat{\pi}))/(2\epsilon_T) - \tfrac{1}{T}\Sigma_1^T \tfrac{\partial}{\partial\beta_j}Er_t(\hat{\beta},\hat{\pi})\right\| \\ & + \left\|\tfrac{1}{T}\Sigma_1^T \tfrac{\partial}{\partial\beta_j}Er_t(\hat{\beta},\hat{\pi}) - R_j(\hat{\beta},\hat{\pi})\right\| + \|R_j(\hat{\beta},\hat{\pi}) - R_j\| \\ = & \, A_{1T} + A_{2T} + A_{3T} + A_{4T} \text{ (say)}.\end{aligned} \tag{A.20}$$

Since

$$A_{1T} = \frac{1}{2\epsilon_T\sqrt{T}}(\nu_T(\hat{\beta}+\epsilon_T e_j,\hat{\pi}) - \nu_T(\beta_0, \pi_0)) - \frac{1}{2\epsilon_T\sqrt{T}}(\nu_T(\hat{\beta}-\epsilon_T e_j,\hat{\pi}) - \nu_T(\beta_0,\pi_0)), \tag{A.21}$$

Assumptions 1*(e), 2*(a), and 2*(c) combine to yield $A_{1T} \xrightarrow{p} 0$. Assumptions 1*(a), 1*(b), 1*(d), 2*(a), and 2*(b) imply $A_{2T} \xrightarrow{p} 0$. Assumption 1*(f) implies $A_{3T} \xrightarrow{p} 0$ and $A_{4T} \xrightarrow{p} 0$. Hence, $\hat{R} \xrightarrow{p} R$ as desired. □

PROOF OF THEOREM 2: First, we prove that under Assumption 1–ℓp, 1*–ℓp, or 1**–ℓp the following result holds:

$$\sqrt{T}\, \bar{r}_T(\hat{\beta},\hat{\pi}) \xrightarrow{d} N(\xi + R(\beta_0, \pi_0)\eta, \Phi). \tag{A.22}$$

Then, Theorem 2 holds by a similar argument to that of the proof of Theorem 1.

To show (A.22), suppose Assumption 1–ℓp holds. Then

$$\sqrt{T}\, \bar{r}_T(\hat{\beta},\hat{\pi}) = [I_m \,\vdots\, R(\beta_0, \pi_0)]\sqrt{T}\, \bar{g}_T(\beta_0, \pi_0) + \xi + R(\beta_0, \pi_0)\eta + o_p(1) \tag{A.23}$$

by a similar argument to that of the proof of Lemma 1, using the fact that $\hat{\beta} \xrightarrow{p} \beta_0$, $\sqrt{T}(\hat{\beta} - \beta_0) = \sqrt{T}(\hat{\beta} - \beta_T) + \eta$, and $\sqrt{T}\, \bar{r}^*_{Tj}(\beta_0, \hat{\pi}) \xrightarrow{p} \xi$. Equation (A.22) follows from Assumption 1–ℓp(d) and (A.23). Equation (A.22) holds under Assumption 1*–ℓp or 1**–ℓp by a similar argument to that given above. □

PROOF OF THEOREM 3: We show that

$$\|\bar{r}_T(\hat{\beta},\hat{\pi}) - r(\bar{\beta},\bar{\pi})\| \xrightarrow{p} 0 . \tag{A.24}$$

Then, Theorem 3 follows since

$$|\,\bar{r}_T(\hat{\beta},\hat{\pi})'\hat{\Phi}^{-1}\bar{r}_T(\hat{\beta},\hat{\pi}) - r(\bar{\beta},\bar{\pi})'\bar{\Phi}^{-1}r(\bar{\beta},\bar{\pi})\,| \xrightarrow{p} 0 , \tag{A.25}$$

$\|r(\bar{\beta},\bar{\pi})\| > 0$, and $\bar{\Phi}^{-1}$ is nonsingular (Assumption 4(d)).

Equation (A.24) holds because

$$\begin{aligned} &\|\bar{r}_T(\hat{\beta},\hat{\pi}) - r(\bar{\beta},\bar{\pi})\| \\ &\le \|\bar{r}_T(\hat{\beta},\hat{\pi}) - \bar{r}^*_T(\hat{\beta},\hat{\pi})\| + \|\bar{r}^*_T(\hat{\beta},\hat{\pi}) - r(\hat{\beta},\hat{\pi})\| + \|r(\hat{\beta},\hat{\pi}) - r(\bar{\beta},\bar{\pi})\| \xrightarrow{p} 0 , \end{aligned} \tag{A.26}$$

where the convergence to zero uses Assumptions 4(a), (b), and (c). □

PROOF OF THEOREM PLO: Below we show that Assumption PLO (hereafter PLO) implies Assumption 1 (hereafter As. 1). Since PLO(g) and (h) imply Assumptions 2 and 3 respectively, the result of Theorem PLO follows from Theorem 1. We note that As. 1(a) is implied by PLO(a). The first part of As. 1(b) and As. 1(e) are implied by PLO (b)–(i) and (e) respectively. The second part of As. 1(b) holds by PLO (b) and (d) using the pseudo–metric ρ_Π defined in (2.2.6) because

$$\rho_{\Pi}(\hat{\pi}, \pi_0) = \Big[E\|\epsilon_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\beta_0][Q_t - \pi_3(Z_t)]$$
$$- \epsilon_t[Q_t - \pi_{30}(Z_t)]\|^2\Big|_{\pi=\hat{\pi}}\Big]^{1/2}$$
$$\leq \Big[\Big[\int|\hat{\pi}_1(z) - \pi_{10}(z)|^4 dP(z)\Big]^{1/4} + \Big[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^4 dP(z)\Big]^{1/4}\|\beta_0\|\Big] \tag{B.1}$$
$$\times \Big[\Big[\int\|\hat{\pi}_3(z) - \pi_{30}(z)\|^4 dP(z)\Big]^{1/4} + 2\Big[E\|Q_t\|^4\Big]^{1/4}\Big]$$
$$+ \Big[E\epsilon_t^4\Big]^{1/4}\Big[\int\|\hat{\pi}_3(z) - \pi_{30}(z)\|^4 dP(z)\Big]^{1/4} \xrightarrow{p} 0 .$$

As. 1(c) follows from PLO(c) and (d), because

$$\|\sqrt{T}\,\bar{r}_T^*(\beta_0, \hat{\pi})\|$$
$$= \|\sqrt{T}E[\epsilon_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\beta_0][Q_t - \pi_3(Z_t)]\|\Big|_{\pi=\hat{\pi}} \tag{B.2}$$
$$\leq \Big\{T^{1/4}\Big[\int|\hat{\pi}_1(z)-\pi_{10}(z)|^2 dP(z)\Big]^{1/2} + T^{1/4}\Big[\int\|\hat{\pi}_2(z)-\pi_{20}(z)\|^2 dP(z)\Big]^{1/2}\|\beta_0\|\Big\}$$
$$\times\, T^{1/4}\Big[\int\|\hat{\pi}_3(z) - \pi_{30}(z)\|^2 dP(z)\Big]^{1/2} \xrightarrow{p} 0 .$$

As. 1(d) is satisfied by Corollary 1 of Herrndorf (1984) and PLO(d), because

$$\sqrt{T}\,\bar{g}_T(\beta_0, \pi_0) = \begin{bmatrix} \frac{1}{\sqrt{T}}\Sigma_1^T \quad (Q_t - \pi_{30}(Z_t))\epsilon_t \\ \frac{1}{\sqrt{T}}\Sigma_1^T J^{-1}(X_t - \pi_{20}(Z_t))\epsilon_t \end{bmatrix} . \tag{B.3}$$

The uniform WLLN condition of As. 1(f) follows from PLO(f). Continuity of $R(\beta,\pi)$ at (β_0, π_0), which is required by As. 1(f), holds using the pseudo–metric ρ^* of (2.2.5) by PLO (b) and (d):

$$\rho^*((\hat{\beta},\hat{\pi}), (\beta_0, \pi_0))$$
$$= E\|(Q_t - \pi_3(Z_t))(X_t - \pi_2(Z_t))' - (Q_t - \pi_{30}(Z_t))(X_t - \pi_{20}(Z_t))'\|\Big|_{\pi=\hat{\pi}} \tag{B.4}$$
$$\leq \Big[\int\|\hat{\pi}_3(z) - \pi_{30}(z)\|^2 dP(z)\Big]^{1/2}\Big[\Big[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^2 dP(z)\Big]^{1/2} + 2\Big[E\|X_t\|^2\Big]^{1/2}\Big]$$
$$+ 2\Big[E\|Q_t\|^2\Big]^{1/2}\Big[\int\|\hat{\pi}_3(z) - \pi_{30}(z)\|^2 dP(z)\Big]^{1/2} \xrightarrow{p} 0 . \quad \square$$

PROOF OF THEOREM PLH: First we verify As.1. As. 1(a) follows from (Assumption) PLH (a). The first part of As. 1(b), As. 1(e), and the uniform WLLN assumption of As. 1(f) follow directly from PLO (b)–(i), (e), and (f) respectively. The second part of As. 1(b) holds by PLO (b) and (d) using the pseudo–metric ρ_Π defined in (2.2.6), because

$$\begin{aligned}\rho_\Pi(\hat{\pi}, \pi_0) &= \Big[E\|Q_t[\pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\theta_0] \\ &\quad \times [\pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\theta_0 + 2\epsilon_t]\|^2\Big|_{\pi=\hat{\pi}}\Big]^{1/2} \\ &\leq \Big[E\|Q_t\|^4\Big]^{1/4}\Big[\Big[\int|\hat{\pi}_1(z) - \pi_{10}(z)|^8 dP(z)\Big]^{1/8} \\ &\quad + \Big[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^8 dP(z)\Big]^{1/8}\|\theta_0\|\Big] \\ &\quad \times \Big[2\Big[E\epsilon_t^8\Big]^{1/8} + \Big[\int|\hat{\pi}_1(z) - \pi_{10}(z)|^8 dP(z)\Big]^{1/8} \\ &\quad + \Big[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^8 dP(z)\Big]^{1/8}\|\theta_0\|\Big] \xrightarrow{p} 0\,. \end{aligned} \tag{C.1}$$

As. 1(c) is implied by PLH (b), (c),and (d), since

$$\begin{aligned}&\|\sqrt{T}\,\bar{r}_T^*(\beta_0, \hat{\pi})\| \\ &= \Big\|\sqrt{T}EQ_t\Big[(\epsilon_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\theta_0)^2 - \sigma_0^2\Big]\Big\|\Big|_{\pi=\hat{\pi}} \\ &= \Big\|\sqrt{T}EQ_t[\pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\theta_0]^2\Big\|\Big|_{\pi=\hat{\pi}} \\ &\leq \Big[E\|Q_t\|^2\Big]^{1/2}\Big[T^{1/4}\Big[\int|\hat{\pi}_1(z) - \pi_{10}(z)|^4 dP(z)\Big]^{1/4} \\ &\quad + T^{1/4}\Big[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^4 dP(z)\Big]^{1/4}\|\theta_0\|\Big]^2 \xrightarrow{p} 0\,. \end{aligned} \tag{C.2}$$

As. 1(d) is verified by Corollary 1 of Herrndorf (1984) and PLH (b) and (d), since

$$\sqrt{T}\,\bar{g}_T(\beta_0, \pi_0) = \begin{bmatrix} \frac{1}{\sqrt{T}}\Sigma_1^T[\epsilon_t^2 - \sigma_0^2]Q_t \\ \frac{1}{\sqrt{T}}\Sigma_1^T J^{-1}(X_t - \pi_{20}(Z_t))\epsilon_t \\ \frac{1}{\sqrt{T}}\Sigma_1^T(\epsilon_t^2 - \sigma_0^2) \end{bmatrix}. \tag{C.3}$$

To verify continuity of $R(\beta,\pi)$ at (β_0, π_0) of As. 1(f), note that

$$R(\beta,\pi) = -\Big[2EQ_t[Y_t - \pi_1(Z_t) - (X_t - \pi_2(Z_t))'\theta][X_t - \pi_2(Z_t)]' \;\vdots\; EQ_t\Big] . \tag{C.4}$$

With the pseudo–metric ρ^* defined in (2.2.5), it suffices to show that

$$\begin{aligned} &\rho^*((\hat\beta,\hat\pi), (\beta_0, \pi_0)) \\ &= E\Big\| Q_t[Y_t - \pi_1(Z_t) - (X_t - \pi_2(Z_t))'\theta][X_t - \pi_2(Z_t)]' - \epsilon_t Q_t[X_t - \pi_{20}(Z_t)]' \Big\| \Big|_{\substack{\theta=\hat\theta \\ \pi=\hat\pi}} \\ &\xrightarrow{p} 0 . \end{aligned} \tag{C.5}$$

Note that (C.5) holds by PLH (a), (b), and (d), since the left–hand side (lhs) of (C.5) is less than or equal to

$$\begin{aligned} &\Big[E\|Q_t\|^2\Big]^{1/2}\Bigg[\Big[E\|\pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'(2\theta_0 - \theta) \\ &\quad + (X_t - \pi_{20}(Z_t))'(\theta_0 - \theta)\|^4\Big]^{1/4}\Big\{\Big[E\|\pi_2(Z_t) - \pi_{20}(Z_t)\|^4\Big]^{1/4} \\ &\quad + \Big[E\|X_t - \pi_{20}(Z_t)\|^4\Big]^{1/4}\Big\} + \Big[E\epsilon_t^4\Big]^{1/4}\Big[E\|\pi_2(Z_t) - \pi_{20}(Z_t)\|^4\Big]^{1/4}\Bigg]\Bigg|_{\substack{\theta=\hat\theta \\ \pi=\hat\pi}} \end{aligned}$$

$$\begin{aligned} &\le \Big[E\|Q_t\|^2\Big]^{1/2}\Bigg[\Bigg\{\Big[\int|\hat\pi_1(z) - \pi_{10}(z)|^4 dP(z)\Big]^{1/4} + (\|\hat\theta - \theta_0\| + \|\theta_0\|) \\ &\quad \times \Big[\int\|\hat\pi_2(z) - \pi_{20}(z)\|^4 dP(z)\Big]^{1/4} + \|\hat\theta - \theta_0\|\Big[E\|X_t\|^4\Big]^{1/4}\Bigg\} \\ &\quad \times \Big\{\Big[\int\|\hat\pi_2(z) - \pi_{20}(z)\|^4 dP(z)\Big]^{1/4} + \Big[E\|X_t\|^4\Big]^{1/4}\Big\} + \Big[E\epsilon_t^4\Big]^{1/4} \\ &\quad \times \Big[\int\|\hat\pi_2(z) - \pi_{20}(z)\|^4 dP(z)\Big]^{1/4}\Bigg] \xrightarrow{p} 0 . \end{aligned} \tag{C.6}$$

Now the result of Theorem follows by PLH (g) and (h) and Theorem 1 by noting that $R(\beta_0, \pi_0) = [0 \;\vdots\; -EQ_t]$. □

PROOF OF THEOREM PLA: First we verify As. 1. As. 1(a), the first part of As. 1(b), As. 1(e), the uniform WLLN of As. 1(f) are verified using similar arguments to those given in the proof of Theorem PLO. The second part of As. 1(b) follows from (Assumption) PLA (b) and (d) using the pseudo–metric ρ_Π defined in (2.2.6), because

$$\begin{aligned}
&\rho_\Pi(\hat{\pi}, \pi_0) \\
&= \Big[E\|[\epsilon_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\beta_0][\epsilon_{t-1} + \pi_{10}(Z_{t-1}) \\
&\quad - \pi_1(Z_{t-1}) + (\pi_2(Z_{t-1}) - \pi_{20}(Z_{t-1}))'\beta_0] - \epsilon_t\epsilon_{t-1}\|^2\Big|_{\pi=\hat{\pi}}\Big]^{1/2} \qquad (D.1)\\
&\le \left[\left[\int|\hat{\pi}_1(z) - \pi_{10}(z)|^4 dP(z)\right]^{1/4} + \left[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^4 dP(z)\right]^{1/4}\|\beta_0\|\right]^2 \\
&\quad + 2\left[E\epsilon_t^4\right]^{1/4}\left[\int|\hat{\pi}_1(z) - \pi_{10}(z)|^4 dP(z)\right]^{1/4}\left[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^4 dP(z)\right]^{1/4}\|\beta_0\| \\
&\xrightarrow{p} 0\,.
\end{aligned}$$

As. 1(c) follows from PLA (c) and (d), since

$$\begin{aligned}
&\|\sqrt{T}\,\bar{r}_T^*(\beta_0, \hat{\pi})\| \\
&= \|\sqrt{T}\,E[\epsilon_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\beta_0][\epsilon_{t-1} + \pi_{10}(Z_{t-1}) \\
&\quad - \pi_1(Z_{t-1}) + (\pi_2(Z_{t-1}) - \pi_{20}(Z_{t-1}))'\beta_0]\|\Big|_{\pi=\hat{\pi}} \qquad (D.2)\\
&\le T^{1/2}\int|\hat{\pi}_1(z) - \pi_{10}(z)|^2 dP(z) + 2\|\beta_0\|T^{1/4}\left[\int|\hat{\pi}_1(z) - \pi_{10}(z)|^2 dP(z)\right]^{1/2} \\
&\quad \times T^{1/4}\left[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^2 dP(z)\right]^{1/2} + \|\beta_0\|^2 T^{1/2}\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^2 dP(z) \\
&\xrightarrow{p} 0\,.
\end{aligned}$$

As. 1(d) follows from PLA (b) and (d) and Corollary 1 of Herrndorf (1984), because

$$\sqrt{T}\,\bar{g}_T(\beta_0, \pi_0) = \begin{bmatrix} \frac{1}{\sqrt{T}}\Sigma_1^T \epsilon_t \epsilon_{t-1} \\ \frac{1}{\sqrt{T}}\Sigma_1^T J^{-1}(X_t - \pi_{20}(Z_t))\epsilon_t \end{bmatrix} \text{ and }$$

$$\Sigma = \begin{bmatrix} \sigma_0^4 & 0 \\ 0 & \sigma_0^2 E(X_t - \pi_{20}(Z_t))(X_t - \pi_{20}(Z_t))' \end{bmatrix}. \tag{D.3}$$

To verify continuity of $R(\beta,\pi)$ at (β_0, π_0) of As. 1(f), note that

$$\begin{aligned} \frac{\partial}{\partial\beta'} r_t(\beta,\pi) = &-[\epsilon_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (X_t - \pi_{20}(Z_t))'\beta_0 - (X_t - \pi_2(Z_t))'\beta] \\ &\times [X_{t-1} - \pi_2(Z_{t-1})]' - [\epsilon_{t-1} + \pi_{10}(Z_{t-1}) - \pi_1(Z_{t-1}) \\ &+ (X_{t-1} - \pi_{20}(Z_{t-1}))'\beta_0 - (X_{t-1} - \pi_2(Z_{t-1}))'\beta][X_t - \pi_2(Z_t)]' . \end{aligned} \tag{D.4}$$

It suffices to show that $\rho^*((\hat\beta,\hat\pi), (\beta_0, \pi_0)) \xrightarrow{p} 0$, where $\rho^*(\cdot,\cdot)$ is as defined in (2.2.5). The latter result holds since under PLA (a), (b), and (d),

$$\begin{aligned} E\Big\|[\epsilon_t + \pi_{10}(Z_t) - \pi_1(Z_t) + X_t'(\beta_0 - \beta) + \pi_2(Z_t)'\beta - \pi_{20}(Z_t)'\beta_0][X_{t-1} - \pi_2(Z_{t-1})] \\ - \epsilon_t[X_{t-1} - \pi_{20}(Z_{t-1})]\Big\|\Big|_{\substack{\beta=\hat\beta \\ \pi=\hat\pi}} \xrightarrow{p} 0 \end{aligned} \tag{D.5}$$

and likewise with the roles of (ϵ_t, X_t, Z_t) and $(\epsilon_{t-1}, X_{t-1}, Z_{t-1})$ reversed.

Note that $R = R(\beta_0, \pi_0) = \underset{\sim}{0}$. Therefore the result of Theorem PLA follows by Theorem 1 and PLA (g), since $\Phi = [1 \;\; 0]\Sigma[1 \;\; 0]' = \sigma_0^4$. □

PROOF OF THEOREM SSO: First we show that Assumption SSO (hereafter SSO) implies As. 1. Below we verify only As. 1(b)–(d) and 1(f). The remaining arguments are similar to those given above (see, for example, the proof of Theorem PLO).

With ρ_Π defined by (2.2.6), the second part of As. 1(b) holds by SSO (b) and (d), since

$$
\begin{aligned}
&\rho_{\Pi}^{2}(\hat{\pi}, \pi_0) \\
&= \int 1(D=1)\Big\| [\mu + \pi_{10}(v) - \hat{\pi}_1(v) + (\hat{\pi}_2(v) - \pi_{20}(v))'\theta_0][q - \hat{\pi}_4(v)]/\hat{\pi}_3(v) \\
&\quad - \mu[q - \pi_{40}(v)]/\pi_{30}(v)\Big\|^2 dP(\mu,D,q,v) \\
&= \int 1(D=1)\Big\| [\pi_{10}(v) - \hat{\pi}_1(v) + (\hat{\pi}_2(v) - \pi_{20}(v))'\theta_0][q - \hat{\pi}_4(v)]/\hat{\pi}_3(v) \\
&\quad + \mu[\pi_{40}(v) - \hat{\pi}_4(v)]/\hat{\pi}_3(v) + \mu[q - \pi_{40}(v)][\pi_{30}(v) - \hat{\pi}_3(v)] \\
&\quad /(\hat{\pi}_3(v)\pi_{30}(v))\Big\|^2 dP(\mu,D,q,v) \xrightarrow{p} 0 ,
\end{aligned}
\tag{E.1}
$$

where $\hat{\pi}_j(v)$ and $\pi_{j0}(v)$ abbreviate $\hat{\pi}_j(\alpha_0, v)$ and $\pi_{j0}(\alpha_0, v)$ respectively for $j = 1, 2, 3, 4$ and $P(\cdot,\cdot,\cdot,\cdot)$ denotes the distribution of $(\mu_t, D_t, Q_t, h(Z_t, \alpha_0))$.

As. 1(c) holds by SSO (b), (c), and (d), since

$$
\begin{aligned}
&\|\sqrt{T}\, \bar{r}_T^*(\beta_0, \hat{\pi})\| \\
&= \|\sqrt{T} E D_t[\mu_t + \pi_{10}(\alpha_0, h(Z_t, \alpha_0)) - \pi_1(\alpha_0, h(Z_t, \alpha_0)) + (\pi_2(\alpha_0, h(Z_t, \alpha_0)) \\
&\quad - \pi_{20}(\alpha_0, h(Z_t, \alpha_0)))'\theta_0][Q_t - \pi_4(\alpha_0, h(Z_t, \alpha_0))]/\pi_3(\alpha_0, h(Z_t, \alpha_0))\| \Big|_{\pi=\hat{\pi}} \\
&\le \Bigg[T^{1/4}\Big[\int 1(D=1)|\hat{\pi}_1(v) - \pi_{10}(v)|^2 dP(D,v)\Big]^{1/2} \\
&\quad + T^{1/4}\Big[\int 1(D=1)\|\hat{\pi}_2(v) - \pi_{20}(v)\|^2 dP(D,v)\Big]^{1/2} \|\theta_0\| \Bigg] \\
&\quad \times T^{1/4}\Big[\int 1(D=1)\|\hat{\pi}_4(v) - \pi_{40}(v)\|^2 dP(D,v)\Big]^{1/2} /\epsilon + o_p(1) \xrightarrow{p} 0 .
\end{aligned}
\tag{E.2}
$$

Let

$$
\begin{aligned}
&\sqrt{T}\, \bar{g}_T(\beta_0, \pi_0) = \tfrac{1}{\sqrt{T}}\Sigma_1^T [r_t(\beta_0, \pi_0)' \quad \psi_{1Tt}(\alpha_0)' \quad \psi_{2Tt}(\beta_0)']' \\
&= \tfrac{1}{\sqrt{T}}\Sigma_1^T \begin{bmatrix} \mu_t[Q_t - \pi_{40}(\alpha_0, h(Z_t, \alpha_0))]/\pi_{30}(\alpha_0, h(Z_t, \alpha_0)) \\ -M_1^{-1}\eta'(D_t, \varphi_t)H_t \\ M_4^{-1}M_3M_1^{-1}\eta'(D_t, \varphi_t)H_t - M_4^{-1}\mu_t[X_t - \pi_{20}(\alpha_0, h(Z_t, \alpha_0))]/\pi_{30}(\alpha_0, h(Z_t, \alpha_0)) \end{bmatrix} .
\end{aligned}
\tag{E.3}
$$

The summands in (E.3) are square integrable by the following arguments. First,

$E\|r_t(\beta_0, \pi_0)\|^2 < \infty$ by SSO (b)–(iii), the Cauchy–Schwarz inequality, and Jensen's inequality, since

$$
\begin{aligned}
E\mu_t^4 &= ED_t[U_t - E(U_t | \epsilon_t > -h(Z_t, \alpha_0))]^4 \\
&\leq \left[\left[ED_tU_t^4\right]^{1/4} + \left[E[E(D_tU_t | \epsilon_t > -h(Z_t, \alpha_0))]^4\right]^{1/4}\right]^4 \qquad \text{(E.4)} \\
&\leq 2^4 ED_tU_t^4 < \infty
\end{aligned}
$$

and, similarly, $E\|Q_t - \pi_{40}(\alpha_0, h(Z_t, \alpha_0))\|^4 \leq 2^4 E\|Q_t\|^4 < \infty$. Second, $E\|\psi_{1Tt}(\alpha_0)\|^2 < \infty$ by SSO (b)–(iii) and (d)–(ii), since $|\eta'(D_t, \varphi_t)|$ is bounded and $E\|H_t\|^2 \leq C\cdot E\left\|\frac{\partial}{\partial\alpha}h(Z_t, \alpha_0)\right\|^2 < \infty$ for some $C < \infty$ by a similar argument to that of (E.4). Finally, note that $E\|\psi_{2Tt}(\beta_0)\|^2 < \infty$ by the same arguments as those used to show $E\|r_t(\beta_0, \pi_0)\|^2 < \infty$ and $E\|\psi_{1Tt}(\beta_0)\|^2 < \infty$. Therefore, As. 1(d) holds by a multivariate CLT for iid rv's and hence $\sqrt{T}\,\bar{g}_T(\beta_0, \pi_0) \xrightarrow{d} N(\underset{\sim}{0},\Sigma)$, where Σ is as defined in (3.2.16). Note that the expression given for Σ uses the fact that $E(\mu_t^2 | h(Z_t, \alpha_0), X_t, Q_t, D_t = 1) = E(\mu_t^2 | h(Z_t, \alpha_0), D_t = 1) = \pi_{30}(\alpha_0, h(Z_t, \alpha_0))$.

With $\rho^*(\cdot,\cdot)$ defined by (2.2.5), the last part of As. 1(f) holds immediately from the following results:

$$
E\|R_{jt}(\beta,\pi) - R_{jt}(\beta_0, \pi_0)\|\Big|_{\substack{\beta=\hat{\beta} \\ \pi=\hat{\pi}}} \xrightarrow{p} 0 \text{ for } j = 1, 2, 3, 4 . \qquad \text{(E.5)}
$$

To show (E.5) holds when $j = 1$, let $\tilde{\pi}_j^{(i)}(z)$, $\hat{\pi}_j^{(i)}(z)$, and $\pi_{j0}^{(i)}(z)$ abbreviate $\hat{\pi}_j^{(i)}(\hat{\alpha}, h(z,\hat{\alpha}))$, $\hat{\pi}_j^{(i)}(\alpha_0, h(z,\alpha_0))$, and $\pi_{j0}^{(i)}(\alpha_0, h(z,\alpha_0))$ for $i = 0, 1, 2$ and $j = 1, 2, 3, 4$. Then,

$$\begin{aligned}
&E\|R_{1t}(\beta,\pi) - R_{1t}(\beta_0, \pi_0)\|\Big|_{\substack{\beta=\hat\beta \\ \pi=\hat\pi}} \\
&= \int 1(D=1)\Big\|[\pi_{10}(z) - \tilde\pi_1(z) + (\tilde\pi_2(z) - \pi_{20}(z))'\theta_0 + (\tilde\pi_2(z) - x)'(\hat\theta - \theta_0)] \\
&\quad \times \Big[\tilde\pi_4^{(1)}(z) + \tilde\pi_4^{(2)}(z)\frac{\partial}{\partial\alpha'}h(z,\hat\alpha)\Big]/\tilde\pi_3(z) + \mu\Big[\tilde\pi_4^{(1)}(z) - \pi_{40}^{(1)}(z) \\
&\quad + \tilde\pi_4^{(2)}(z)\frac{\partial}{\partial\alpha'}h(z,\hat\alpha) - \pi_{40}^{(2)}(z)\frac{\partial}{\partial\alpha'}h(z,\alpha_0)\Big]/\tilde\pi_3(z) + \mu\Big[\pi_{40}^{(1)}(z) \\
&\quad + \pi_{40}^{(2)}(z)\frac{\partial}{\partial\alpha'}h(z,\alpha_0)\Big]\ [\pi_{30}(z) - \tilde\pi_3(z)]/(\tilde\pi_3(z)\pi_{30}(z))\Big\|dP(\mu,D,x,z) \xrightarrow{p} 0\ ,
\end{aligned} \tag{E.6}$$

where $P(\cdot,\cdot,\cdot,\cdot)$ denotes the distribution of (μ_t, D_t, X_t, Z_t). The convergence to zero in (E.6) uses the fact that

$$\sup_{z\in\mathcal{Z}}\Big\|\frac{\partial}{\partial\alpha}h(z,\hat\alpha) - \frac{\partial}{\partial\alpha}h(z, \alpha_0)\Big\| \xrightarrow{p} 0 \text{ and} \tag{E.7}$$

$$\begin{aligned}
&\Big[\int 1(D=1)\|\tilde\pi_j^{(i)}(z) - \pi_{j0}(z)\|^k dP(D,z)\Big]^{1/k} \\
&\le \Big[\int 1(D=1)\|\tilde\pi_j^{(i)}(z) - \hat\pi_j^{(i)}(z)\|^k dP(D,z)\Big]^{1/k} \\
&\quad + \Big[\int 1(D=1)\|\hat\pi_j^{(i)}(v) - \pi_{j0}^{(i)}(v)\|^k dP(D,v)\Big]^{1/k} \\
&\le \sup_{\pi\in\Pi}\sup_{z\in\mathcal{Z}}\|\pi_j^{(i)}(\hat\alpha, h(z,\hat\alpha)) - \pi_j^{(i)}(\alpha_0, h(z, \alpha_0))\| \\
&\quad + \Big[\int 1(D=1)\|\hat\pi_j^{(i)}(v) - \pi_{j0}^{(i)}(v)\|^k dP(D,v)\Big]^{1/k} + o_p(1) \xrightarrow{p} 0\ ,
\end{aligned} \tag{E.8}$$

for $i = 0, 1, 2$ and $j = 1, 2, 3, 4$ and for some $k > 0$ (as assumed in SSO (b)), where $P(D,z)$ and $P(D,v)$ denote the distributions of (D_t, Z_t) and $(D_t, h(Z_t, \alpha_0))$ respectively and $\hat\pi_j^{(i)}(v)$ and $\pi_{j0}^{(i)}(v)$ abbreviate $\hat\pi_j^{(i)}(\alpha_0, v)$ and $\pi_{j0}^{(i)}(\alpha_0, v)$ respectively.

The proof of (E.5) for $j = 2, 3, 4$ is similar. □

PROOF OF THEOREM SSH: The proof of Theorem SSH is similar to that of Theorem SSO. Here we also verify As. 1(b)–(d) and 1(f). The notation used below is same as in the proof of Theorem SSO. The second part of As. 1(b) holds by SSH (b) and (d) with ρ_Π defined by (2.2.6), since

$$
\begin{aligned}
&\rho_{\Pi}^2(\hat{\pi}, \pi_0) \\
&= \int 1(D=1)\Big\| [\mu + \pi_{10}(v) - \hat{\pi}_1(v) + (\hat{\pi}_2(v) - \pi_{20}(v))'\theta_0]^2 - \mu^2 \\
&\quad + \pi_{30}(v) - \hat{\pi}_3(v)][q - \hat{\pi}_4(v)]/\hat{\pi}_3(v) + [\mu^2 - \pi_{30}(v)][\pi_{40}(v) - \hat{\pi}_4(v)]/\hat{\pi}_3(v) \\
&\quad + [\mu^2 - \pi_{30}(v)][q - \pi_{40}(v)][\pi_{30}(v) - \hat{\pi}_3(v)]/(\hat{\pi}_3(v)\pi_{30}(v))\Big\|^2 dP(\mu,D,q,v) \\
&\xrightarrow{p} 0 .
\end{aligned} \tag{F.1}
$$

As. 1(c) holds by SSH (c) and (d), since

$$
\begin{aligned}
&\|\sqrt{T}\,\bar{r}_T^*(\beta_0, \hat{\pi})\| \\
&\leq T^{1/4}\left[\int 1(D=1)\|\hat{\pi}_4(v) - \pi_{40}(v)\|^2 dP(D,v)\right]^{1/2} \\
&\quad \times \Bigg[T^{1/4}\left[\int 1(D=1)|\hat{\pi}_3(v) - \pi_{30}(v)|^2 dP(D,v)\right]^{1/2} \\
&\quad + T^{1/4}\left[\int 1(D=1)|\hat{\pi}_1(v) - \pi_{10}(v)|^4 dP(D,v)\right]^{1/2} \\
&\quad + T^{1/4}\left[\int 1(D=1)\|\hat{\pi}_2(v) - \pi_{20}(v)\|^4 dP(D,v)\right]^{1/2}\|\theta_0\|^2 \\
&\quad + 2T^{1/8}\left[\int 1(D=1)|\hat{\pi}_1(v) - \pi_{10}(v)|^4 dP(D,v)\right]^{1/4} \\
&\quad \times T^{1/8}\left[\int 1(D=1)\|\hat{\pi}_2(v) - \pi_{20}(v)\|^4 dP(D,v)\right]^{1/4}\|\theta_0\|\Bigg]/\epsilon + o_p(1) \\
&\xrightarrow{p} 0 .
\end{aligned} \tag{F.2}
$$

As. 1(d) is verified as in the proof of Theorem SSO except that $r_t(\beta_0, \pi_0)$ is now defined by

$$
r_t(\beta_0,\pi_0) = D_t[\mu_t^2 - \pi_{30}(\alpha_0,h(Z_t,\alpha_0))][Q_t - \pi_{40}(\alpha_0,h(Z_t,\alpha_0))]/\pi_{30}(\alpha_0,h(Z_t,\alpha_0)) . \tag{F.3}
$$

Therefore, $\sqrt{T}\,\bar{g}_T(\beta_0, \pi_0) \xrightarrow{d} N(\underset{\sim}{0},\Sigma)$, where

$$\Sigma = \begin{bmatrix} \Sigma_{11} & 0 & \Sigma_{13} \\ 0 & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{13}' & \Sigma_{23}' & \Sigma_{33} \end{bmatrix}, \ \Sigma_{22} = M_1^{-1} S_1 M_1^{-1}, \ \Sigma_{23} = -M_1^{-1} S_1 M_1^{-1} M_3' M_4^{-1},$$

$$\Sigma_{13} = -\Big[E D_t \mu_t^3 [Q_t - \pi_{40}(\alpha_0, h(Z_t, \alpha_0))][X_t - \pi_{20}(\alpha_0, h(Z_t, \alpha_0))]' \\ / \pi_{30}(\alpha_0, h(Z_t, \alpha_0)) \Big] M_4^{-1},$$

$$\Sigma_{33} = M_4^{-1}(S_2 + M_3 M_1^{-1} S_1 M_1^{-1} M_3') M_4^{-1}, \text{ and}$$

M_1, M_3, M_4, S_1, and S_2 are as defined in (3.2.16). Note that the above expression for Σ uses the fact that $E(\mu_t^2 | D_t = 1, Z_t, Q_t) = E(\mu_t^2 | D_t = 1, h(Z_t, \alpha_0))$.

If As. 1(f) holds, then we have $R = E\frac{\partial}{\partial \beta'} r_t(\beta_0, \pi_0) = [R_1 \ \ 0]$. Therefore, we have $\Phi = [I_m \vdots R] \Sigma [I_m \vdots R]' = \Sigma_{11} + R_1 M_1^{-1} S_1 M_1^{-1} R_1'$ as required.

Now the last part of As. 1(f) can be verified using similar arguments to those given in (E.6)–(E.8). □

PROOF OF THEOREM CLAD: First we show that Assumption CLAD (hereafter CLAD) implies Assumption 1* (hereafter As. 1*). As. 1*(a) holds by CLAD (a) and (b). Note that $E\psi_{1Tt}(\beta_0) = \underset{\sim}{0}$ and $E\psi_{2Tt}(\beta_0) = 0$ because $E(D_t | X_t) = F_{0t}$ a.s., $E(D_t U_t | X_t) = \sigma_0^2 f_{0t}$ a.s., and $E(D_t U_t^2 | X_t) = \sigma_0^2 (F_{0t} - X_t' \beta_0 f_{0t})$ a.s.

To verify As. 1*(b), consider the pseudo–metric of (2.2.9) with π_1 and π_2 eliminated. Using CLAD (a) and (c), we have

$$\begin{aligned}
\rho_B^2(\hat\beta, \beta_0) &= \sup_{N\geq 1} \tfrac{1}{N}\Sigma_1^N E\|r_t(\beta) - r_t(\beta_0)\|^2 \Big|_{\beta=\hat\beta} \\
&= \sup_{N\geq 1} \tfrac{1}{N}\Sigma_1^N E\Big\|\Big[\tfrac{1}{2} - 1(U_t < X_t'(\beta-\beta_0))\Big]1(X_t'\beta > 0)X_t - \Big[\tfrac{1}{2} - 1(U_t < 0)\Big] \\
&\quad \times 1(X_t'\beta_0 > 0)X_t\Big\|^2 \Big|_{\beta=\hat\beta} \\
&= \sup_{N\geq 1} \tfrac{1}{N}\Sigma_1^N E\Big\|[1(U_t < 0) - 1(U_t < X_t'(\beta-\beta_0))]1(X_t'\beta > 0)X_t \\
&\quad + \Big[\tfrac{1}{2} - 1(U_t < 0)\Big][1(X_t'\beta > 0) - 1(X_t'\beta_0 > 0)]X_t\Big\|^2 \Big|_{\beta=\hat\beta} \qquad \text{(G.1)} \\
&\leq \sup_{N\geq 1} \tfrac{2}{N}\Sigma_1^N E\|[1(U_t < 0) - 1(U_t < X_t'(\beta-\beta_0))]X_t\|^2 \Big|_{\beta=\hat\beta} \\
&\quad + \sup_{N\geq 1} \tfrac{1}{2N}\Sigma_1^N E\|[1(X_t'\beta > 0) - 1(X_t'\beta_0 > 0)]X_t\|^2 \Big|_{\beta=\hat\beta} \\
&\leq \sup_{N\geq 1} \tfrac{2}{N}\Sigma_1^N E\Big[1(|U_t| \leq \|X_t\|\cdot\|\beta - \beta_0\|)\|X_t\|^2\Big] \Big|_{\beta=\hat\beta} \\
&\quad + \sup_{N\geq 1} \tfrac{1}{2N}\Sigma_1^N E\Big[1(|X_t'\beta_0| \leq \|X_t\|\cdot\|\beta - \beta_0\|)\|X_t\|^2\Big] \Big|_{\beta=\hat\beta} .
\end{aligned}$$

The first term on the rhs of the last inequality in (G.1) is $o_p(1)$ because it equals

$$\begin{aligned}
&\sup_{N\geq 1} \tfrac{1}{N}\Sigma_1^N E\Big[[2F(\|X_t\|\cdot\|\beta - \beta_0\|, \sigma_0^2) - 1]\|X_t\|^2\Big] \Big|_{\beta=\hat\beta} \\
&= \sup_{N\geq 1} \tfrac{1}{N}\Sigma_1^N E\Big[f(X_t^*, \sigma_0^2)\|X_t\|^3\Big] \|\hat\beta - \beta_0\| \qquad \text{(G.2)} \\
&\leq C_1\|\hat\beta - \beta_0\| \xrightarrow{p} 0
\end{aligned}$$

by CLAD (a) and the facts that $\hat\beta \xrightarrow{p} \beta_0$ and $f(X_t^*, \sigma_0^2)$ is bounded a.s. uniformly in t, where X_t^* is a rv lying between 0 and $\|X_t\|\cdot\|\hat\beta - \beta_0\|$ and C_1 is some finite constant.

The second term in (G.1) is $o_p(1)$ by CLAD (a) and (c) and the fact that $\hat\beta \xrightarrow{p} \beta_0$:

$$\sup_{N\geq 1} \frac{1}{N}\Sigma_1^N E\left[1(|X_t'\beta_0| \leq \|X_t\|\cdot\|\beta-\beta_0\|)\|X_t\|^2\right]\Big|_{\beta=\hat\beta}$$

$$\leq \sup_{N\geq 1}\frac{1}{N}\Sigma_1^N E\left[1(|X_t'\beta_0| \leq \|X_t\|\cdot z)\|X_t\|^2\right] + \left[\sup_{N\geq 1}\frac{1}{N}\Sigma_1^N E\|X_t\|^2\right]1(\|\hat\beta-\beta_0\| \geq z) \qquad \text{(G.3)}$$

$$\leq K_1 z + K_2 1(\|\hat\beta-\beta_0\| \geq z)$$

for some finite constants K_1 and K_2. The rhs of (G.3) can be made arbitrarily small by choosing z sufficiently small and then T sufficiently large.

As. 1*(c) holds because

$$\begin{aligned}\sqrt{T}\,\bar r_T^*(\beta_0) &= \frac{1}{\sqrt{T}}\Sigma_1^T E\left[\frac{1}{2} - 1(Y_t - X_t'\beta_0 < 0)\right]1(X_t'\beta_0 > 0)X_t \\ &= \frac{1}{\sqrt{T}}\Sigma_1^T E\left[\frac{1}{2} - 1(\max\{-X_t'\beta_0, U_t\} < 0)\right]1(X_t'\beta_0 > 0)X_t \qquad \text{(G.4)} \\ &= \frac{1}{\sqrt{T}}\Sigma_1^T E\left[\frac{1}{2} - 1(U_t < 0)\right]1(X_t'\beta_0 > 0)X_t = \underset{\sim}{0}\,.\end{aligned}$$

Let

$$\sqrt{T}\,\bar g_T(\beta_0) = \frac{1}{\sqrt{T}}\Sigma_1^T \begin{bmatrix} \left[\frac{1}{2} - 1(Y_t - X_t'\beta_0 < 0)\right]1(X_t'\beta_0 > 0)X_t \\ [I_q \ \vdots\ 0]\,J^{-1}\begin{bmatrix}\psi_{1Tt}(\beta_0) \\ \psi_{2Tt}(\beta_0)\end{bmatrix}\end{bmatrix}. \qquad \text{(G.5)}$$

As. 1*(d) holds by CLAD (a) because $\sqrt{T}\,\bar g_T(\beta_0) \xrightarrow{d} N(\underset{\sim}{0},\Sigma)$ by Liapounov's CLT, where $\Sigma = \begin{bmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22}\end{bmatrix}$, $\Sigma_{11} = M/4$, $\Sigma_{22} = [I_q \vdots 0]J^{-1}[I_q \vdots 0]'$, and $\Sigma_{12} = f_0 M[I_q \vdots 0]J^{-1}[I_q \vdots 0]'$.

The stochastic equicontinuity of

$$\left\{\frac{1}{\sqrt{T}}\Sigma_1^T\left[\frac{1}{2} - 1(Y_t - X_t'\beta < 0)\right]1(X_t'\beta > 0)X_t : T \geq 1\right\} \qquad \text{(G.6)}$$

minus its mean at β_0 (As. 1*(e)) using the pseudo–metric defined in equation (G.1) holds by an application of Theorems II.1–II.3 of Andrews (1990b). This is because $\left\{\frac{1}{2} - 1(y - x'\beta < 0) : \beta \in R^q\right\}$, $\{1(x'\beta > 0) : \beta \in R^q\}$, and $\{g : g(x) = x\}$ are type I classes and $\{X_t : t \geq 1\}$ satisfies the envelope condition by CLAD (a).

To verify As. 1*(f), note that

$$\mathrm{Er}_t(\beta) = \mathrm{E}\left[\frac{1}{2} - \mathrm{F}(\mathrm{X}_t'(\beta - \beta_0),\ \sigma_0^2)\right] 1(\mathrm{X}_t'\beta > 0)\mathrm{X}_t\,. \tag{G.7}$$

$\mathrm{Er}_t(\beta)$ is differentiable in $\beta \in \mathrm{B}_0$ $\forall t$ under CLAD (c) with

$$\frac{\partial}{\partial \beta'}\mathrm{Er}_t(\beta) = -\mathrm{Ef}(\mathrm{X}_t'(\beta - \beta_0),\ \sigma_0^2)1(\mathrm{X}_t'\beta > 0)\mathrm{X}_t\mathrm{X}_t'\,. \tag{G.8}$$

Equation (G.8) holds since: $\forall j = 1, \ldots, q$, $\forall t$, $\forall \beta \in \mathrm{B}_0$,

$$\begin{aligned}
&\lim_{z\to 0}\left\|\frac{\mathrm{Er}_t(\beta + ze_j) - \mathrm{Er}_t(\beta)}{z} - \frac{\partial}{\partial \beta_j}\mathrm{Er}_t(\beta)\right\| \\
&\le \lim_{z\to 0}\left\|\frac{\mathrm{E}[\mathrm{F}(\mathrm{X}_t'(\beta-\beta_0),\sigma_0^2) - \mathrm{F}(\mathrm{X}_t'(\beta-\beta_0+ze_j),\sigma_0^2)]1(\mathrm{X}_t'(\beta+ze_j)>0)\mathrm{X}_t}{z} - \frac{\partial}{\partial \beta_j}\mathrm{Er}_t(\beta)\right\| \\
&\quad + \lim_{z\to 0}\left\|\frac{\mathrm{E}[1/2 - \mathrm{F}(\mathrm{X}_t'(\beta-\beta_0),\sigma_0^2)][1(\mathrm{X}_t'(\beta+ze_j)>0) - 1(\mathrm{X}_t'\beta>0)]\mathrm{X}_t}{z}\right\| \qquad \text{(G.9)} \\
&\le \lim_{z\to 0}\frac{\mathrm{E}|1(\mathrm{X}_t'(\beta+ze_j)>0) - 1(\mathrm{X}_t'\beta>0)|\,\|\mathrm{X}_t\|}{|z|} \\
&\le \lim_{z\to 0}\frac{\mathrm{E}1(|\mathrm{X}_t'\beta| \le \|\mathrm{X}_t\|\,|z|)\|\mathrm{X}_t\|}{|z|} = 0\,.
\end{aligned}$$

The second inequality in (G.9) holds by a mean value expansion and the dominated convergence theorem.

By CLAD (d), $\mathrm{R}(\beta) = -\lim_{T\to\infty}\frac{1}{T}\Sigma_1^T\mathrm{Ef}(\mathrm{X}_t'(\beta - \beta_0),\ \sigma_0^2)1(\mathrm{X}_t'\beta > 0)\mathrm{X}_t\mathrm{X}_t'$ exists uniformly over B_0. To verify continuity of $\mathrm{R}(\beta)$ at $\beta = \beta_0$, consider the pseudo-metric in (2.2.11) with π_1 and π_2 eliminated. It suffices to show that $\rho^*(\hat\beta, \beta_0) \xrightarrow{p} 0$. The latter holds by CLAD (a) and (c) using the following arguments. Let $\mathrm{f}(\cdot)$ abbreviate $\mathrm{f}(\cdot,\ \sigma_0^2)$.

$$
\begin{aligned}
\rho^*(\hat\beta, \beta_0) &= \overline{\lim_{N\geq 1}} \tfrac{1}{N}\Sigma_1^N \left\| \tfrac{\partial}{\partial\beta'}\mathrm{Er}_t(\beta) - \tfrac{\partial}{\partial\beta'}\mathrm{Er}_t(\beta_0) \right\| \Big|_{\beta=\hat\beta} \\
&= \overline{\lim_{N\to\infty}} \tfrac{1}{N}\Sigma_1^N \| \mathrm{E} f(X_t'(\beta-\beta_0)) 1(X_t'\beta > 0) X_t X_t' - \mathrm{E} f(0) 1(X_t'\beta_0 > 0) X_t X_t' \| \Big|_{\beta=\hat\beta} \\
&\leq \overline{\lim_{N\to\infty}} \tfrac{1}{N}\Sigma_1^N \mathrm{E}\Big[| f(X_t'(\beta-\beta_0)) - f(0) | \cdot \|X_t\|^2 \Big] \Big|_{\beta=\hat\beta} \qquad \text{(G.10)} \\
&\quad + f(0)\, \overline{\lim_{N\to\infty}} \tfrac{1}{N}\Sigma_1^N \mathrm{E}\Big[| 1(X_t'\beta > 0) - 1(X_t'\beta_0 > 0) | \cdot \|X_t\|^2 \Big] \Big|_{\beta=\hat\beta} .
\end{aligned}
$$

The second term on the rhs of (G.10) is $o_p(1)$ by the same argument as in (G.3). The first term on the rhs of (G.10) is also $o_p(1)$ since it is bounded by

$$
C_2 \Big[\overline{\lim_{N\to\infty}} \tfrac{1}{N}\Sigma_1^N \mathrm{E}\|X_t\|^3 \Big] \|\hat\beta - \beta_0\| = o_p(1) , \qquad \text{(G.11)}
$$

using a mean value expansion, where C_2 is some finite constant.

Finally, note that

$$
\begin{aligned}
\|\hat M - M\| &\leq \left\| \tfrac{1}{T}\Sigma_1^T 1(X_t'\hat\beta > 0) X_t X_t' - \tfrac{1}{T}\Sigma_1^T 1(X_t'\beta_0 > 0) X_t X_t' \right\| \qquad \text{(G.12)} \\
&\quad + \left\| \tfrac{1}{T}\Sigma_1^T 1(X_t'\beta_0 > 0) X_t X_t' - \lim_{T\to\infty} \tfrac{1}{T}\Sigma_1^T \mathrm{E} 1(X_t'\beta_0 > 0) X_t X_t' \right\| .
\end{aligned}
$$

The second term above is $o_p(1)$ by the WLLN. The first term is $o_p(1)$, since

$$
\begin{aligned}
&\left\| \tfrac{1}{T}\Sigma_1^T [1(X_t'\hat\beta > 0) - 1(X_t'\beta_0 > 0)] X_t X_t' \right\| \qquad \text{(G.13)} \\
&\leq \tfrac{1}{T}\Sigma_1^T 1(|X_t'\beta_0| \leq \|X_t\| \cdot \|\hat\beta - \beta_0\|) \cdot \|X_t\|^2 \xrightarrow{p} 0
\end{aligned}
$$

using a similar argument to that used in (G.3). The result of Theorem CLAD now follows from Theorem 1. □

PROOF OF THEOREM LPL: Assumption LPL (hereafter LPL) (g) implies Assumptions 2 and 3. LPL (b)–(i), (e), and (f) imply the first part of As. 1(b), As. 1(e), and the uniform WLLN of As. 1(f) respectively. If we verify the remaining parts of As. 1, then the result of Theorem LPL holds by Theorem 1.

As. 1(a) holds with $\psi_{Tt}(W_{Tt}, \beta_0) = [I_q \vdots 0]J^{-1}Q_tU_t$ because LPL (a) and (f) imply that $\hat{J} = (\Sigma_1^T Q_t Q_t')/T \xrightarrow{p} J$, J is nonsingular, $\frac{1}{\sqrt{T}}\Sigma_1^T Q_t U_t = O_p(1)$, and $E\psi_{Tt}(W_{Tt}, \beta_0) = \underset{\sim}{0}$, where the consistency of $\hat{J}$ for J under LPL (a) holds by Theorem 2 of Andrews (1988).

The second part of As. 1(b) holds by LPL (a) and (b) using the pseudo–metric ρ_Π defined in (2.2.6), because

$$\begin{aligned}
\rho_\Pi(\hat{\pi}, \pi_0) &= \Big[E\|[U_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\beta_0] \\
&\quad \times [X_t - \pi_2(Z_t)]/\pi_3(Z_t) - U_t[X_t - \pi_{20}(Z_t)]/\pi_{30}(Z_t)\|^2\Big|_{\pi=\hat{\pi}}\Big]^{1/2} \\
&\le \left[\left[\int|\hat{\pi}_1(z) - \pi_{10}(z)|^4 dP(z)\right]^{1/4} + \left[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^4 dP(z)\right]^{1/4}\|\beta_0\|\right] \\
&\quad \times \left[2\left[E\|X_t\|^4\right]^{1/4} + \left[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^4 dP(z)\right]^{1/4}\right]/\epsilon + \left[EU_t^4\right]^{1/4} \\
&\quad \times \left[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^4 dP(z)\right]^{1/4}/\epsilon + \left[EU_t^4\|X_t - \pi_{20}(Z_t)\|^4\right]^{1/4} \\
&\quad \times \left[\int|\hat{\pi}_3(z) - \pi_{30}(z)|^4 dP(z)\right]^{1/4}/\epsilon^2 + o_p(1) \xrightarrow{p} 0 .
\end{aligned} \tag{H.1}$$

As. 1(c) holds by LPL (a), (b)–(i), and (c), because

$$\begin{aligned}
&\|\sqrt{T}\,\bar{r}_T^*(\beta_0, \hat{\pi})\| \\
&\le T^{1/4}\left[\int|\hat{\pi}_1(z) - \pi_{10}(z)|^2 dP(z)\right]^{1/2} T^{1/4}\left[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^2 dP(z)\right]^{1/2}/\epsilon \\
&\quad + \|\beta_0\|T^{1/2}\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^2 dP(z)/\epsilon + o_p(1) \xrightarrow{p} 0 .
\end{aligned} \tag{H.2}$$

As. 1(d) holds by LPL (a) and (d) and Corollary 1 of Herrndorf (1984), since

$$\sqrt{T}\,\bar{g}_T(\beta_0, \pi_0) = \begin{bmatrix} \frac{1}{\sqrt{T}}\Sigma_1^T U_t[X_t - \pi_{20}(Z_t)]/\pi_{30}(Z_t) \\ \frac{1}{\sqrt{T}}\Sigma_1^T[I_q \ \vdots\ 0]J^{-1}Q_tU_t \end{bmatrix} . \tag{H.3}$$

The last part of As. 1(f) holds by LPL (a) and (b) using the pseudo–metric ρ^* defined in (2.2.5), because

$$\rho^*((\hat\beta,\hat\pi),(\beta_0,\pi_0)) = E\|[X_t-\pi_2(Z_t)][X_t-\pi_2(Z_t)]'/\pi_3(Z_t)$$
$$-[X_t-\pi_{20}(Z_t)][X_t-\pi_{20}(Z_t)]'/\pi_{30}(Z_t)\|\Big|_{\pi=\hat\pi}$$
$$\le \left[\int\|\hat\pi_2(z)-\pi_{20}(z)\|^2 dP(z)\right]/\epsilon + 4\left[E\|X_t\|^2\right]^{1/2} \tag{H.4}$$
$$\times\left[\int\|\hat\pi_2(z)-\pi_{20}(z)\|^2 dP(z)\right]^{1/2}/\epsilon$$
$$+4\left[E\|X_t\|^4\right]^{1/2}\left[\int|\hat\pi_3(z)-\pi_{30}(z)|^2 dP(z)\right]^{1/2}/\epsilon^2 + o_p(1) \xrightarrow{p} 0\,. \square$$

PROOF OF THEOREM PLN: First we show that Assumption PLN (hereafter PLN) implies As. 1. Below we verify only the second part of As. 1(b), As. 1(c), As. 1(d), and the last part of As. 1(f). The remaining parts of As. 1 are implied directly by PLN.

The second part of As. 1(b) holds with ρ_Π defined by (2.2.6) by PLN (b)–(ii), (d), and (g), since

$$\rho_\Pi(\hat\pi,\pi_0) = \Bigg[E\Big\|[U_t+\pi_{10}(Z_t)-\pi_1(Z_t)+(\pi_2(Z_t)-\pi_{20}(Z_t))'\beta_0]^2$$
$$-[U_t^*+\pi_{30}(X_t^*,Z_t^*)-\pi_3(X_t^*,Z_t^*)]^2-(U_t^2-U_t^{*2})\Big\|^2\Big|_{\pi=\hat\pi}\Bigg]^{1/2}$$
$$\le\left[\left[\int|\hat\pi_1(z)-\pi_{10}(z)|^4 dP(z)\right]^{1/4}+\|\beta_0\|\left[\int\|\hat\pi_2(z)-\pi_{20}(z)\|^4 dP(z)\right]^{1/4}\right]^2$$
$$+2\left[EU_t^4\right]^{1/4}\left[\left[\int|\hat\pi_1(z)-\pi_{10}(z)|^4 dP(z)\right]^{1/4} \tag{I.1}$$
$$+\|\beta_0\|\left[\int\|\hat\pi_2(z)-\pi_{20}(z)\|^4 dP(z)\right]^{1/4}\right]$$
$$+\left[\int|\hat\pi_3(x,z)-\pi_{30}(x,z)|^4 dP(x,z)\right]^{1/2}$$
$$+2\left[EU_t^4\right]^{1/4}\left[\int|\hat\pi_3(x,z)-\pi_{30}(x,z)|^4 dP(x,z)\right]^{1/4} \xrightarrow{p} 0\,.$$

As. 1(c) holds by PLN (c) and (d), because

$$
\begin{aligned}
|\sqrt{T}\,\bar{r}_T^*(\beta_0, \hat{\pi})| &= \Big|\sqrt{T}\,E\Big[[U_t + \pi_{10}(Z_t) - \pi_1(Z_t) + (\pi_2(Z_t) - \pi_{20}(Z_t))'\beta_0]^2 \\
&\qquad - [U_t^* + \pi_{30}(X_t^*, Z_t^*) - \pi_3(X_t^*, Z_t^*)]^2\Big]\Big|_{\pi=\hat{\pi}}\Big| \\
&\le T^{1/2}\int|\hat{\pi}_1(z) - \pi_{10}(z)|^2 dP(z) + \|\beta_0\|^2 T^{1/2}\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^2 dP(z) \\
&\quad + 2\|\beta_0\|T^{1/4}\Big[\int|\hat{\pi}_1(z) - \pi_{10}(z)|^2 dP(z)\Big]^{1/2} \qquad (I.2) \\
&\quad \times T^{1/4}\Big[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^2 dP(z)\Big]^{1/2} + T^{1/2}\int|\hat{\pi}_3(x,z) - \pi_{30}(x,z)|^2 dP(x,z) \\
&\xrightarrow{p} 0\,.
\end{aligned}
$$

As. 1(d) holds by a multivariate CLT for iid rv's using PLN (d) and (g), because

$$
\sqrt{T}\,\bar{g}_T(\beta_0, \pi_0) = \begin{bmatrix} \frac{1}{\sqrt{T}}\Sigma_1^T(U_t^2 - U_t^{*2}) \\ \frac{1}{\sqrt{T}}\Sigma_1^T J^{-1}(X_t - \pi_{20}(Z_t))U_t \end{bmatrix}. \qquad (I.3)
$$

The last part of As. 1(f) holds by PLN (a), (b), (d), and (g) using the pseudo-metric ρ^* defined in (2.2.5), because

$$
\begin{aligned}
\rho^*((\hat{\beta},\hat{\pi}), (\beta_0, \pi_0)) &\le \Bigg[\Big[\int|\hat{\pi}_1(z) - \pi_{10}(z)|^2 dP(z)\Big]^{1/2} \\
&\quad + 2\|\hat{\beta} - \beta_0\|\Big[E\|X_t\|^2\Big]^{1/2} + \|\hat{\beta} - \beta_0\|\Big[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^2 dP(z)\Big]^{1/2} \\
&\quad + \|\beta_0\|\Big[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^2 dP(z)\Big]^{1/2}\Bigg] \qquad (I.4) \\
&\quad \times \Bigg[\Big[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^2 dP(z)\Big]^{1/2} + \Big[E\|X_t\|^2\Big]^{1/2}\Bigg] \\
&\quad + \Big[EU_t^2\Big]^{1/2}\Big[\int\|\hat{\pi}_2(z) - \pi_{20}(z)\|^2 dP(z)\Big]^{1/2} \xrightarrow{p} 0\,.
\end{aligned}
$$

Lemma 1 and PLN (h) now give the desired result noting that $R(\beta_0, \pi_0) = E\frac{\partial}{\partial\beta'}r_t(\beta_0, \pi_0) = -2EU_t[X_t - \pi_{20}(Z_t)] = \underset{\sim}{0}$. □

FOOTNOTES

1. The first and second authors gratefully acknowledge the research support of the Cowles Foundation via the Carl Arvid Anderson Prize Fellowship and the National Science Foundation via grant number SES–8821021 respectively.

2. We say that $\left\{\frac{\partial}{\partial\beta'}r_t(\beta,\pi) : t \geq 1\right\}$ satisfies a uniform WLLN over $B_0 \times \Pi$ if $E\frac{\partial}{\partial\beta'}r_t(\beta,\pi)$ exists $\forall\beta \in B_0$, $\forall\pi \in \Pi$, $\forall t \geq 1$ and $\sup_{(\beta,\pi)\in B_0\times\Pi}\left\|\frac{1}{T}\Sigma_1^T\left[\frac{\partial}{\partial\beta'}r_t(\beta,\pi) - E\frac{\partial}{\partial\beta'}r_t(\beta,\pi)\right]\right\| \xrightarrow{p} 0$. We say that $R(\beta,\pi) = \lim_{T\to\infty}\frac{1}{T}\Sigma_1^T E\frac{\partial}{\partial\beta'}r_t(\beta,\pi)$ exists uniformly over $B_0 \times \Pi$ if $\sup_{(\beta,\pi)\in B_0\times\Pi}\left\|\frac{1}{T}\Sigma_1^T E\frac{\partial}{\partial\beta'}r_t(\beta,\pi) - R(\beta,\pi)\right\| \longrightarrow 0$.

3. It appears that Koenker's (1981) result is not correct as stated in his Theorem because the covariance matrix of the limit distribution of $\sqrt{n}\,\hat{\alpha}$ does not equal ϕD^{-1} in general, where $\hat{\alpha}$ is defined in his equation (2.8), $\phi = Var(\epsilon^2)$, and $D = \lim_{n\to\infty}\frac{1}{n}\Sigma_1^n z_i z_i'$. This occurs because $\sqrt{n}(Z'Z)^{-1}Z'v_4$ is not $o_p(1)$. This term affects the limit distribution of $\sqrt{n}\,\hat{\alpha}$ and hence the asymptotic expansion given in his equation (2.9) is not justified in general. In fact, one can show that $\sqrt{n}\,\hat{\alpha} \xrightarrow{d} N(\alpha_0, V)$, where $V = \phi[D^{-1} - D^{-1}AA'D^{-1}]$ and $A = \lim_{n\to\infty}\frac{1}{n}\Sigma_1^n z_i$. The latter result implies that nR^2 has an asymptotic χ^2 distribution with p degrees of freedom, where p is the dimension of z_i and R^2 is the coefficient of determination from a regression of $\hat{U}_i^2$ against unity and z_i.

REFERENCES

Amemiya, T., 1973, Regression analysis when the dependent variable is truncated normal, Econometrica 41, 997–1016.

Andrews, D. W. K., 1987, Consistency in nonlinear econometric models: A generic uniform law of large numbers, Econometrica 55, 1465–1472.

Andrews, D. W. K., 1988, Laws of large numbers for dependent non–identically distributed random variables, Econometric Theory 4, 458–467.

Andrews, D. W. K., 1990a, Asymptotics for semiparametric econometric models: I. Estimation and testing, Cowles Foundation Discussion Paper No. 908R (Yale University, New Haven, CT).

Andrews, D. W. K., 1990b, Asymptotics for semiparametric econometric models: II. Stochastic equicontinuity and nonparametric kernel estimation, Cowles Foundation Discussion Paper No. 909R (Yale University, New Haven, CT).

Andrews, D. W. K., 1990c, Generic uniform convergence, Cowles Foundation Discussion Paper No. 940 (Yale University, New Haven, CT).

Andrews, D. W. K., 1990d, Heteroskedasticity and autocorrelation consistent covariance matrix estimators for semiparametric models, manuscript in preparation (Cowles Foundation, Yale University, New Haven, CT).

Andrews, D. W. K., 1991a, Asymptotic normality of series estimators for nonparametric and semiparametric regression models, Econometrica 59, forthcoming.

Andrews, D. W. K., 1991b, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59, forthcoming.

Andrews, D. W. K. and J. C. Monahan, 1990, An improved heteroskedasticity and auto–correlation consistent covariance matrix estimator, Cowles Foundation Discussion Paper No. 942 (Yale University, New Haven, CT).

Arabmazar, A. and P. Schmidt, 1982, An investigation of the robustness of the Tobit estimator to non–normality, Econometrica 50, 1055–1063.

Breusch, T. S. and A. R. Pagan, 1979, A simple test for heteroskedasticity and random coefficient variation, Econometrica 47, 1287–1294.

Chamberlain, G., 1986, Notes on semiparametric regression, unpublished manuscript (Department of Economics, Harvard University, Cambridge, MA).

Cox, D. R., 1961, Tests of separate families of hypotheses, in: Proceedings of the Fourth berkeley symposium in mathematical statistics and probability (University of California Press, Berkeley, CA) 105–123.

Cox, D. R., 1962, Further results on tests of separate families of hypotheses, Journal of the Royal Statistical Society, Series B 24, 406–424.

Gallant, A. R., 1987, Nonlinear statistical models (Wiley, New York, NY).

Goldberger, A. S., 1983, Abnormal selection bias, in: S. Karlin, T. Amemiya and L. A. Goodman, eds., Studies in econometrics, time series and multivariate statistics (Academic Press, New York, NY).

Hall, P. and C. C. Heyde, 1980, Martingale limit theory and its applications (Academic Press, New York, NY).

Hausman, J. A., 1978, Specification tests in econometrics, Econometrica 46, 1251–1272.

Heckman, J. J., 1979, Sample selection bias as a specification error, Econometrica 47, 153–161.

Herrndorf, N., 1984, An invariance principle for weakly dependent sequences of random variables, Annals of Probability 12, 141–153.

Ichimura, H., 1985, Semiparametric least squares estimation of single index models, unpublished manuscript (Department of Economics, University of Minnesota, Minneapolis, MN).

Jennrich, R., 1969, Asymptotic properties of nonlinear least squares estimators, Annals of Mathematical Statistics 40, 633–643.

Klein, R. W. and R. H. Spady, 1987, An efficient semiparametric estimator for discrete choice models, unpublished manuscript (Economics Research Group, Bell Communications Research, Morristown, NJ).

Koenker, R., 1981, A note on studentizing a test for heteroscedasticity, Journal of Econometrics 17, 107–112.

McLeish, D. L., 1975, Invariance principles for dependent variables, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 32, 165–178.

McLeish, D. L., 1977, On the invariance principle for nonstationary mixingales, Annals of Probability 5, 616–621.

Newey, W. K., 1985a, Maximum likelihood specification testing and conditional moment tests, Econometrica 53, 1047–1070.

Newey, W. K., 1985b, Generalized method of moments specification testing, Journal of Econometrics 29, 229–256.

Newey, W. K., 1988, Two step estimation of sample selection models, unpublished manu– script (Department of Economics, Princeton University, Princeton, NJ).

Newey, W. K., 1989, Uniform convergence in probability and stochastic equicontinuity, unpublished manuscript (Department of Economics, Princeton University, Princeton, NJ).

Newey, W. K. and K. D. West, 1987, A simple positive semi–definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703–708.

Pagan, A. R. and A. D. Hall, 1983, Diagnostic tests as residual analysis, Econometric Reviews 2, 159–218.

Pagan, A. R. and F. Vella, 1989, Diagnostic tests for models based on individual data: A survey, Journal of Applied Econometrics 4, S29–S60.

Pollard, D., 1984, Convergence of stochastic processes (Springer–Verlag, New York, NY).

Pötscher, B. M. and I. R. Prucha, 1989, A uniform law of large numbers for dependent and heterogeneous data processes, Econometrica 57, 675–683.

Powell, J. L., 1984, Least absolute deviations estimation for the censored regression model, Journal of Econometrics 25, 303–325.

Powell, J. L., 1987, Semiparametric estimation of bivariate latent variable models, SSRI Working Paper No. 8704 (University of Wisconsin, Madison, WI).

Robinson, P. M., 1988, Root–N–consistent semiparametric regression, Econometrica 56, 931–954.

Ruud, P. A., 1984, Tests of specification in econometrics, Econometric Reviews 3, 211–242.

Tauchen, G., 1985, Diagnostic testing and evaluation of maximum likelihood models, Journal of Econometrics 30, 415–443.

Whang, Y.–J. and D. W. K. Andrews, 1990, Tests of specification for parametric and semiparametric models: Theory and examples, unpublished manuscript (Cowles Foundation, Yale University, New Haven, CT).

White, H., 1982, Maximum likelihood estimation of misspecified models, Econometrica 50, 1–24.

White, H., 1984, Asymptotic theory for econometricians (Academic Press, Orlando, FL).

White, H., 1987, Specification testing in dynamic models, in: T. Bewley, ed., Advances in econometrics: Fifth world congress, Vol. I (Cambridge University Press, Cambridge).

Wooldridge, J. M., 1989, Some results on specification testing against nonparametric alternatives, unpublished manuscript (Department of Economics, M.I.T., MA).

Wooldridge, J. M. and H. White, 1988a, Central limit theorems for dependent, hetero–geneous processes with trending moments, unpublished manuscript (Department of Economics, M.I.T., MA).

Wooldridge, J. M. and H. White, 1988b, Some invariance principles and central limit theorems for dependent heterogeneous processes, Econometric Theory 4, 210–230.

Yatchew, A. J., 1988, Nonparametric regression tests based on an infinite dimensional least squares procedure, unpublished manuscript (Department of Economics, University of Toronto, Ontario, Canada).