COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 908R

ASYMPTOTICS FOR SEMIPARAMETRIC ECONOMETRIC MODELS: I. ESTIMATION AND TESTING

by Donald W. K. Andrews

August 1988

Revised: May 1990

ASYMPTOTICS FOR SEMIPARAMETRIC ECONOMETRIC MODELS: I. ESTIMATION AND TESTING

Donald W. K. Andrews[1]

Cowles Foundation for Research in Economics
Yale University

August 1988

Revised: May 1990

ABSTRACT

This paper provides a general framework for proving the $\sqrt{T}$–consistency and asymptotic normality of a wide variety of semiparametric estimators. The results apply in time series and cross–sectional modelling contexts. The class of estimators considered consists of estimators that can be defined as the solution to a minimization problem based on a criterion function that may depend on a preliminary infinite dimensional nuisance parameter estimator. The criterion function need not be differentiable. The method of proof exploits results concerning the stochastic equicontinuity or weak convergence of normalized sums of stochastic processes.

This paper also considers tests of nonlinear parametric restrictions in semiparametric econometric models. To date, only Wald tests of such restrictions have been considered in the literature. Here, Wald, Lagrange multiplier, and likelihood ratio–like tests statistics are considered. A general framework is provided for proving that these test statistics have asymptotic chi–square distributions under the null hypothesis and local alternatives. The results hold for a wide variety of underlying estimation techniques and in a wide variety of model scenarios.

JEL Classification Number: 211.

Keywords: Asymptotic normality, empirical process, infinite dimensional nuisance parameter, Lagrange multiplier test, likelihood ratio–like test, nonparametric estimation, semiparametric estimation, semiparametric model, semiparametric test, stochastic equicontinuity, Wald test, weak convergence.

1. INTRODUCTION

Semiparametric models and estimation procedures have become increasingly popular in econometrics in recent years. A large number of semiparametric estimators have been introduced and many have been shown to be $\sqrt{T}$–consistent and asymptotically normal. The proofs of such results are given in the literature on a case by case basis. No general results are available. The purpose of this paper is to provide a general framework for establishing the $\sqrt{T}$–consistency and asymptotic normality of a wide class of semiparametric estimators for time series, cross–section, and panel data models.

A consequence of the recent development of estimation results for semiparametric models is an increase in the need for tests of parametric restrictions in such models. To date, only Wald tests of such restrictions have been considered in the literature (e.g., see Robinson (1989)). A second purpose of this paper is to provide a general framework for establishing the asymptotic chi–square distributions of Wald (W), Lagrange multiplier (LM), and likelihood ratio–like (LR) tests of nonlinear parametric restrictions in a wide variety of different semiparametric models.

The estimators considered in this paper are called MINPIN estimators. They are estimators that MINimize a criterion function that may depend on a Preliminary Infinite dimensional Nuisance parameter estimator. The criterion function need not be differentiable. As it happens, many of the semiparametric (and parametric) estimators in the literature are MINPIN estimators. Examples are given below.

The test statistics considered in this paper are W, LM, and LR statistics that are based on MINPIN estimators and MINPIN criterion functions. In consequence, all three tests apply in a wide variety of different underlying estimation contexts. In particular, the "LR" test applies in non–maximum likelihood scenarios. The testing results, like the estimation results, apply in time series, cross–section, and panel data scenarios. An attractive feature of the testing results is that few assumptions are needed beyond those used to obtain asymptotic normality of the underlying MINPIN estimators.

The method of proof used here employs a condition called stochastic equicontinuity. This condition can be verified directly or by establishing a central limit theorem (CLT) for the normalized sums of stochastic processes, such as empirical processes. An important feature of the method used is its generality. The same method can be used with a wide variety of estimators in different semiparametric models. The same method can be applied with independent identically distributed (iid), independent non-identically distributed (inid), and dependent non-identically distributed (dnid) random variables (rv's). Thus, the method used here can be utilized for attacking new semiparametric estimation and testing problems and for extending the domain for applicability of existing results, e.g., to dnid scenarios.[2]

A second feature of the method used is that the assumptions on the infinite dimensional nuisance parameter estimator and on the random criterion function are separated. There is no need to use sample splitting procedures or to introduce complicated arguments to circumvent the problem caused by the dependence between the nuisance parameter estimator and the criterion function. In consequence, the results given below are flexible regarding the choice of estimator of the infinite dimensional nuisance parameter. In contrast, most results in the econometrics literature specify a particular estimator for the nuisance parameter even though other estimators are just as suitable.

A third feature of the method used here is the simplicity of the structure of the proof. Many proofs in the literature are long and complicated. In some cases, the key steps of the proof are obscured by the details. With the method used here, the key steps are highlighted and compartmentalized. The results given here, however, do not provide a complete proof of asymptotic normality—they assume, rather than prove, that an estimator of an infinite dimensional nuisance parameter has certain properties. In consequence, direct comparisons of the complexity of proofs may be misleading.

A fourth feature of the method used here is the flexibility it affords with respect to the type of estimator considered. Many results in the semiparametric literature apply only to one-step estimators because of their technical tractability, among other reasons (e.g., see examples (1), (4), (7), and (8) below). The results of this paper apply to one-step versions of estimators as well as to the pure minimization versions of the estimators. One consequence of this is that LM and LR tests of parametric restrictions can be constructed in semiparametric contexts.

On the other hand, a drawback of the method used here is that in some examples it requires more smoothness conditions on certain underlying unknown functions than are necessary for $\sqrt{T}$-consistency and asymptotic normality of the estimator in question. (The class of examples for which this is true is characterized below.) For some of these examples, alternative proofs are available in the literature that do not rely on such smoothness conditions. Examples are given below.

A second drawback of the method used here arises in those examples where trimming of nonparametric function estimators is required. In such examples, the method used here places more restrictions on the form of trimming that can be used than is necessary. This drawback and the previous one are consequences of the stochastic equicontinuity results that are currently available; they are not intrinsic to the method used here. It is possible that future developments of stochastic equicontinuity results will ameliorate these drawbacks.

A third drawback of the method used here is that, while the method is quite general, it is not applicable to all semiparametric estimators that are $\sqrt{T}$-consistent and asymptotically normal. Examples are given below. A method for extending the results to cover some of these examples is mentioned briefly.

The results given in this paper are proved under a set of "high-level" assumptions. 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 CLT, and a stochastic equicontinuity condition for certain random variables and/or stochastic processes. The reasons for adopting such assumptions, rather than more primitive assumptions, are the following:

First, the high-level assumptions clarify those features of the infinite dimensional nuisance parameter estimator that are important for its successful use in the semiparametric estimation procedure.

Second, the high–level assumptions provide for greater flexibility regarding the choice of the preliminary nuisance parameter estimator than would be obtained otherwise.

Third, the isolation of the role played by the ULLN, CLT, and stochastic equicontinuity condition helps one to understand the scope of the results, and in particular, the effect of temporal dependence on the results.

Fourth, the flexibility to choose from numerous existing ULLNs, CLTs, and stochastic equicontinuity results yields greater generality and/or simplicity than the alternative of specifying a single ULLN, CLT, and stochastic equicontinuity result and specifying primitive assumptions under which it holds. The reason is simply that depending upon the modelling context – time series, cross section, or panel data – and upon the application of interest, different ULLNs, CLTs, and stochastic equicontinuity results may be the most suitable for use.

Fifth, the results given here using high–level assumptions can exploit continual improvements in asymptotic results for infinite dimensional nuisance parameter estimators, such as nonparametric regression and density estimators, and improvements in ULLNs, CLTs, and stochastic equicontinuiuty results. Especially for dnid contexts, these improvements are likely to be substantial in the future.

A consequence of the use of high–level assumptions is that the results given here provide a <u>general</u> <u>framework</u> for proving the asymptotic results rather than a <u>complete</u> <u>proof</u> of such results. To obtain the latter for a given example of interest, one needs to specify primitive conditions under which the high–level assumptions are satisfied. For the

ULLN and CLT conditions this is relatively easy, because there are numerous ULLN and CLT results in the literature that are suitable without alteration. In addition, a sequel to this paper Andrews (1989a), hereafter referred to as ASEM:II, provides primitive conditions under which the stochastic equicontinuity condition holds.

Thus, the remaining "high–level" assumptions that require verification concern the properties of the infinite dimensional nuisance parameter estimator. For the case of kernel regression and density estimators, ASEM:II provides results that establish the requisite properties. Thus, when the infinite dimensional nuisance parameter is estimated using kernel estimators, the present paper and ASEM:II provide all the ingredients necessary to establish the consistency and asymptotic normality of MINPIN estimators and the asymptotic chi–square distribution of the corresponding W, LM, and LR test statistics. When the nuisance parameter is estimated by some method other than kernel estimation, a wide variety of results in the literature on nonparametric regression and density estimation can be exploited for this purpose. Some references are given below. These results alone, however, are not always sufficient to complete the proof of asymptotic normality. In many cases, special tailoring of existing nonparametric results is required to verify the desired properties of the preliminary infinite dimensional nuisance parameter estimators.

We now specify a number of examples of estimators that fall within the MINPIN class. Those marked with an asterisk are discussed in the paper.

(1)* Efficient generalized method of moments (GMM) and one–step GMM estimators of parameters defined by conditional moment restrictions. For the latter, see Newey (1987, 1990a). Included in this class are weighted least squares (LS) estimators for linear and nonlinear regression models that adapt to heteroskedasticity of unknown form, see Carroll (1982), Robinson (1987), and Delgado (1988a,b), and weighted instrumental variable estimators for simultaneous equations models.

(2)* Semiparametric LS and weighted LS estimators of partially linear regression models. For the former, see Robinson (1988), Chamberlain (1986), and Andrews (1991a).

(3)* Semiparametric instrumental variable estimators for regression models with unobserved risk variables, see Pagan and Ullah (1988).

(4)* Efficient weighted censored least absolute deviations (WC–LAD) estimators and one–step versions of them for the censored regression model. For the latter, see Newey and Powell (1990).

(5)* MAD–DUC estimators of index regression models. Included in this class are Klein and Spady's (1987) efficient semiparametric estimator of the binary choice model and Ichimura's (1985) and Ichimura and Lee's (1990) LS estimators of single and multiple index models.

(6)* Two–step and three–step estimators of the sample selection model. For the former, see Powell (1987) and Newey (1988).

(7)* Adaptive estimators for regression models with errors of unknown distribution. Included in this class are pure minimization and one–step estimators of linear and nonlinear regression models with (i) iid errors, (ii) independent symmetrically distributed errors, and (iii) stationary, homoskedastic, r–th order Markov errors. See Bickel (1982) and Manski (1984) for one–step estimators for the independent error models.

(8) Pure minimization and one–step adaptive estimators of autoregressive moving average models with innovations with unknown distribution. For one–step estimators, see Kreiss (1987).

(9) Profile likelihood estimators for semiparametric models, see Severini and Wong (1987b) and Lee (1989).

(10) M–estimators with non–differentiable ψ functions, including Huber (1973) ψ function regression estimators, Koenker and Bassett's (1978) regression quantiles, Ruppert and Carroll's (1980) trimmed LS regression estimators, Powell's (1984, 1986a,b, 1990) least absolute deviations (LAD), quantile, and trimmed LS

estimators of censored, monotonic, and truncated regression models, and Bates and White's (1988) weighted M–estimators for regression models.

(11) Method of simulated moments estimators, including those of McFadden (1989), Pakes and Pollard (1989), and Laroque and Salanie (1989).

(12) Parametric estimators that minimize criterion functions that are differentiable and that may depend on a finite dimensional nuisance parameter estimator, including the classes of estimators considered by Bierens (1981), Burguete, Gallant, and Souza (1982), Gallant (1987b), Gallant and White (1988), Andrews and Fair (1988), and Pōtscher and Prucha (1990).

A useful feature of the results given here is that in the common parametric case described in example (12), the results are comparable to existing results in the econometrics literature (e.g., Andrews and Fair (1988))—the assumptions are no more difficult to verify.

The results of this paper do not cover most estimators of time series models that have deterministic or stochastic trends. The results do not cover nonparametric estimators. In special cases, the results cover some finite dimensional sub–vectors of the seminonparametric estimators considered by Gallant (see Gallant (1987a) and references therein), but in general Gallant's seminonparametric estimators are not covered.

In addition, except for the consistency results, the results do not cover Manski's (1975, 1985) maximum score estimator, Horowitz's (1989) smoothed maximum score estimator, or Cosslett's (1983) semiparametric maximum likelihood estimator of the binary choice model, Cox's (1975) partial likelihood estimator of the proportional hazard model, Han's (1987) maximum rank correlation estimator of generalized regression models, Horowitz's (1988) M–estimator of the censored regression model, or Powell, Stock, and Stoker's (1989) or Andrews' (1991a) estimators of index regression models.

The maximum score estimator is not covered because it does not solve a set of first order conditions with probability that goes to one as $T \to \infty$, as is required for the asymptotic normality results. In fact, this estimator is not $\sqrt{T}$–consistent or asymptotically normal, see Kim and Pollard (1990). In consequence, the failure of the method used in this paper is to be expected. Horowitz's smoothed maximum score estimator is not covered, because the sample average of summands that comprise the first order conditions for this estimator is not asymptotically normal with a $\sqrt{T}$–normalization. This estimator as well is not $\sqrt{T}$–consistent.[3]

The other estimators listed in the paragraph above are not covered by the results of this paper because they fail an asymptotic orthogonality condition (Assumption 2(c) below) between the finite dimensional and infinite dimensional parameter estimators. This condition is the most restrictive of the conditions imposed below to obtain asymptotic normality. It is sufficiently general, however, to cover many estimators, as examples (1)–(10) above illustrate. In particular, it does not restrict the results to adaptive estimators of adaptive models. The orthogonality condition is discussed at length below.

This paper does not discuss semiparametric asymptotic efficiency bounds. The bounds in some semiparametric models are obvious, e.g., see examples (7) and (8) above. In other models, such as those of examples (1), (2), (4), (5), and (6), the asymptotic efficiency bounds are not obvious, but have been determined in the literature. See Begun, Hall, Huang, and Wellner (1983), Bickel, Klaasen, Ritov, and Wellner (1988), and references therein for general results concerning semiparametric asymptotic efficiency bounds. See Newey (1990b) for a review of such bounds with an emphasis on econometric models.

This paper also does not cover specification tests, except those that are of the W, LM, or LR type. A treatment of specification tests in semiparametric models is currently being developed by Whang and Andrews (1990) and will be reported elsewhere.

The remainder of this paper is organized as follows: Section 2 defines the class of MINPIN estimators and provides consistency results for them. Section 3 gives $\sqrt{T}$–consistency and asymptotic normality results. Section 4 introduces consistent covariance matrix estimators for MINPIN estimators. Section 5 introduces the W, LM, and LR test statistics, presents conditions under which they have chi–square asymptotic distributions under the null hypothesis, outlines conditions under which their defining expressions simplify, and gives local power results for the tests. Examples (1) and (2) above are used throughout Sections 2–5 to illustrate the results. Section 6 discusses several additional examples, viz., examples (3)–(7) 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. CONSISTENCY

In this section we define the MINPIN class of estimators and give sufficient conditions for their consistency. We also introduce two examples that are used throughout the paper to illustrate the results given and discuss sufficient conditions for consistency in these examples.

2.1. Definition of MINPIN Estimators

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

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

$$(2.1) \qquad d(\bar{m}_T(\hat{\theta},\hat{\tau}), \hat{\gamma}) = \inf_{\theta\in\Theta} d(\bar{m}_T(\theta,\hat{\tau}), \hat{\gamma})$$

with probability $\rightarrow$ 1, where $\bar{m}_T(\theta,\tau) = \frac{1}{T}\Sigma_1^T m_t(\theta,\tau)$, $m_t(\theta,\tau)$ denotes $m_{Tt}(W_{Tt}, \theta, \tau)$, $m_{Tt}(\cdot,\cdot,\cdot)$ is a function from $R^{k_{Tt}} \times \Theta \times \mathcal{T}$ to R^v, k_{Tt} is a positive integer $\leq \infty$, $\hat{\tau}$ is a random element of $\mathcal{T}$ with probability $\rightarrow$ 1, $\hat{\gamma}$ is a random element of Γ (and $\hat{\tau}$ and $\hat{\gamma}$ depend on T in general), Θ, $\mathcal{T}$, and Γ are pseudo–metric spaces, and $d(\cdot,\cdot)$ is a non–random, real–valued function (which does not depend on T).[4]

Throughout this paper, all functions that are introduced (such as $\hat{\theta}$, $\hat{\tau}$, $\hat{\gamma}$, $m_{Tt}(\cdot,\cdot,\cdot)$, and $d(\cdot,\cdot)$) 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.

Note that $\hat{\tau}$ and $\hat{\gamma}$ are preliminary, possibly infinite dimensional, estimators used in the definition of $\hat{\theta}$. In almost all examples, however, either no preliminary estimator $\hat{\gamma}$ appears or it is finite dimensional. For the asymptotic normality and testing results given below, Θ is taken to be a subset of R^p. This structure is not needed, however, for the consistency results of this section.

2.2. GMM/CMR and WLS/PLR Examples

We now discuss two estimators in terms of the above definition of MINPIN estimators. Other examples are given in Section 6 below. The estimators that we discuss here are a GMM estimator of parameters defined by conditional moment restrictions (CMR) and a weighted least squares (WLS) estimator of the partially linear regression (PLR) model. These estimators are chosen for several reasons. First, consistency and asymptotic normality results are not available in the literature for either of them. Second, they can be used below to illustrate the two different sets of testing results given in Section 5. Third, they illustrate the two different cases that arise regarding a key orthogonality condition that is used to obtain asymptotic normality.

For each of the two examples, and in the additional examples given in Section 6 below, we distinguish between "assumptions for which the estimator is designed" and assumptions that are imposed for the asymptotic results to hold. The former refer to assumptions that are not necessarily imposed but that motivate the choice of estimator and sometimes correspond to assumptions under which the estimator is asymptotically efficient.

We consider first the CMR model. Chamberlain (1987a) establishes the semiparametric asymptotic efficiency bound for this model when the observations are iid. Newey (1987) establishes the asymptotic normality and efficiency of a "one-step" GMM estimator of this model for the iid case. Here, we consider the "pure" GMM estimator, which is designed for independent observations, and consider its behavior for *dnid* observations.

In the CMR model, θ_0 is defined to be the unique parameter vector that solves the equations:

$$E(\psi(Z_t, \theta) | X_t) = \underset{\sim}{0} \text{ a.s. } \forall t \geq 1 \tag{2.2}$$

for some specified R^α-valued function $\psi(\cdot,\cdot)$, where $X_t \in R^{k_a}$. The observations $W_t = (Z_t', X_t')'$ may be dnid. Examples of this model in econometrics are quite numerous, see Chamberlain (1987a) and Newey (1987). One example is the parametric nonlinear regression model with errors that are heteroskedastic of unknown form:

$$Y_t = f(X_t, \theta_0) + U_t,\ E(U_t | X_t) = 0 \text{ a.s., and } Var(U_t | X_t) = \Omega_0(X_t) \text{ a.s.} \tag{2.3}$$

for $t = 1, \ldots, T$, for some known function $f(\cdot,\cdot)$ and some unknown function $\Omega_0(\cdot)$. In this case, $Z_t = (Y_t, X_t')'$ and $\psi(Z_t, \theta) = Y_t - f(X_t, \theta)$. Carroll (1982), Robinson (1987), and Delgado (1988b) consider weighted LS estimators of different versions of this model with weights given by a preliminary estimator $\hat{\tau}(X_t) = \frac{\partial}{\partial\theta}f(X_t, \hat{\theta})/\hat{\Omega}(X_t)$ of $\tau_0(X_t) = \frac{\partial}{\partial\theta}f(X_t, \theta_0)/\Omega_0(X_t)$.

For the CMR model, we define

$$
\begin{aligned}
&\Omega_0(X_t) = E(\psi(Z_t,\theta_0)\psi(Z_t,\theta_0)'\,|\,X_t) \in R^{\alpha} \times R^{\alpha}, \\
&\Delta_0(X_t) = E\left[\frac{\partial}{\partial\theta'}\psi(Z_t,\theta_0)\,|\,X_t\right] \in R^{\alpha} \times R^{p}, \\
(2.4)\qquad &\tau_0(X_t) = \Delta_0(X_t)'\Omega_0^{-1}(X_t) \in R^{p} \times R^{\alpha}, \text{ and} \\
&\gamma_0 = \left[\lim_{T\to\infty} \tfrac{1}{T}\Sigma_1^T E\Delta_0(X_t)'\Omega_0^{-1}(X_t)\Delta_0(X_t)\right]^{-1} \in R^{p} \times R^{p}.
\end{aligned}
$$

By assumption, the functions $\Omega_0(\cdot)$, $\Delta_0(\cdot)$, and $\tau_0(\cdot)$ do not depend on t.

We define a GMM estimator $\hat{\theta}$ as a MINPIN estimator by taking

$$
\begin{aligned}
&d(m,\gamma) = m'\gamma m/2,\ m_t(\theta,\tau) = \tau(X_t)\psi(Z_t,\theta), \\
(2.5)\qquad &\hat{\tau}(X_t) = \hat{\Delta}(X_t)'\hat{\Omega}^{-1}(X_t), \text{ and } \hat{\gamma} = \left[\tfrac{1}{T}\Sigma_1^T\hat{\Delta}(X_t)'\hat{\Omega}^{-1}(X_t)\hat{\Delta}(X_t)\right]^{-1},
\end{aligned}
$$

where $\hat{\Delta}(\cdot)$ and $\hat{\Omega}(\cdot)$ are estimators of $\Delta_0(\cdot)$ and $\Omega_0(\cdot)$ respectively.[5] For example, Newey (1987) considers nearest neighbor and instrumental variable estimators of $\Delta_0(\cdot)$ and $\Omega_0(\cdot)$. In the nonlinear regression model of (2.3), $\Delta_0(X_t) = \frac{\partial}{\partial\theta'}f(X_t,\theta_0)$, γ_0 is as above or is equal to I_p, and the above GMM estimator is the weighted LS estimator.

For the case of iid observations, the GMM estimator attains the semiparametric efficiency bound for the CMR model (under suitable assumptions).

In some cases, it is convenient to trim the nonparametric estimator $\hat{\tau}(x)$ such that it equals zero outside a bounded set $\mathcal{X}^*$. In this case, $\Delta_0(X_t)$ and $\tau_0(X_t)$ of (2.4) each need to be multiplied by $1(X_t \in \mathcal{X}^*)$. There are two reasons for trimming. First, trimming can eliminate observations from the computation of $\hat{\theta}$ for which the nuisance parameter estimator $\hat{\tau}(X_t)$ is estimated with relatively large error in comparison to the non-trimmed observations. Second, trimming makes it much simpler to establish the consistency and asymptotic normality of $\hat{\theta}$, because one can obtain uniform consistency of $\hat{\tau}(x)$ for $\tau_0(x)$ over a bounded set $\mathcal{X}^*$ under suitable conditions, but not over unbounded sets in general. On the other hand, trimming using a single fixed set $\mathcal{X}^*$ affects the asymptotic distribution of a MINPIN estimator $\hat{\theta}$ and usually sacrifices some asymptotic efficiency.

For simplicity, we do not make trimming explicit in expressions given for $m_t(\theta,\tau)$ and other quantities in this example or in others below. If trimming is carried out, then indicator functions need to be added in the appropriate places.

Next, we consider the PLR model:

$$Y_t = X_t'\theta_0 + g(Z_t) + U_t \text{ and } E(U_t|X_t, Z_t) = 0 \text{ a.s.} \tag{2.6}$$

for $t = 1, \ldots, T$, where the real function g is unknown, $W_t = (Y_t, X_t', Z_t', U_t)'$ is identically distributed for $t \geq 1$, $Y_t, U_t \in R$, $X_t, \theta_0 \in R^p$, and $Z_t \in R^{k_a}$. Chamberlain (1987b) establishes the semiparametric asymptotic efficiency bound for estimating θ_0 in this model when the observations are iid and the errors are square integrable. Robinson (1988), Chamberlain (1986), and Andrews (1991a) establish the asymptotic normality of different LS estimators of θ_0 for this model. These LS estimators are obtained by regressing Y_t on X_t after Y_t and X_t have been purged of their correlation with Z_t by subtracting off nonparametric estimates of their conditional expectations given Z_t.

Here we consider a WLS estimator of θ_0 that is defined analogously to the LS estimator mentioned above, but is designed for the case where the conditional variance of U_t given (X_t, Z_t) depends on Z_t. To motivate this estimator, we note that the PLR model with heteroskedasticity of this form is generated by a sample selection model. In particular, suppose

$$\begin{aligned} &\tilde{Y}_t = \tilde{X}_t'\theta_0 + \tilde{Z}_t'\phi_0 + \tilde{U}_t, \ D_t = 1(h(\tilde{Z}_t, \tilde{\epsilon}_t) > 0), \text{ and} \\ &(Y_t, D_t, X_t, Z_t) = (\tilde{Y}_t D_t, D_t, \tilde{X}_t D_t, \tilde{Z}_t D_t) \text{ are observed for } t = 1, \ldots, T, \end{aligned} \tag{2.7}$$

where the real function $h(\cdot,\cdot)$ is unknown, $(\tilde{U}_t, \tilde{\epsilon}_t, \tilde{X}_t, \tilde{Z}_t)$ is identically distributed for $t \geq 1$, and $(\tilde{U}_t, \tilde{\epsilon}_t)$ is independent of $(\tilde{X}_t, \tilde{Z}_t)$ and has unknown distribution. By multiplying the first equation of (2.7) by D_t, one sees that the sample selection model (2.7) generates the PLR model (2.6) with the unknown function $g(\cdot)$ of (2.6) given by $g(v) = v'\phi_0 + E(D_t\tilde{U}_t | Z_t = v)$ and with the error U_t of (2.6) given by $U_t = D_t\tilde{U}_t - E(D_t\tilde{U}_t | Z_t)$. In this case, the error U_t has conditional variance given (X_t, Z_t) that depends on Z_t alone, which motivates the use of a WLS estimator with weights that depend only on Z_t.

For the PLR model of (2.6), define

$$\begin{aligned} &\tau_{10}(Z_t) = E(Y_t | Z_t), \ \tau_{20}(Z_t) = E(X_t | Z_t), \ \tau_{30}(Z_t) = E(U_t^2 | Z_t), \text{ and} \\ &\tau_0 = (\tau_{10}, \tau_{20}', \tau_{30})' . \end{aligned} \tag{2.8}$$

Let $\hat{\tau}_1(\cdot)$, $\hat{\tau}_2(\cdot)$, and $\hat{\tau}_3(\cdot)$ be estimators of $\tau_{10}(\cdot)$, $\tau_{20}(\cdot)$, and $\tau_{30}(\cdot)$ respectively. We consider the following semiparametric WLS estimator for the PLR model:

$$\hat{\theta} = \left[\Sigma_1^T (X_t - \hat{\tau}_2(Z_t))(X_t - \hat{\tau}_2(Z_t))' / \hat{\tau}_3(Z_t)\right]^{-1} \Sigma_1^T (X_t - \hat{\tau}_2(Z_t))(Y_t - \hat{\tau}_1(Z_t)) / \hat{\tau}_3(Z_t) . \tag{2.9}$$

This estimator can be written as a MINPIN estimator in either of two ways—as an estimator that minimizes a weighted sum of squared residuals or as one that solves the first order conditions of this minimization problem. Correspondingly, for the consistency results below, we can take either

$$d(m,\gamma) = m \text{ and } m_t(\theta,\tau) = (Y_t - \tau_1(Z_t) - (X_t - \tau_2(Z_t))'\theta)^2 / \tau_3(Z_t) , \text{ or} \tag{2.10}$$

$$d(m,\gamma) = m'm/2 \text{ and } m_t(\theta,\tau) = [Y_t - \tau_1(Z_t) - (X_t - \tau_2(Z_t))'\theta][X_t - \tau_2(Z_t)] / \tau_3(Z_t) , \tag{2.11}$$

whichever makes the assumptions for consistency easier to verify.[6] Having obtained consistency, the definition of (2.11) must be used for the asymptotic normality and testing results given below.

We note that Newey (1989a) also discusses WLS estimators for PLR models. He does not establish their asymptotic normality, but derives what their asymptotic covariance matrix must be if they are regular and asymptotically normal and if $\hat{\tau}_3$ and τ_{30} are allowed to depend on both Z_t and X_t.

2.3. Consistency of MINPIN Estimators

We now return to the general case. In what follows we avoid imposing conditions that are used just to ensure measurability of $\hat{\theta}$ by stating results that hold for any sequence of rv's $\{\hat{\theta}\}$. Such results have content only if such a sequence exists. Clearly, sequences $\{\hat{\theta}\}$ that satisfy (2.1), but are not necessarily measurable, exist if Θ is assumed to be compact. Further, we note that one set of sufficient conditions for the existence of a measurable sequence $\{\hat{\theta}\}$ is that $d(\bar{m}_T(\theta,\hat{\tau}), \hat{\gamma})$ viewed as a function from $\Omega \times \Theta$ to R is continuous in θ for each $\omega \in \Omega$ and is measurable for each fixed $\theta \in \Theta$, and Θ is a compact subset of some Euclidean space (see Jennrich (1969), Lemma 2).

Let Θ/Θ_0 denote the set of points θ that are in Θ, but are not in Θ_0.

For consistency of MINPIN estimators we assume the following.

ASSUMPTION 1: (a) *There exists a function* $m(\cdot,\cdot) : \Theta \times \mathcal{T} \to R^v$ *such that* $\bar{m}_T(\theta,\tau) \xrightarrow{p} m(\theta,\tau)$ *uniformly over* $(\theta,\tau) \in \Theta \times \mathcal{T}$.

(b) $\sup_{\theta\in\Theta} \|m(\theta,\hat{\tau}) - m(\theta, \tau_0)\| \xrightarrow{p} 0$ *for some* $\tau_0 \in \mathcal{T}$, $P(\hat{\tau} \in \mathcal{T}) \to 1$, *and* $\hat{\gamma} \xrightarrow{p} \gamma_0$ *for some* $\gamma_0 \in \Gamma$.

(c) $d(m,\gamma)$ *is uniformly continuous over* $\mathcal{M} \times \Gamma_0$, *where* $\mathcal{M} = \{m \in R^v : m = m(\theta,\tau)$ *for some* $\theta \in \Theta, \tau \in \mathcal{T}\}$ *and* Γ_0 $(\subset \Gamma)$ *contains a neighborhood of* γ_0.

(d) *For every neighborhood* Θ_0 $(\subset \Theta)$ *of* θ_0, $\inf_{\theta\in\Theta/\Theta_0} d(m(\theta,\tau_0),\gamma_0) > d(m(\theta_0,\tau_0),\gamma_0)$.[7]

THEOREM I.1: *Under Assumption 1, every sequence of* MINPIN *estimators* $\{\hat{\theta}\}$ *satisfies* $\hat{\theta} \xrightarrow{p} \theta_0$ *under* **P**.

The proof of Theorem I.1 is similar to many other consistency proofs in the literature. It is given in the Appendix along with the proofs of the other results stated below.

2.4. Discussion of Assumption 1

We now discuss Assumption 1. The function $m(\theta,\tau)$ of Assumption 1(a) usually is given by $\lim_{T\to\infty}\frac{1}{T}\Sigma_1^T Em_t(\theta,\tau)$. Thus, Assumption 1(a) holds if $m(\theta,\tau) = \lim_{T\to\infty}\frac{1}{T}\Sigma_1^T Em_t(\theta,\tau)$ exists uniformly over $\Theta \times \mathcal{T}$ (i.e., $\sup_{(\theta,\tau)\in\Theta\times\mathcal{T}}\left\|\frac{1}{T}\Sigma_1^T Em_t(\theta,\tau) - m(\theta,\tau)\right\| \to 0$) and $\{m_t(\theta,\tau) : t \leq T, T \geq 1\}$ satisfies a uniform weak law of large numbers (WLLN) over $\Theta \times \mathcal{T}$ (i.e., $\sup_{\theta\in\Theta, \tau\in\mathcal{T}}\left\|\frac{1}{T}\Sigma_1^T(m_t(\theta,\tau) - Em_t(\theta,\tau))\right\| \xrightarrow{p} 0$). The latter can be verified using stochastic equicontinuity results, such as those given in ASEM:II. Alternatively, it can be verified using the generic uniform WLLN results of Andrews (1987b, 1989b), Pötscher and Prucha (1989), or Newey (1989b) combined with a pointwise WLLN, such as that of Andrews (1988) or McLeish (1975a) for dnid rv's. As a third alternative, it can be verified using empirical process or Banach space WLLN results, such as those of Pollard (1984, Theorems II.2, II.24, and II.25).

The first part of Assumption 1(b) specifies the manner in which $\hat{\tau}$ must converge to τ_0. The condition shows that $\hat{\tau}$ must estimate τ_0 well only in so far as $m(\theta,\hat{\tau})$ estimates $m(\theta, \tau_0)$ well uniformly over $\theta \in \Theta$ for large T. When τ_0 is a function, the latter usually requires L^Q consistency of $\hat{\tau}(\cdot)$ for $\tau_0(\cdot)$ for some $1 \leq Q \leq \infty$, e.g., see the discussion below of the CMR and PLR examples.

If $\hat{\tau}$ is a nonparametric regression or density estimator, then consistency results in the literature for such estimators and/or their proofs can be exploited when verifying the first part of Assumption 1(b). There are numerous L^Q and uniform consistency results available for a variety of different nonparametric estimators for temporally independent and dependent scenarios. For example, for results and references concerning nonparametric regression estimation via (i) kernel, (ii) series, (iii) nearest neighbor, (iv) sieve, (v) locally weighted polynomial regression, and (vi) splines, see (i) ASEM:II, Devroye and Wagner (1980), Prakasa Rao (1983), Bierens (1987), and Györfi, Härdle, Sarda, and Vieu (1989), (ii) Cox (1988) and Andrews (1991a), (iii) Stone (1977), (iv) Grenander (1981), Severini and Wong (1987a), and Wooldridge and White (1990), (v) Stone (1982), and (vi) Eubank (1988) and Wahba (1990) respectively. For results and references concerning nonparametric density estimators, see ASEM:II, Devroye and Györfi (1985), and Silverman (1986).

Note that in many cases the establishment of the first part of Assumption 1(b) involves an additional step that is not treated in the literature concerning consistent nonparametric estimation (although it is treated in ASEM:II for kernel estimators). This additional step arises when the preliminary estimator $\hat{\tau}$ is based on estimated variables rather than the true variables themselves. For example, for the GMM estimator of the CMR model, one cannot regress the elements of $\psi(Z_t, \theta_0)\psi(Z_t, \theta_0)'$ on X_t to estimate $\Omega_0(X_t)$, since θ_0 is unknown. Rather, one has to regress $\psi(Z_t, \theta^*)\psi(Z_t, \theta^*)'$ on X_t, where θ^* is some preliminary consistent estimator of θ_0. In such cases, one has to show that the error introduced by using θ^* rather than θ_0 is $o_p(1)$. This can be done directly on a case by case basis, or by using the results of ASEM:II when kernel estimators are employed, or by using a discretization/contiguity argument as in Bickel (1982, p. 657). (The latter method is also discussed in Manski (1984, pp. 173–178).)

The second part of Assumption 1(b) requires that with probability $\to 1$ $\hat{\tau}$ lies in the set $\mathcal{T}$ over which $\bar{m}_T(\theta,\tau)$ converges uniformly to $m(\theta,\tau)$. There is a tension between this condition and Assumption 1(a), since the larger is $\mathcal{T}$ the easier it is to verify this condition, but the more difficult it is to verify Assumption 1(a) and vice versa. If Assumption 1(a) is verified using a smoothness condition on all $\tau \in \mathcal{T}$, as is the case when τ is infinite dimensional and the stochastic equicontinuity results of ASEM:II are used, then the second part of Assumption 1(b) requires that $\hat{\tau}$ satisfies this smoothness condition with probability $\to 1$. See the discussion of Assumption 2(b) below for further details. Again, consistency results in the nonparametric literature, including those for derivatives, can be exploited when verifying such a condition.

The third part of Assumption 1(b) often holds trivially since no estimator $\hat{\gamma}$ arises. When $\hat{\gamma}$ does arise, it is almost always finite dimensional.

In almost all examples, $d(m,\gamma) = m$, $m'm/2$, or $m'\gamma m/2$. In these cases, a sufficient condition for Assumption 1(c) is $\sup_{\theta\in\Theta,\tau\in\mathcal{T}} \|m(\theta,\tau)\| < \infty$. The latter condition usually is not overly restrictive.

Assumption 1(d) is the uniqueness/identification assumption that ensures that $\{\hat{\theta} : T \geq 1\}$ neither converges to a multi–element subset of Θ nor diverges to "∞." This condition is the same as a uniqueness/identification condition that is often used for nonlinear parametric models. Sufficient conditions for Assumption 1(d) are: Θ is compact and $d(m(\theta, \tau_0), \gamma_0)$ is continuous in θ on Θ and is uniquely minimized at θ_0 .

2.5. Consistency in the GMM/CMR Example

For the GMM estimator of the CMR model, the following assumption implies Assumption 1, and hence, is sufficient for consistency of $\hat{\theta}$:

ASSUMPTION GMM/CMR 1: (a) $\lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E\Delta(X_t)'\Omega^{-1}(X_t)\psi(Z_t,\theta)$ and $\lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E\Delta(X_t)'\Omega^{-1}(X_t)\Delta(X_t)$ exist uniformly over $(\theta,\tau) = (\theta, \Delta'\Omega^{-1}) \in \Theta \times \mathcal{T}$.

(b) $\{\Delta(X_t)'\Omega^{-1}(X_t)\psi(Z_t, \theta) : t \geq 1\}$ and $\{\Delta(X_t)'\Omega^{-1}(X_t)\Delta(X_t) : t \geq 1\}$ satisfy uniform WLLNs over $(\theta,\tau) = (\theta, \Delta'\Omega^{-1}) \in \Theta \times \mathcal{T}$.

(c) $\sup_{N\geq 1} \frac{1}{N}\Sigma_1^N \int \|\hat{\Delta}(x) - \Delta_0(x)\|^r dP_t(x) \xrightarrow{p} 0$ and $\sup_{N\geq 1} \frac{1}{N}\Sigma_1^N \int \|\hat{\Omega}(x) - \Omega_0(x)\|^u dP_t(x) \xrightarrow{p} 0$ for some $2 \leq r \leq \infty$ and $2 \leq u \leq \infty$, where $P_t(\cdot)$ denotes the distribution of X_t and $\int \|h(x)\|^\infty dP_t(x)$ denotes $\sup_{x\in\mathcal{X}} \|h(x)\|$.

(d) $\sup_{t\geq 1} \sup_{\theta\in\Theta} E\|\psi(Z_t,\theta)\|^a < \infty$ and $\sup_{t\geq 1} \sup_{\theta\in\Theta} E\|\Delta_0(X_t)\|^a < \infty$ for $a = \max\{r/(r-1), 2u/(u-1)\}$.

(e) $\mathcal{T} \subset \{\tau : \tau(x) = \Delta(x)'\Omega^{-1}(x)$ for some functions $\Delta(x)$ and $\Omega(x)$ such that $\inf_{x\in\mathcal{X}} \lambda_{\min}(\Omega(x)) \geq \epsilon$ and $\overline{\lim}_{T\to\infty} \frac{1}{T}\Sigma_1^T E\|\Delta(X_t)\|^r \leq C\}$ for some $\epsilon > 0$ and $C < \infty$. In addition, $P(\hat{\tau} \in \mathcal{T}) \to 1$.

(f) $\lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E\Delta_0(X_t)'\Omega_0^{-1}(X_t)\Delta_0(X_t)$ is nonsingular and

$$\left[\lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E\Delta_0(X_t)'\Omega_0^{-1}(X_t)\psi(Z_t, \theta)\right]'\left[\lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E\Delta_0(X_t)'\Omega_0^{-1}(X_t)\Delta_0(X_t)\right]^{-1}$$

$\times \left[\lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E\Delta_0(X_t)'\Omega_0^{-1}(X_t)\psi(Z_t, \theta)\right]$ is bounded away from zero for all $\theta \in \Theta$ outside any given neighborhood of θ_0.

Part (a) of GMM/CMR 1 holds automatically in the case of identically distributed (Z_t, X_t) (given parts (d) and (e)). Also, note that the restrictions on $\mathcal{T}$ given in part (e) are not necessarily exhaustive. In order to obtain the uniform WLLNs of part (b), one imposes additional conditions on $\mathcal{T}$ which depend on the uniform WLLN to be used. (If the results of Section 3 below are to be applied to establish the asymptotic normality of $\hat{\theta}$, then it may be convenient to define $\mathcal{T}$ from the outset to include restrictions that allow one to verify a stochastic equicontinuity condition, Assumption 2(e), given below.)

To see that Assumption GMM/CMR 1 implies Assumption 1, we proceed as follows: Assumption 1(a) holds with

$$m(\theta,\tau) = \lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E\Delta(X_t)'\Omega^{-1}(X_t)\psi(Z_t, \theta) \tag{2.12}$$

by GMM/CMR 1(a) and (b). The first part of Assumption 1(b) holds by GMM/CMR 1(c), (d), and (e), because

$$\begin{aligned}
\sup_{\theta\in\Theta}\|m(\theta,\hat{\tau})-m(\theta, \tau_0)\| &\le \sup_{\theta\in\Theta} \overline{\lim_{N\to\infty}} \frac{1}{N}\Sigma_1^N\int \|\psi(\theta)\|\|(\hat{\Delta}-\Delta_0)'\hat{\Omega}^{-1} + \Delta_0'\hat{\Omega}^{-1}(\Omega_0-\hat{\Omega})\Omega_0^{-1}\|dP_t \\
&\le C_1\left[\sup_{t\ge1}\sup_{\theta\in\Theta} E\|\psi(X_t, \theta)\|^a\right]^{1/a}\overline{\lim_{N\to\infty}} \frac{1}{N}\Sigma_1^N\left[\int\|\hat{\Delta}-\Delta_0\|^r dP_t\right]^{1/r} \\
&\quad + C_2\left[\sup_{t\ge1}\sup_{\theta\in\Theta} E(\|\psi(X_t, \theta)\|\|\Delta_0(X_t)\|)^{u/(u-1)}\right]^{(u-1)/u} \\
&\quad \times \overline{\lim_{N\to\infty}} \frac{1}{N}\Sigma_1^N\left[\int\|\hat{\Omega} - \Omega_0\|^u dP_t\right]^{1/u}
\end{aligned} \tag{2.13}$$

for some constants $C_1, C_2 < \infty$, where $\psi(\theta)$, $\hat{\Delta}$, Δ_0, $\hat{\Omega}$, Ω_0, and P_t abbreviate $\psi(x,\theta)$, $\hat{\Delta}(x)$, $\Delta_0(x)$, $\hat{\Omega}(x)$, $\Omega_0(x)$, and $P_t(x)$ respectively. The second part of

Assumption 1(b) holds by GMM/CMR 1(e). The third part of Assumption 1(b) can be shown to hold using GMM/CMR 1(a), (b), (c), (d), (e), and (f). Assumption 1(c) holds by GMM/CMR 1(d) and (e), since the latter imply that $\sup_{\theta\in\Theta,\,\tau\in\mathcal{T}} \|m(\theta,\tau)\| < \infty$. Assumption 1(d) holds by GMM/CMR 1(f).

2.6. Consistency in the WLS/PLR Example

For the WLS estimator of the PLR model of (2.6), the following assumption implies Assumption 1 with $d(m,\gamma)$ and $m_t(\theta,\tau)$ as defined in (2.11). In consequence, this assumption is sufficient for consistency of $\hat{\theta}$:

ASSUMPTION WLS/PLR 1: (a) $\{[U_t+\tau_{10}(Z_t)-\tau_1(Z_t)+X_t'(\theta-\theta_0)+(\tau_2(Z_t)-\tau_{20}(Z_t))'\theta] \times [X_t-\tau_2(Z_t)]/\tau_3(Z_t) : t \geq 1\}$ satisfies a uniform WLLN over $(\theta,\tau) \in \Theta \times \mathcal{T}$.

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

(c) $E\|X_t\|^4 < \infty$.

(d) Θ is bounded.

(e) $\mathcal{T} \subset \{\tau : \tau = (\tau_1, \tau_2', \tau_3)',\ \inf_{z\in\mathcal{Z}}|\tau_3(z)| \geq \epsilon,\ E\|\tau_j(Z_t)\|^2 \leq C \text{ for } j = 1, 2, 3\}$ for some $\epsilon > 0$ and $C < \infty$. In addition, $P(\hat{\tau} \in \mathcal{T}) \to 1$.

(f) $E(X_t - E(X_t|Z_t))(X_t - E(X_t|Z_t))'/E(U_t^2|Z_t)$ is nonsingular.

To see that Assumption WLS/PLR 1 implies Assumption 1, we proceed as follows: Assumption 1(a) holds by WLS/PLR 1(a) with

$$\begin{aligned} m(\theta,\tau) = {} & E[\tau_{10} - \tau_1][\tau_{20} - \tau_2]/\tau_3 + EX_t'(\theta - \theta_0)[X_t - \tau_2]/\tau_3 \\ & + E(\tau_2-\tau_{20})'\theta[\tau_{20}-\tau_2]/\tau_3 , \end{aligned} \tag{2.14}$$

where τ_{j0} and τ_j abbreviate $\tau_{j0}(Z_t)$ and $\tau_j(Z_t)$ respectively for $j = 1, 2, 3$. The first part of Assumption 1(b) holds by WLS/PLR 1(b), (c), (d), and (e), because

$$\|m(\theta,\tau)-m(\theta,\tau_0)\| \le E|\tau_1-\tau_{10}|\,\|\tau_2-\tau_{20}\|/\tau_3 + \|EX_t'(\theta_0-\theta)\tau_2/\tau_3 \tag{2.15}$$
$$- EX_t'(\theta_0-\theta)\tau_{20}/\tau_{30}\| + E|(\tau_2-\tau_{20})'\theta|\,\|\tau_2-\tau_{20}\|/\tau_3\,,$$

and so,

$$\sup_{\theta\in\Theta}\|m(\theta,\hat\tau)-m(\theta,\tau_0)\| \le \left[\int(\hat\tau_1-\tau_{10})^2dP\cdot\int\|\hat\tau_2-\tau_{20}\|^2dP\right]^{1/2}/\epsilon$$
$$+\sup_{\theta\in\Theta}\|\theta-\theta_0\|\left[\left[E\|X_t\|^2\right]^{1/2}\left[\int\|\hat\tau_2-\tau_{20}\|^2dP\right]^{1/2}/\epsilon \tag{2.16}$$
$$+\left[E\|X_t\|^2\|\tau_{20}(Z_t)\|^2\right]^{1/2}\left[\int(\hat\tau_3-\tau_{30})^2dP\right]^{1/2}/\epsilon^2\right]$$
$$+\sup_{\theta\in\Theta}\|\theta\|\int\|\hat\tau_2-\tau_{20}\|^2dP/\epsilon + o_p(1)\,,$$

where $\hat\tau_j$, τ_{j0}, and dP abbreviate $\hat\tau_j(z)$, $\tau_{j0}(z)$, and dP(z), respectively, for $j = 1, 2, 3$. The second part of Assumption 1(b) holds by WLS/PLR 1(e). The third part of Assumption 1(b) holds trivially, since there is no preliminary estimator $\hat\gamma$. Assumption 1(c) holds by WLS/PLR 1(c), (d), and (e), since the latter imply that $\sup_{\theta\in\Theta,\,\tau\in\mathcal{T}}\|m(\theta,\tau)\| < \infty$. Assumption 1(d) holds by WLS/PLR 1(f).

3. ASYMPTOTIC NORMALITY

We now give sufficient conditions for the asymptotic normality of sequences of R^p–valued MINPIN estimators $\{\hat\theta\}$. Three alternative assumptions are introduced – Assumptions 2, 2*, and 2**. Each is sufficient for asymptotic normality of $\{\hat\theta\}$. Each involves a different tradeoff in the assumptions it imposes. In particular, Assumptions 2 and 2* allow τ to be infinite dimensional whereas Assumption 2** requires it to be finite dimensional. Also, Assumptions 2 and 2** assume $m_t(\theta,\tau)$ is twice differentiable in θ, whereas Assumption 2* only requires $Em_t(\theta,\tau)$ to be twice differentiable in θ. For example, Assumption 2* allows one to consider least absolute deviations (LAD), censored LAD, method of simulated moments, and Huber ψ–function M–estimators. On the other hand, most estimators do satisfy the differentiability condition of Assumptions 2 and 2** and the imposition of this condition allows other conditions in Assumptions 2 and 2** to be weakened.

We note that when the asymptotic normality results based on Assumption 2* are specialized to the case of a finite dimensional nuisance parameter τ, they improve various general results in the nonlinear econometrics literature with regard to the smoothness required of $m_t(\theta,\tau)$ in θ and τ. For example, Bierens (1981), Burguete *et al.* (1982), Domowitz and White (1982), Gallant (1987b), Andrews and Fair (1988), and Gallant and White (1988) all assume $m_t(\theta,\tau)$ (or its equivalent) is twice differentiable. On the other hand, the results of Huber (1967), Pollard (1985), and Pakes and Pollard (1989) allow for non–differentiable $m_t(\theta,\tau)$ functions. The latter results, however, do not accommodate dependent observations or preliminary nuisance parameter estimators whether of finite or infinite dimension.

Before stating Assumptions 2, 2* and 2**, we define the asymptotic covariance matrix of $\{\hat{\theta}\}$, we introduce some notation and definitions used in the assumptions, and we give a brief description of how the property of stochastic equicontinuity is used in establishing the asymptotic normality of $\hat{\theta}$.

3.1. The Asymptotic Covariance Matrix of MINPIN Estimators

The asymptotic covariance matrix V of $\{\hat{\theta}\}$ is defined as follows. Let $d(m,\gamma)$ and $m_t(\theta,\tau)$ be defined such that the dimension v of $\bar{m}_T(\theta,\tau)$ is at least as large as the dimension p of θ. For example, for the WLS estimator of the PLR model, $d(m,\gamma)$ and $m_t(\theta,\tau)$ are as defined in (2.11) rather than as in (2.10). Let

$$S = \lim_{T\to\infty} \mathrm{Var}_P(\sqrt{T}\,\bar{m}_T(\theta_0, \tau_0)),\ M = \lim_{T\to\infty} \tfrac{1}{T}\Sigma_1^T \tfrac{\partial}{\partial\theta'}Em_t(\theta_0, \tau_0)\,,$$

(3.1) $$D = \frac{\partial^2}{\partial m \partial m'}d(m(\theta_0, \tau_0), \gamma_0)\ ,\ \mathcal{J} = M'DM\ ,\ \mathcal{I} = M'DSDM\ ,\ \text{and}$$

$$V = \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}\,,$$

where $\frac{\partial^2}{\partial m \partial m'}d(\cdot,\cdot)$ denotes the matrix of second partial derivatives of $d(\cdot,\cdot)$ with respect to its first argument and $m(\theta_0, \tau_0) = \lim_{T\to\infty} \frac{1}{T}\Sigma_1^T Em_t(\theta_0, \tau_0)$. In the common case where $p = v$ (i.e., the dimension of θ equals the dimension of $\bar{m}_T(\theta,\tau)$), the covariance matrix V simplifies to

(3.2) $$V = M^{-1}S(M^{-1})'\,.$$

For example, for the GMM estimator of the CMR model, we have $p = v$ and the asymptotic covariance matrix of $\hat{\theta}$ is

(3.3)
$$V = M^{-1}SM^{-1}\,,\ \text{where}$$
$$M = \lim_{T\to\infty} \tfrac{1}{T}\Sigma_1^T E\tau_0(X_t)\Delta_0(X_t) = \lim_{T\to\infty} \tfrac{1}{T}\Sigma_1^T E\Delta_0(X_t)'\Omega_0^{-1}(X_t)\Delta_0(X_t)\ \ (= \gamma_0^{-1})\,,$$
$$D = \gamma_0\ \ (= M^{-1})\,,\ \text{and}$$
$$S = \lim_{T\to\infty} \tfrac{1}{T}\Sigma_1^T E\Delta_0(X_s)'\Omega_0^{-1}(X_s)\psi(Z_s, \theta_0)\psi(Z_t, \theta_0)'\Omega_0^{-1}(X_t)\Delta_0(X_t)\,.$$

If W_t is independent across t, S simplifies to M and $V = M^{-1}$. If W_t is iid for $t \geq 1$, $\hat{\theta}$ obtains the semiparametric asymptotic efficiency bound of Chamberlain (1987a). In addition, if $\hat{\gamma}$ is replaced by I_p in the definition of $\hat{\theta}$, the asymptotic covariance matrix of $\{\hat{\theta}\}$ is unchanged. In this case, however, Assumption 6a introduced below does not hold and the likelihood ratio–like test statistic defined below is not necessarily asymptotically chi–square under the null hypothesis.

For the WLS estimator of the PLR model, using the definition of $d(m,\gamma)$ and $m_t(\theta,\tau)$ given in (2.11), we have $p = v$ and the asymptotic covariance matrix of $\hat{\theta}$ is

$V = M^{-1}SM^{-1}$, where

(3.4) $\quad M = -E(X_t - E(X_t|Z_t))(X_t - E(X_t|Z_t))'/E(U_t^2|Z_t)$, $D = I_p$, and

$$S = \lim_{T\to\infty} \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} EU_sU_t(X_s - E(X_s|Z_s))(X_t - E(X_t|Z_t))'/(E(U_s^2|Z_s)E(U_t^2|Z_t)) .$$

If the errors U_s and U_t are uncorrelated conditional on (X_s, Z_s, X_t, Z_t) $\forall s \neq t$ and $E(U_t^2|X_t, Z_t) = E(U_t^2|Z_t)$ a.s. $\forall t$, then $S = -M$ and $V = -M^{-1}$. If the observations are iid and $E(U_t^2|X_t, Z_t) = E(U_t^2|Z_t)$ a.s. $\forall t$ (but the latter is not part of the prior restrictions on the model), then the WLS estimator attains the semiparametric asymptotic efficiency bound given by Chamberlain (1987b) for the PLR model.

3.2. The Definition of Stochastic Equicontinuity

Next, let $\rho_{\mathcal{T}}(\cdot,\cdot)$ and $\rho_{\Theta\times\mathcal{T}}(\cdot,\cdot)$ denote pseudo–metrics on $\mathcal{T}$ and $\Theta \times \mathcal{T}$ respectively. The former is used with Assumption 2 and the latter with Assumption 2*. Examples of such pseudo–metrics are given in (3.23), (3.24), (3.32), and (3.33) below. Convergence in probability of $\hat{\tau}$ to τ_0 and $(\hat{\theta},\hat{\tau})$ to (θ_0, τ_0) means convergence with respect to $\rho_{\mathcal{T}}$ and $\rho_{\Theta\times\mathcal{T}}$ respectively.

We introduce two sequences of stochastic processes, $\{\nu_T(\cdot)\colon T \geq 1\}$ and $\{\nu_T(\cdot,\cdot) : T \geq 1\}$, which are indexed by $\tau \in \mathcal{T}$ and $(\theta,\tau) \in \Theta \times \mathcal{T}$ respectively. The first sequence is used only in Assumption 2 below and the second only in Assumption 2*. Neither is assumed to be measurable. By definition,

$$\nu_T(\tau) = \sqrt{T}(\bar{m}_T(\theta_0, \tau) - \bar{m}_T^*(\theta_0, \tau)) \text{ and}$$

(3.5) $$\nu_T(\theta,\tau) = \sqrt{T}(\bar{m}_T(\theta,\tau) - \bar{m}_T^*(\theta,\tau)) \text{ , where}$$

$$\bar{m}_T^*(\theta,\tau) = \frac{1}{T}\Sigma_1^T Em_t(\theta,\tau) .$$

DEFINITION: $\{\nu_T(\cdot) : T \geq 1\}$ is *stochastically equicontinuous at* τ_0 if: For all $\epsilon > 0$ and $\eta > 0$ there exists $\delta > 0$ such that

$$(3.6) \qquad \overline{\lim_{T\to\infty}} P^*\Big(\sup_{\tau\in\mathcal{T}:\rho_{\mathcal{T}}(\tau,\tau_0)\leq\delta} |\nu_T(\tau) - \nu_T(\tau_0)| > \eta\Big) < \epsilon\,,$$

where P^* denotes outer probability.

(If the rv in parentheses is measurable, then P^* can be replaced by P.) Stochastic equicontinuity of $\{\nu_T(\cdot,\cdot) : T \geq 1\}$ at (θ_0, τ_0) is defined analogously with τ replaced by (θ,τ) and $\rho_{\mathcal{T}}$ replaced by $\rho_{\Theta\times\mathcal{T}}$.

The stochastic equicontinuity condition is discussed in detail in ASEM:II. Primitive sufficient conditions are given there for stochastic equicontinuity in iid, inid, and dnid scenarios. For the time being, note that stochastic equicontinuity of $\{\nu_T(\cdot)\}$ at τ_0 is implied by the weak convergence of $\{\nu_T(\cdot)\}$ to some stochastic process whose sample paths are continuous with probability one (and likewise for $\{\nu_T(\cdot,\cdot)\}$). Thus, if the process $\{\nu_T(\cdot)\}$ or $\{\nu_T(\cdot,\cdot)\}$ satisfies a CLT (with Gaussian limit process that is continuous with probability one), then it is stochastically equicontinuous at τ_0.

3.3. Description of the Method of Proof of Asymptotic Normality

One of the main purposes of this paper is to show how stochastic equicontinuity can be used to establish the $\sqrt{T}$–consistency and asymptotic normality of a large class of semiparametric estimators. We now give a brief heuristic description of how this is done. The assumptions, theorems, and proofs given above and below regarding MINPIN estimators provide a more detailed account.

For the time being, suppose $\hat{\theta}$ is a consistent estimator of $\theta_0 \in R^p$ that satisfies the first order conditions

$$(3.7) \qquad \sqrt{T}\,\bar{m}_T(\hat{\theta},\hat{\tau}) = \underset{\sim}{0}$$

with probability $\to 1$, where $\bar{m}_T(\hat{\theta},\hat{\tau}) = \frac{1}{T}\Sigma_1^T m_t(\hat{\theta},\hat{\tau}) \in R^p$. Let $\hat{\tau}$ be an infinite dimensional parameter estimator that lies in $\mathcal{T}$ with probability $\to 1$ and is consistent for $\tau_0 \in \mathcal{T}$ with respect to the pseudo–metric $\rho_{\mathcal{T}}$. Assume the population first order conditions, $\bar{m}_T^*(\theta_0, \tau_0) = E\bar{m}_T(\theta_0, \tau_0) = \underset{\sim}{0}$, hold.

We consider first the case where $m_t(\theta,\tau)$ is differentiable in θ. If τ was finite dimensional one could establish the asymptotic normality of $\hat{\theta}$ by expanding $\sqrt{T}\,\bar{m}_T(\hat{\theta},\hat{\tau})$ about (θ_0, τ_0) using element by element mean value expansions. Since τ is infinite dimensional, however, mean value expansions in (θ,τ) are not available. In consequence, we expand $\sqrt{T}\,\bar{m}_T(\hat{\theta},\hat{\tau})$ about θ_0 only (using element by element mean value expansions) and use stochastic equicontinuity and an asymptotic orthogonality condition to handle $\hat{\tau}$:

$$(3.8) \qquad o_p(1) = \sqrt{T}\,\bar{m}_T(\hat{\theta},\hat{\tau}) = \sqrt{T}\,\bar{m}_T(\theta_0, \hat{\tau}) + \frac{\partial}{\partial\theta'}\bar{m}_T(\theta^*, \hat{\tau})\sqrt{T}(\hat{\theta}_0 - \theta_0)\,,$$

where θ^* lies on the line segment joining $\hat{\theta}$ and θ_0 (and takes different values in each row of $\frac{\partial}{\partial\theta'}\bar{m}_T(\theta^*, \hat{\tau})$). Under suitable assumptions on $\{m_t(\theta,\tau) : t \geq 1\}$, one can show that $\frac{\partial}{\partial\theta'}\bar{m}_T(\theta^*, \hat{\tau}) \xrightarrow{p} M = \lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E\frac{\partial}{\partial\theta'}m_t(\theta_0, \tau_0)$. Thus, provided M is nonsingular, we have

$$(3.9) \qquad \sqrt{T}(\hat{\theta} - \theta_0) = (M^{-1} + o_p(1))\sqrt{T}\bar{m}_T(\theta_0, \hat{\tau})\,.$$

If $\hat{\tau}$ is replaced by τ_0 in (3.9), the right–hand side of (3.9) is asymptotically normal, say $N(\underset{\sim}{0},S)$, under general conditions by the application of a CLT, since $\sqrt{T}\,\bar{m}_T(\theta_0, \tau_0)$ is a mean zero sample average normalized by $\sqrt{T}$. Hence, if we can show that

$$(3.10) \qquad \sqrt{T}\,\bar{m}_T(\theta_0, \hat{\tau}) - \sqrt{T}\,\bar{m}_T(\theta_0, \tau_0) \xrightarrow{p} \underset{\sim}{0}\,,$$

then we will have established that $\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\underset{\sim}{0}, M^{-1}S(M^{-1})')$. Note that in this case the estimation of τ_0 by $\hat{\tau}$ does not affect the asymptotic distribution of $\hat{\theta}$.

Stochastic equicontinuity is useful in establishing (3.10). In particular, stochastic equicontinuity of $\nu_T(\tau) = \sqrt{T}(\bar{m}_T(\theta_0, \tau) - \bar{m}_T^*(\theta_0, \tau))$ (indexed by $\tau \in \mathcal{T}$) at τ_0, consistency of $\hat{\tau}$ for τ_0 with respect to the pseudo–metric $\rho_{\mathcal{T}}$, and $P(\hat{\tau} \in \mathcal{T}) \to 1$ yield

$$(3.11) \qquad \nu_T(\hat{\tau}) - \nu_T(\tau_0) \xrightarrow{p} \underset{\sim}{0} \,.$$

This follows because given any $\eta, \epsilon > 0$ there exists a $\delta > 0$ such that

$$(3.12) \qquad \begin{aligned} &\overline{\lim_{T\to\infty}} P(|\nu_T(\hat{\tau}) - \nu_T(\tau_0)| > \eta) \\ &\leq \overline{\lim_{T\to\infty}} P(|\nu_T(\hat{\tau}) - \nu_T(\tau_0)| > \eta, \hat{\tau} \in \mathcal{T}, \rho_{\mathcal{T}}(\hat{\tau},\tau_0) \leq \delta) + \overline{\lim_{T\to\infty}} P(\hat{\tau} \notin \mathcal{T} \text{ or } \rho_{\mathcal{T}}(\hat{\tau},\tau_0) > \delta) \\ &\leq \overline{\lim_{T\to\infty}} P^*\Big(\sup_{\tau\in\mathcal{T}:\rho_{\mathcal{T}}(\tau,\tau_0)\leq\delta} |\nu_T(\tau) - \nu_T(\tau_0)| > \eta\Big) \\ &< \epsilon \,. \end{aligned}$$

Since

$$(3.13) \qquad \sqrt{T}\,\bar{m}_T(\theta_0, \hat{\tau}) - \sqrt{T}\,\bar{m}_T(\theta_0, \tau_0) = \nu_T(\hat{\tau}) - \nu_T(\tau_0) - \sqrt{T}\,\bar{m}_T^*(\theta_0, \hat{\tau}) \,,$$

equation (3.10) now holds if and only if

$$(3.14) \qquad \sqrt{T}\,\bar{m}_T^*(\theta_0, \hat{\tau}) \xrightarrow{p} \underset{\sim}{0} \,.$$

The latter is an asymptotic orthogonality condition between $\hat{\theta}$ and $\hat{\tau}$ that is analogous to the block diagonality of the information matrix between θ_0 and τ_0 in the case of maximum likelihood estimation with finite dimensional τ. This condition is usually satisfied by adaptive estimators of adaptive models, but is also satisfied by numerous semiparametric estimators of non–adaptive models (for suitable estimators $\hat{\tau}$). See the discussion of this condition given below.

In sum, given stochastic equicontinuity of $\nu_T(\tau)$, consistency of $\hat{\tau}$ for τ_0, the asymptotic orthogonality condition (3.14), and the other conditions referred to above, equation (3.10) holds and we obtain the asymptotic normality of $\hat{\theta}$, as desired.

Next, we briefly consider the case where $m_t(\theta,\tau)$ is not differentiable in θ, but $Em_t(\theta,\tau)$ is. This occurs with LAD estimators, Huber M–estimators, and weighted censored LAD estimators (under appropriate conditions) among other parametric and semiparametric estimators. Let $\hat{\theta}$, $\hat{\tau}$, etc. be as described in the paragraph containing (3.7) except that consistency of $\hat{\tau}$ is replaced by consistency of $(\hat{\theta},\hat{\tau})$ for (θ_0, τ_0) with respect to some pseudo–metric $\rho_{\Theta \times \mathcal{T}}$ on $\Theta \times \mathcal{T}$.

Since $\sqrt{T}\,\bar{m}_T(\theta,\hat{\tau})$ is not differentiable in θ, we cannot carry out mean value expansions of $\sqrt{T}\,\bar{m}_T(\hat{\theta},\hat{\tau})$ as in (3.8). Instead, we expand $\sqrt{T}\,\bar{m}_T^*(\theta_0, \hat{\tau})$ about $\hat{\theta}$ using element by element mean value expansions:

$$(3.15) \qquad o_p(1) = \sqrt{T}\,\bar{m}_T^*(\theta_0, \hat{\tau}) = \sqrt{T}\,\bar{m}_T^*(\hat{\theta},\hat{\tau}) - \frac{\partial}{\partial\theta'}\bar{m}_T^*(\theta^*, \hat{\tau})\sqrt{T}(\hat{\theta} - \theta_0)\,,$$

where θ^* is as above. The first equality of (3.15) holds by the asymptotic orthogonality condition (3.14). Under suitable assumptions on $\{m_t(\theta,\tau) : t \geq 1\}$, $\frac{\partial}{\partial\theta'}\bar{m}_T^*(\theta^*, \hat{\tau}) \xrightarrow{p} M = \lim_{T\to\infty} \frac{1}{T}\Sigma_1^T \frac{\partial}{\partial\theta'}Em_t(\theta_0, \tau_0)$. Thus, provided M is nonsingular, we have

$$(3.16) \qquad \sqrt{T}(\hat{\theta} - \theta_0) = (M^{-1} + o_p(1))\sqrt{T}\,\bar{m}_T^*(\hat{\theta},\hat{\tau})\,.$$

Thus, we obtain the same asymptotic distribution for $\sqrt{T}(\hat{\theta} - \theta_0)$ as was obtained above if we can show that

$$(3.17) \qquad \sqrt{T}\,\bar{m}_T^*(\hat{\theta},\hat{\tau}) + \sqrt{T}\,\bar{m}_T(\theta_0, \tau_0) \xrightarrow{p} \underset{\sim}{0}\,,$$

since $-\sqrt{T}\,\bar{m}_T(\theta_0, \tau_0)$ is asymptotically normal under general conditions by a CLT. Let $\nu_T(\theta,\tau) = \sqrt{T}(\bar{m}_T(\theta,\tau) - \bar{m}_T^*(\theta,\tau))$. Using (3.7) and $\sqrt{T}\,\bar{m}_T^*(\theta_0, \tau_0) = \underset{\sim}{0}$, we have

$$(3.18) \qquad \sqrt{T}\,\bar{m}_T^*(\hat{\theta},\hat{\tau}) + \sqrt{T}\,\bar{m}_T(\theta_0, \tau_0) = -\nu_T(\hat{\theta},\hat{\tau}) + \nu_T(\theta_0, \tau_0)$$

with probability $\to 1$. In addition, in analogy to (3.12), stochastic equicontinuity of $\nu_T(\cdot,\cdot)$ (indexed by $(\theta,\tau) \in \Theta \times \mathcal{T}$) at (θ_0, τ_0), consistency of $(\hat{\theta},\hat{\tau})$ for (θ_0, τ_0) with respect to the pseudo–metric $\rho_{\Theta\times\mathcal{T}}$, and $P((\hat{\theta},\hat{\tau}) \in \Theta \times \mathcal{T}) \to 1$ yield

$$(3.19) \qquad \nu_T(\hat{\theta},\hat{\tau}) - \nu_T(\theta_0, \tau_0) \xrightarrow{p} \underset{\sim}{0}\,.$$

In consequence, $\sqrt{T}(\hat{\theta} - \theta_0)$ is asymptotically normal, as desired.

3.4. Asymptotic Normality of MINPIN Estimators

We now return to the discussion of MINPIN estimators and state the Assumptions 2, 2*, and 2** that are sufficient for asymptotic normality of such estimators. The discussion above of the use of stochastic equicontinuity should serve to explain its appearance in Assumptions 2 and 2*. Let $\hat\theta$, $\hat\tau$, etc. be as defined in the definition of MINPIN estimators rather than as defined in Section 3.3. Let Θ_0 be a subset of Θ ($\subset R^p$) that contains a neighborhood of θ_0. Define

$$\text{(3.20)}\qquad \begin{aligned} B_t &= \sup_{\theta\in\Theta_0,\, \tau\in\mathcal{T}} \left\| \frac{\partial^2}{\partial\theta\partial\theta'} m_t(\theta,\tau) \right\| \quad \text{and} \\ B_t^* &= \sup_{\theta\in\Theta_0,\, \tau\in\mathcal{T}} \left\| \frac{\partial^2}{\partial\theta\partial\theta'} E m_t(\theta,\tau) \right\| . \end{aligned}$$

The rv B_t is assumed to exist only in Assumptions 2 and 2** below.

We now state Assumptions 2, 2*, and 2** and the main result of this section:

ASSUMPTION 2: (a) $\hat\theta \xrightarrow{p} \theta_0 \in \Theta \subset R^p$ *and* θ_0 *is in the interior of* Θ.

(b) $P(\hat\tau \in \mathcal{T}) \to 1$, $\hat\tau \xrightarrow{p} \tau_0$, *and* $\hat\gamma \xrightarrow{p} \gamma_0$ *for some* $\tau_0 \in \mathcal{T}$ *and* $\gamma_0 \in \Gamma$.

(c) $\sqrt{T}\, \frac{\partial}{\partial m} d(\bar m_T^*(\theta_0, \hat\tau), \hat\gamma) \xrightarrow{p} \underset{\sim}{0}$.

(d) $\nu_T(\tau_0) \xrightarrow{d} N(\underset{\sim}{0}, S)$.

(e) $\{\nu_T(\cdot)\}$ *is stochastically equicontinuous at* τ_0.

(f) $\frac{\partial}{\partial m} d(m,\gamma)$ *and* $\frac{\partial^2}{\partial m \partial m'} d(m,\gamma)$ *exist for* $(m,\gamma) \in \mathcal{M}_0 \times \Gamma_0$ *and are continuous at* $(m,\gamma) = (m(\theta_0, \tau_0), \gamma_0)$, *where* $\mathcal{M}_0$ *and* Γ_0 *are subsets of* R^v *and* Γ *that contain neighborhoods of* $m(\theta_0, \tau_0)$ *and* γ_0 *respectively (using the Euclidean norm on* R^v *and the pseudo–metric on* Γ *).*

(g) $m_t(\theta,\tau)$ *is twice continuously differentiable in* θ *on* Θ_0, $\forall \tau \in \mathcal{T}$, $\forall t \geq 1$, $\forall \omega \in \Omega$. $\{m_t(\theta,\tau)\}$ and $\left\{\frac{\partial}{\partial\theta'} m_t(\theta,\tau)\right\}$ *satisfy uniform WLLNs over* $\Theta_0 \times \mathcal{T}$. $m(\theta,\tau) = \lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E m_t(\theta,\tau)$ *and* $M(\theta,\tau) = \lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E \frac{\partial}{\partial\theta'} m_t(\theta,\tau)$ *each exist uniformly over*

$\Theta_0 \times \mathcal{T}$ *and are continuous at* (θ_0, τ_0) *with respect to some pseudo–metric on* $\Theta_0 \times \mathcal{T}$ *for which* $(\hat{\theta},\hat{\tau}) \xrightarrow{p} (\theta_0, \tau_0)$. $\overline{\lim}_{T\to\infty} \frac{1}{T}\Sigma_1^T EB_t < \infty$.

(h) $M'DM$ *is nonsingular.*

ASSUMPTION 2*: (a) *Assumption* 2(a) *holds and* $\sqrt{T}\left[\frac{\partial}{\partial\theta'}\bar{m}_T^*(\hat{\theta},\hat{\tau})\right]'\frac{\partial}{\partial m}d(\bar{m}_T(\hat{\theta},\hat{\tau}), \hat{\gamma}) = o_p(1)$.

(b) $P(\hat{\tau} \in \mathcal{T}) \to 1$, $(\hat{\theta},\hat{\tau}) \xrightarrow{p} (\theta_0, \tau_0)$, and $\hat{\gamma} \xrightarrow{p} \gamma_0$ *for some* $\tau_0 \in \mathcal{T}$ *and* $\gamma_0 \in \Gamma$.

(c) *Assumption* 2(c) *holds.*

(d) *Assumption* 2(d) *holds.*

(e) $\{\nu_T(\cdot,\cdot)\}$ *is stochastically equicontinuous at* (θ_0, τ_0).

(f) *Assumption* 2(f) *holds.*

(g) $Em_t(\theta,\tau)$ *is twice continuously differentiable in* θ *on* Θ_0, $\forall\tau \in \mathcal{T}$, $\forall t \geq 1$. $m(\theta,\tau) = \lim_{T\to\infty} \frac{1}{T}\Sigma_1^T Em_t(\theta,\tau)$ *and* $M(\theta,\tau) = \lim_{T\to\infty} \frac{1}{T}\Sigma_1^T \frac{\partial}{\partial\theta'}Em_t(\theta,\tau)$ *each exist uniformly over* $\Theta_0 \times \mathcal{T}$ *and are continuous at* (θ_0, τ_0) *with respect to some pseudo–metric on* $\Theta_0 \times \mathcal{T}$ *for which* $(\hat{\theta},\hat{\tau}) \xrightarrow{p} (\theta_0, \tau_0)$. $\overline{\lim}_{T\to\infty} \frac{1}{T}\Sigma_1^T B_t^* < \infty$.

(h) *Assumption* 2(h) *holds.*

ASSUMPTION 2**: *Assumption* 2 *holds with* $\mathcal{T} \subset R^u$ *for some* $u < \infty$, *with* $\rho_{\mathcal{T}}$ *given by the Euclidean metric on* $\mathcal{T}$, *and with Assumption* 2(e) *replaced by*

2**(e) $\sqrt{T}(\hat{\tau} - \tau_0) = O_p(1)$, $\frac{\partial}{\partial\tau'}m_t(\theta_0, \tau)$ *exists* $\forall\tau \in \mathcal{T}$, $\forall t \geq 1$, $\forall\omega \in \Omega$, $\left\{\frac{\partial}{\partial\tau'}m_t(\theta_0,\tau)\right\}$ *satisfies a uniform WLLN over* $\tau \in \mathcal{T}$, *and* $E\sup_{\tau\in\mathcal{T}}\left\|\frac{\partial}{\partial\tau'}m_t(\theta_0,\tau)\right\| < \infty$ $\forall t \geq 1$.

THEOREM I.2: *Under Assumption 2, 2*, or 2**, every sequence of* MINPIN *estimators* $\{\hat{\theta}\}$ *satisfies*

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\underset{\sim}{0},V).$$

COMMENT: We note that a useful feature of the method used here for establishing the asymptotic normality of semiparametric estimators is that the assumptions on $(\hat{\tau},\hat{\gamma})$ and on the random criterion function $d(\bar{m}_T(\theta,\tau), \gamma)$ are split apart. This simplifies their verification and allows for greater generality in some dimensions. Assumption 2(b) only involves $(\hat{\tau},\hat{\gamma})$ and Assumption 2(c) only involves $(\hat{\tau},\hat{\gamma})$ and the *non-random function* $d(\bar{m}_T^*(\theta_0, \tau), \gamma)$. Thus, the fact that $(\hat{\tau},\hat{\gamma})$ and the random function $d(\bar{m}_T(\theta_0, \tau), \gamma)$ are defined using the same underlying random variables does not present a problem. In particular, sample splitting is not needed to establish asymptotic normality of $\hat{\theta}$.

3.5. Discussion of Assumption 2

We now discuss Assumption 2. Assumption 2(a) can be established by Theorem I.1 or some other consistency proof. Assumptions 2(b), (c), and (e) are key assumptions – they are discussed below. Assumption 2(d) can be verified using a CLT for a sequence (or triangular array) of rv's. In the independent case, the Lindeberg–Lévy CLT can be used. In the dependent case, numerous CLTs are available that differ according to the dependence, moment, and identical distributions assumptions that they impose. For example, see McLeish (1975b, 1977), Hall and Heyde (1980, Chs. 3–5), Herrndorf (1984), Gallant (1987b), and Wooldridge and White (1988a, b).

Assumption 2(f) usually is not restrictive and is easy to verify. Assumption 2(g) requires $m_t(\theta,\tau)$ to be twice differentiable in θ. This assumption can be avoided, if necessary, by using Assumption 2*. Assumption 2(g) requires certain uniform WLLNs to hold. They can be verified in the same manner as Assumption 1(a) above. Assumption 2(g) also requires continuity of $m(\theta,\tau)$ and $M(\theta,\tau)$ with respect to some pseudo–metrics on $\Theta_0 \times \mathcal{T}$ for which $(\hat{\theta},\hat{\tau}) \xrightarrow{p} (\theta_0, \tau_0)$. The most convenient choices of pseudo–metrics are

$$(3.21) \qquad \rho((\theta_1, \tau_1), (\theta_2, \tau_2)) = \overline{\lim_{N\to\infty}} \tfrac{1}{N}\Sigma_1^N E\|m_t(\theta_1, \tau_1) - m_t(\theta_2, \tau_2)\|$$

for establishing continuity of $m(\theta,\tau)$ and

$$(3.22) \qquad \rho((\theta_1, \tau_1), (\theta_2, \tau_2)) = \overline{\lim_{N\to\infty}} \tfrac{1}{N}\Sigma_1^N E\left\|\frac{\partial}{\partial\theta'} m_t(\theta_1, \tau_1) - \frac{\partial}{\partial\theta'} m_t(\theta_2, \tau_2)\right\|$$

for establishing continuity of $M(\theta,\tau)$.[8] With these choices, continuity of $m(\theta,\tau)$ and $M(\theta,\tau)$ automatically holds, so it suffices to verify that $\rho((\hat{\theta},\hat{\tau}),(\theta_0,\tau_0)) \xrightarrow{p} 0$ for each choice of $\rho(\cdot,\cdot)$. As the examples below show, the latter usually holds under similar assumptions to those used to verify the condition of Assumption 2(b) that $\rho_{\mathcal{T}}(\hat{\tau},\tau_0) \xrightarrow{p} 0$. Assumption 2(h) is a standard condition that is often closely related to the identification of θ_0. For example, it reduces to nonsingularity of the information matrix in iid ML contexts.

The stochastic equicontinuity assumption, Assumption 2(e), is basic to the approach taken here. It can be verified using the stochastic equicontinuity results given in ASEM:II or by using other results in the literature.

In order to obtain stochastic equicontinuity, the index set $\mathcal{T}$ needs to satisfy some conditions. This creates a tension between Assumption 2(e) and the first part of Assumption 2(b), since the more restricted is $\mathcal{T}$, the more difficult it is to show that $P(\hat{\tau} \in \mathcal{T}) \to 1$. For example, if $\mathcal{T}$ is an infinite dimensional class of functions, the stochastic equicontinuity results of ASEM:II require the functions in $\mathcal{T}$ to satisfy certain smoothness conditions. When $\mathcal{T}$ is defined as such, one has to show that the nonparametric function estimator $\hat{\tau}$ also satisfies these smoothness conditions with probability $\to 1$ to verify the first part of Assumption 2(b).

For example, if τ_0 is a function of x for $x \in \mathcal{X}$, τ_0 satisfies the smoothness conditions of ASEM:II, and $\hat{\tau}$ and a suitable number of its derivatives converge in probability uniformly over $x \in \mathcal{X}$ to τ_0 and its corresponding derivatives, then the first part of Assumption 2(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 (at the expense, of course, of reducing asymptotic efficiency by some $\epsilon > 0$). 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, one can exploit existing consistency results and proofs for nonparametric estimators of regression and density functions and their derivatives, when establishing the first part of Assumption 2(b).

We note that the above mentioned smoothness requirements on the infinite dimensional parameter τ_0 and the estimator $\hat{\tau}$ of it are sometimes stronger than necessary for asymptotic normality of $\hat{\theta}$. Whether they are depends primarily on Assumption 2(c). As discussed below, Assumption 2(c) requires L^Q–consistency of $\hat{\tau}$ for τ_0 at rate $T^{1/4}$ for some $1 \leq Q < \infty$ in many examples, including the WLS/PLR example.[9] If τ_0 is a nonparametric regression function, then the smoothness conditions on τ_0 required for the existence of an estimator that is L^Q–consistent at rate $T^{1/4}$ (see Stone (1982)) are essentially the same as those imposed in ASEM:II for Assumption 2(e). (More specifically, the condition is that $q > k_a/2$, where q is a measure of smoothness based on the number of derivatives that exist and k_a is the dimension of the regressor variable, see ASEM:II.)

On the other hand, the orthogonality condition 2(c) is satisfied trivially in a number of examples, see the discussion below. For these examples, the smoothness conditions of ASEM:II that are used to verify Assumption 2(e) when τ_0 is infinite dimensional are stronger than necessary. For example, Robinson's (1987) results for the WLS estimator of a linear regression model with heteroskedasticity of unknown form and Newey's (1987) results for the one–step GMM estimator of the CMR model establish asymptotic normality of $\hat{\theta}$ without any smoothness conditions at all on τ_0.

Next we discuss the pseudo–metric $\rho_{\mathcal{T}}$ on $\mathcal{T}$. As with the choice of $\mathcal{T}$, there is a tradeoff between Assumptions 2(b) and (e) with regard to the choice of $\rho_{\mathcal{T}}$. The stronger is the pseudo–metric, the easier it is to verify Assumption 2(e), but the more difficult it is to verify the condition of Assumption 2(b) that $\rho_{\mathcal{T}}(\hat{\tau}, \tau_0) \xrightarrow{p} 0$. It is this tradeoff and the availability of stochastic equicontinuity results for different pseudo–metrics that determine the most appropriate choice of pseudo–metric. For the stochastic equicontinuity results of ASEM:II, several different pseudo–metrics are considered. One is

$$(3.23) \qquad \rho_{\mathcal{T}}(\tau_1, \tau_2) = \sup_{N \geq 1} \left[\frac{1}{N}\Sigma_1^N E\|m_t(\theta_0, \tau_1) - m_t(\theta_0, \tau_2)\|^2\right]^{1/2} .$$

A second is

$$(3.24) \qquad \rho_{\mathcal{T}}(\tau_1, \tau_2) = \left[\int_{\mathcal{W}} \|m(w, \theta_0, \tau_1) - m(w, \theta_0, \tau_2)\|^2 dw\right]^{1/2} ,$$

which applies when $m_{Tt}(\cdot,\cdot,\cdot)$ does not depend on T or t and W_{Tt} takes values in a bounded set $\mathcal{W}$. A variation of (3.24) is also considered in which some elements of W_t may be unbounded while others must be bounded. Stochastic equicontinuity results using the pseudo–metric of (3.23) are given in ASEM:II for the case of independent or m–dependent rv's. For rv's that exhibit more general forms of temporal dependence, such as strong mixing dependence or near–epoch dependence, one of the other pseudo–metrics must be used.

Consistency of $\hat{\tau}$ for τ_0 with respect to a pseudo–metric such as that of (3.23) or (3.24) (as required by Assumption 2(b)) can usually be reduced to L^Q–consistency of $\hat{\tau}$ for some $2 \leq Q \leq \infty$ when $\mathcal{T}$ is an infinite dimensional class of functions, plus some moment conditions on certain functions of W_{Tt} and sometimes some uniform boundedness condition on τ_0 and $\hat{\tau}$ that must hold with probability $\to 1$. For example, see the following discussion of the GMM estimator of the CMR model and of the WLS estimator of the PLR model. As with Assumption 1(b), one can exploit existing consistency results and proofs for nonparametric estimators when verifying such conditions on $\hat{\tau}$.

Assumption 2(c) is a key assumption. It is an asymptotic orthogonality condition between the estimators $\hat{\theta}$ and $\hat{\tau}$. It is needed to show that preliminary estimation of τ_0 does not affect the asymptotic distribution of $\hat{\theta}$. In almost all examples, $d(m,\gamma) = m'm/2$ or $m'\gamma m/2$ and $\hat{\gamma} = O_p(1)$. In consequence, Assumption 2(c) usually reduces to the requirement that

$$(3.25) \qquad \sqrt{T}\,\bar{m}_T^*(\theta_0, \hat{\tau}) \xrightarrow{p} \underset{\sim}{0}\,.$$

Note that $\bar{m}_T^*(\theta_0, \tau)$ evaluated at $\tau = \tau_0$ generally equals zero, because $\bar{m}_T^*(\theta_0, \tau_0) = \underset{\sim}{0}$ are the population first order conditions for the estimator $\hat{\theta}$ and θ_0 is an interior point. Thus, condition (3.25) requires that the replacement of τ_0 by $\hat{\tau}$ in $\bar{m}_T^*(\theta_0, \tau)$ has an effect that is at most $o_p(T^{-1/2})$.

Condition (3.25) (and hence, Assumption 2(c)) is trivial to verify whenever

$$(3.26) \qquad \bar{m}_T^*(\theta_0, \tau) = \underset{\sim}{0} \ \ \forall\tau \text{ in some neighborhood of } \tau_0$$

for all T sufficiently large. The reason is, in this case,

$$\sqrt{T}\,\bar{m}_T^*(\theta_0,\hat{\tau}) = \sqrt{T}\bar{m}_T^*(\theta_0,\tau)1(\rho_{\mathcal{T}}(\tau,\tau_0) < \epsilon)\Big|_{\tau=\hat{\tau}} + \sqrt{T}\bar{m}_T^*(\theta_0,\hat{\tau})1(\rho_{\mathcal{T}}(\hat{\tau},\tau_0) \geq \epsilon)$$

$$(3.27) \qquad = \underset{\sim}{0} + o_p(1)$$

for some $\epsilon > 0$, using (3.26) and Assumption 2(b). In many cases, condition (3.26) is equivalent to the condition: The derivative of $\bar{m}_T^*(\theta_0, \tau(\zeta))$ with respect to ζ equals a zero matrix *for all* ζ *in a neighborhood of* ζ_0 for T large, for any finite dimensional parametrization $\{\tau(\zeta) : \zeta \in Z \subset R^{\ell}\}$ of a subset of $\mathcal{T}$ for which (i) $\tau_0 = \tau(\zeta_0)$ for some $\zeta_0 \in Z$ and (ii) $\bar{m}_T^*(\theta_0, \tau(\zeta))$ is differentiable in ζ in a neighborhood of ζ_0.

It is easy to see that (3.26) holds for the GMM estimator of the CMR model. It also holds for the weighted censored LAD estimator of the censored regression model, see Example 6.2 below. Neither of these models is an adaptive model. Condition (3.26) also holds for most adaptive estimators of adaptive models, such as Bickel's (1982) and Manski's (1984) adaptive estimators of linear and nonlinear regression models with errors of unknown distribution, see Example 6.5 below. Note that when (3.26) holds, Assumption 2(c) places no additional restrictions on $\hat{\tau}$ beyond those of Assumption 2(b).

On the other hand, Assumption 2(c) and (3.25) do not require (3.26) to hold. In fact, there are numerous examples where Assumption 2(c) and (3.25) hold (for suitable $\hat{\tau}$) but (3.26) does not hold. The WLS estimator of the PLR model is one example of this.

Section 6 discusses several other examples. In such cases, the estimator $\hat{\tau}$ usually needs to converge to τ_0 at a particular rate, with the rate often being $T^{1/4}$. Note that nonparametric estimators of regression and density functions exist that are L^Q or uniformly consistent at rate $T^{1/4}$ provided the function satisfies certain smoothness conditions among other assumptions, see Stone (1982).

To see why a convergence rate on $\hat{\tau}$ may be needed to verify Assumption 2(c), to see why the required rate is often $T^{1/4}$, and to see what Assumption 2(c) reduces to in the case of finite dimensional τ, we consider the case where τ is finite dimensional and $\bar{m}_T^*(\theta_0, \tau)$ is twice differentiable in τ. In this case, a two term Taylor expansion of the j-th element of $\sqrt{T}\,\bar{m}_T^*(\theta_0, \hat{\tau})$ about τ_0 yields

$$\begin{aligned}\sqrt{T}\,\bar{m}_{Tj}^*(\theta_0, \hat{\tau}) &= \sqrt{T}\,\bar{m}_{Tj}^*(\theta_0, \tau_0) + \frac{\partial}{\partial \tau'}\bar{m}_{Tj}^*(\theta_0, \tau_0)\sqrt{T}(\hat{\tau} - \tau_0) \\ &\quad + T^{1/4}(\hat{\tau} - \tau_0)'\frac{\partial^2}{\partial \tau \partial \tau'}\bar{m}_{Tj}^*(\theta_0, \bar{\tau}_j)T^{1/4}(\hat{\tau} - \tau_0)/2\end{aligned} \tag{3.28}$$

for $j = 1, \ldots, v$, where $\bar{\tau}_j$ lies between $\hat{\tau}$ and τ_0.

The right–hand side of (3.28) is $o_p(1)$, as required by Assumption 2(c), if (i) $\bar{m}_T^*(\theta_0, \tau_0) = \underset{\sim}{0}$, (ii) $\frac{\partial}{\partial \tau'}\bar{m}_T^*(\theta_0, \tau_0) = \underset{\sim}{0}$, (iii) $T^{1/4}(\hat{\tau} - \tau_0) = o_p(1)$, and (iv) $\frac{\partial^2}{\partial \tau \partial \tau'}\bar{m}_{Tj}^*(\theta_0, \bar{\tau}_j) = O_p(1)\ \ \forall j$. As mentioned above, (i) holds quite generally whether or not $\hat{\theta}$ and $\hat{\tau}$ are asymptotically orthogonal. Condition (ii) is the requisite asymptotic orthogonality condition between $\hat{\theta}$ and $\hat{\tau}$. For a maximum likelihood estimator, it reduces to block diagonality of the information matrix between the parameters θ_0 and τ_0. Condition (iii) is the rate of convergence condition on $\hat{\tau}$ referred to above. It is satisfied by most consistent parametric estimators, since such estimators are usually $\sqrt{T}$–consistent.

When τ is infinite dimensional, one cannot carry out the expansion given in (3.28). Nevertheless, it is usually straightforward to obtain conditions on $\hat{\tau}$ that are sufficient for (3.25) and Assumption 2(c) by using standard inequalities (such as the Cauchy–Schwartz and Hölder's inequalities) to obtain a suitable upper bound on $\|\sqrt{T}\,\bar{m}_T^*(\theta_0, \hat{\tau})\|$. This is best seen by looking at the examples given below and in Section 6. In those cases where (3.26) does not hold, but where it is possible for Assumption 2(c) to hold, such upper bounds usually involve terms such as $\sqrt{T}\int\|\hat{\tau}(w) - \tau_0(w)\|^2 dP(w)$. Hence, L^2–consistency of $\hat{\tau}$ at rate $T^{1/4}$ is a common requirement on $\hat{\tau}$ for Assumption 2(c).

Next, we discuss the relationship between Assumption 2(c) and the property of adaptation of an estimator. When Assumption 2(c) holds, a MINPIN estimator $\hat{\theta}$ has the same asymptotic distribution as the estimator that minimizes the same criterion function as $\hat{\theta}$ but with $\hat{\tau}$ replaced by τ_0. If, in addition, the latter estimator is an asymptotically efficient estimator of θ_0 for the case where θ_0 is the only unknown parameter in the problem, then $\hat{\theta}$ is an adaptive estimator.

The latter condition only holds in special cases, so it is not the case that Assumption 2(c) only holds for adaptive estimators. For example, with the WLS estimator in the PLR model, Assumption 2(c) holds for suitable $\hat{\tau}$, as shown below. Nevertheless, if one knows the function $g(\cdot)$ in (2.6), then one can form an asymptotically more efficient estimator than the WLS estimator, and hence, the WLS estimator is not adaptive. In particular, if $g(\cdot)$ is known, one can move $g(Z_t)$ to the left–hand side of (2.6), define the dependent variable to be $Y_t - g(Z_t)$, and estimate θ_0 via a WLS regression of $Y_t - g(Z_t)$ on X_t. On the other hand, if one tried to carried out the latter estimation procedure with $g(\cdot)$ replaced by an estimator $\hat{g}(\cdot)$, then one would find that the estimation of $g(\cdot)$ affects the limit distribution of the estimator and Assumption 2(c) fails for this estimator. This must be the case, since the PLR model is not an adaptive model. (That is, it is a model for which no adaptive estimator exists.)

To conclude, Assumption 2(c) implies that for *the criterion function at hand* the preliminary estimation of τ_0 does not affect the asymptotic distribution of the resulting estimator. It does not imply the same result for *any* criterion function that might be used to estimate θ_0. In consequence, Assumption 2(c) is a much weaker requirement than is adaptability. In addition, as noted above, adaptability is not a necessary condition for Assumption 2(c) to hold trivially via (3.26).

We now consider the case where Assumption 2(c) fails, and hence, the estimation of τ_0 has an effect on the asymptotic distribution of the MINPIN estimator, also see the discussion of Newey (1990b, Sec. 4.4). In such cases, the estimator in question may or may not be asymptotically normal. Some examples of MINPIN estimators where Assumption 2(c) fails but the estimator is still asymptotically normal include Powell, Stock, and Stoker's (1989) and Andrews' (1991a) estimators of weighted average derivatives and index regression models, Han's (1987) maximum rank correlation estimator of generalized regression models, Cox's (1975) partial likelihood estimator of the proportional hazard model, and Horowitz's (1988) M–estimators of the censored regression model. To fit such estimators into the framework developed here, one would have to present conditions under which $\sqrt{T}\,\bar{m}_T^*(\theta_0, \hat{\tau})$ is asymptotically normal jointly with $\nu_T(\tau_0)$ (say, $\left[\sqrt{T}\,\bar{m}_T^*(\theta_0, \hat{\tau})', \nu_T(\tau_0)'\right]' \xrightarrow{d} N\left[\underset{\sim}{0}, \begin{bmatrix} A & A_{12} \\ A_{12}' & S \end{bmatrix}\right]$) rather than $o_p(1)$, when $d(m,\gamma) = m'\gamma m/2$. For example, Newey (1989a, Sec. 6) presents a method of establishing the asymptotic normality of $\sqrt{T}\,\bar{m}_T^*(\theta_0, \hat{\tau})$ for some cases, using the stochastic equicontinuity results of ASEM:II. His method can be extended straightforwardly to obtain the joint normality of $\sqrt{T}\,\bar{m}_T^*(\theta_0, \hat{\tau})$ and $\nu_T(\tau_0)$ for these cases. Under conditions for the joint normality of $\sqrt{T}\,\bar{m}_T^*(\theta_0, \hat{\tau})$ and $\nu_T(\tau_0)$, the limit distribution of $\hat{\theta}$ is given by

$$\text{(3.29)} \qquad \sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\underset{\sim}{0}, V + (M'DM)^{-1}M'D(A + A_{12} + A_{12}')DM(M'DM)^{-1}) .$$

We leave the development of results along these lines to future research.

3.6. Asymptotic Normality in the GMM/CMR Example

Here, we discuss the verification of Assumption 2 for the GMM estimator of the CMR model. The following assumption in conjunction with Assumption GMM/CMR 1 is sufficient for Assumption 2 (with ρ_T given by (3.23)), and hence, for the asymptotic normality of the GMM estimator $\hat{\theta}$:

ASSUMPTION GMM/CMR 2: (a) θ_0 is in the interior of Θ_0.

(b) $\lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} E\Delta_0(X_s)'\Omega_0^{-1}(X_s)\psi(Z_s, \theta_0)\psi(Z_t, \theta_0)'\Omega_0^{-1}(X_t)\Delta_0(X_t)$ exists, $\sup_{t\geq 1} E\|\psi(Z_t, \theta_0)\|^{2+\delta} < \infty$ for some $\delta > 0$, and $\{(Z_t, X_t) : t \geq 1\}$ is a sequence of independent rv's or strong mixing rv's with mixing numbers that satisfy $\sum_{s=1}^{\infty} \alpha(s)^{\delta/(2+\delta)} < \infty$ for δ as above.

(c) $\left\{\frac{1}{\sqrt{T}} \Sigma_1^T \tau(X_t)\psi(Z_t, \theta_0) : T \geq 1\right\}$ is stochastically equicontinuous at $\tau = \tau_0$ with $\rho_{\mathcal{T}}$ defined by $\rho_{\mathcal{T}}(\tau, \tau_0) = \sup_{N\geq 1}\left[\frac{1}{N}\Sigma_1^N E\|(\tau(X_t) - \tau_0(X_t))\psi(Z_t, \theta_0)\|^2\right]^{1/2}$.

(d) $\psi(z,\theta)$ is twice continuously differentiable in θ for all z in the support of Z_t $\forall t \geq 1$, $\sup_{t\geq 1} E\|\psi(Z_t, \theta_0)\|^{2a} < \infty$, $\sup_{t\geq 1} E \sup_{\theta\in\Theta_0} \left\|\frac{\partial}{\partial\theta'}\psi(Z_t, \theta)\right\|^{a} < \infty$,

$\overline{\lim}_{T\to\infty} \frac{1}{T}\Sigma_1^T E \sup_{\theta\in\Theta_0} \left\|\frac{\partial^2}{\partial\theta_j \theta_\ell}\psi(Z_t, \theta)\right\|^{2a/(2a-1)} < \infty \quad \forall j, \ell \leq p$, and $\sup_{t\geq 1} E\|\Delta_0(X_t)\|^{2a} < \infty$,

where Θ_0 is some neighborhood of θ_0 and a is as in Assumption GMM/CMR 1.

(e) $\lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E\Delta(X_t)'\Omega^{-1}(X_t)\frac{\partial}{\partial\theta'}\psi(Z_t, \theta)$ exists uniformly over $(\theta,\tau) = (\theta, \Delta'\Omega^{-1}) \in \Theta_0 \times \mathcal{T}$ and $\{\Delta(X_t)'\Omega^{-1}(X_t)\frac{\partial}{\partial\theta'}\psi(Z_t, \theta) : t \geq 1\}$ satisfies a uniform WLLN over $(\theta,\tau) = (\theta, \Delta'\Omega^{-1}) \in \Theta_0 \times \mathcal{T}$.

To see that Assumptions GMM/CMR 1 and 2 imply Assumption 2 for the GMM estimator, we proceed as follows: Assumption 2(a) follows from GMM/CMR 1, Theorem I.1, and GMM/CMR 2(a). The first part of Assumption 2(b) follows from GMM/CMR 1(e). The second part of Assumption 2(b) can be shown to hold (in a similar fashion to the first part of Assumption 1(b)) when $\rho_{\mathcal{T}}$ is given by (3.23) using GMM/CMR 1(c), 1(e), and 2(d). The third part of Assumption 2(b) is the same as that of Assumption 1(b) and holds by GMM/CMR 1 as discussed above. The asymptotic orthogonality Assumption 2(c) holds trivially for the GMM estimator because (3.26) holds. In particular, we have

$$(3.30) \qquad \sqrt{T}\,\bar{m}_T^*(\theta_0, \tau) = \frac{1}{\sqrt{T}}\Sigma_1^T E\tau(X_t)\psi(Z_t, \theta_0) = \underset{\sim}{0} \ \ \forall\tau, \ \forall T \geq 1,$$

since $E(\psi(Z_t, \theta_0)|X_t) = \underset{\sim}{0}$ a.s. $\forall t \geq 1$ by definition of the CMR model in (2.2).

Assumption 2(d) requires $\frac{1}{\sqrt{T}}\Sigma_1^T \Delta_0(X_t)'\Omega_0^{-1}(X_t)\psi(Z_t, \theta_0)$ to satisfy a CLT. It does, given GMM/CMR 2(b), by Corollary 1 of Herrndorf (1984). Assumption 2(e) holds by GMM/CMR 2(c). By restricting $\mathcal{T}$, the latter can be verified using the results of ASEM:II. Note that the condition $P(\hat{\tau} \in \mathcal{T}) \to 1$ of GMM/CMR 1(e) must hold with $\mathcal{T}$ restricted in the manner used to obtain the stochastic equicontinuity condition GMM/CMR 2(c). Assumption 2(f) holds trivially since $d(m,\gamma) = m'\gamma m/2$. Assumption 2(g) can be shown to hold using GMM/CMR 1(a), 1(b), 1(e), 2(d), and 2(e) with the pseudo–metrics on $\Theta_0 \times \mathcal{T}$ given by (3.21) and (3.22). Assumption 2(h) holds by GMM/CMR 1(f).

Note that the strong mixing assumption of GMM/CMR 2(b) can be relaxed if need be (to near–epoch dependence, for example), provided sufficient conditions are added for $\frac{1}{\sqrt{T}}\Sigma_1^T \Delta_0(X_t)'\Omega_0^{-1}\psi(Z_t, \theta_0)$ to satisfy a CLT. Also, the pseudo–metric $\rho_{\mathcal{T}}$ used in GMM/CMR 2(c) can be replaced by some other pseudo–metric provided one checks that GMM/CMR 1 and 2 are sufficient for $\rho_{\mathcal{T}}(\hat{\tau}, \tau_0) \xrightarrow{p} 0$ with this pseudo–metric or provided additional conditions are added to GMM/CMR 1 and 2 such that $\rho_{\mathcal{T}}(\hat{\tau}, \tau_0) \xrightarrow{p} 0$.

3.7. Asymptotic Normality in the WLS/PLR Example

For the WLS estimator of the PLR model of (2.6), the following assumption plus Assumption WLS/PLR 1 are sufficient for Assumption 2 (with $d(m,\gamma)$ and $m_t(\theta,\tau)$ as defined in (2.11) and $\rho_{\mathcal{T}}$ given by (3.23)), and hence, for the asymptotic normality of the WLS estimator $\hat{\theta}$:

ASSUMPTION WLS/PLR 2: (a) θ_0 is in the interior of Θ.

(b) $\int \|\hat{\tau}_j(z) - \tau_{j0}(z)\|^4 dP(z) \xrightarrow{p} 0$ for $j = 1, 2, 3$.

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

(d) $S = \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} EU_s U_t [X_s - E(X_s|Z_s)][X_t - E(X_t|Z_t)]' / [E(U_s^2|Z_s)E(U_t^2|Z_t)]$ exists, $E\|U_t X_t\|^{2+\delta} < \infty$ for some $\delta > 0$, and $\{(U_t, X_t, Z_t) : t \geq 1\}$ is a sequence of independent rv's or strong mixing rv's with mixing numbers that satisfy $\sum_{s=1}^{\infty} \alpha(s)^{\delta/(2+\delta)} < \infty$ for δ as above.

(e) $\left\{\frac{1}{\sqrt{T}}\Sigma_1^T(m_t(\theta_0, \tau) - Em_t(\theta_0, \tau)) : T \geq 1\right\}$ is stochastically equicontinuous at $\tau = \tau_0$ with $\rho_{\mathcal{T}}$ defined by (3.23), where $m_t(\theta_0, \tau) = [U_t + \tau_{10}(Z_t) - \tau_1(Z_t) + (\tau_2(Z_t) - \tau_{20}(Z_t))'\theta_0][X_t - \tau_2(Z_t)]/\tau_3(Z_t)$.

(f) $EU_t^4 < \infty$ and $EU_t^4\|X_t\|^4 < \infty$.

(g) $\{[X_t - \tau_2(Z_t)][X_t - \tau_2(Z_t)]'/\tau_3(Z_t) : t \geq 1\}$ satisfies a uniform WLLN over $\tau \in \mathcal{T}$.

To see that Assumptions WLS/PLR 1 and 2 are sufficient for Assumption 2, we proceed as follows: Assumption 2(a) follows from WLS/PLR 1, Theorem I.1, and WLS/PLR 2(a). The first part of Assumption 2(b) follows from WLS/PLR 1(e). The second part of Assumption 2(b) follows from WLS/PLR 1(c), 1(e), 2(b), and 2(f) in a similar fashion to that used to establish the first part of Assumption 1(b).

The orthogonality condition, Assumption 2(c), holds using WLS/PLR 1(e) and 2(c), since

$$\begin{aligned} \|\sqrt{T}\,\bar{m}_T^*(\theta_0, \hat{\tau})\| &= \|\sqrt{T}\int(\tau_{10} - \hat{\tau}_1)(\tau_{20} - \hat{\tau}_2)/\hat{\tau}_3 dP + \sqrt{T}\int(\hat{\tau}_2 - \tau_{20})'\theta_0(\tau_{20} - \hat{\tau}_2)/\hat{\tau}_3 dP\| \\ &\leq T^{1/4}\left[\int(\hat{\tau}_1 - \tau_{10})^2 dP\right]^{1/2} T^{1/4}\left[\int\|\hat{\tau}_2 - \tau_{20}\|^2 dP\right]^{1/2}/\epsilon \\ &\quad + \|\theta_0\|\sqrt{T}\int\|\hat{\tau}_2 - \tau_{20}\|^2 dP/\epsilon + o_p(1), \end{aligned} \tag{3.31}$$

where $\hat{\tau}_j$, τ_{j0}, and dP abbreviate $\hat{\tau}_j(z)$, $\tau_{j0}(z)$, and $dP(z)$, respectively, for $j = 1, 2, 3$.

Assumption 2(d) requires that $\frac{1}{\sqrt{T}}\Sigma_1^T U_t[X_t - \tau_{20}(Z_t)]/\tau_{30}(Z_t)$ satisfies a CLT. It does, given WLS/PLR 2(d), by Corollary 1 of Herrndorf (1984). Assumption 2(e) holds by WLS/PLR 2(e). Note that $\mathcal{T}$ must be defined in the same way in WLS/PLR 1(e) and

2(e). Assumption 2(f) holds trivially since $d(m,\gamma) = m'm/2$. Assumption 2(g) can be shown to hold using WLS/PLR 1, 2(f), and 2(g) with the pseudo–metrics on $\Theta_0 \times \mathcal{T}$ given by (3.21) and (3.22). Assumption 2(h) holds by WLS/PLR 1(f).

*3.8. Discussion of Assumption 2**

We now discuss Assumption 2* and compare it with Assumption 2. First, consider Assumption 2*(a). The second part of this assumption is a necessary condition for the asymptotic normality of $\hat{\theta}$ with covariance matrix V. Essentially, it requires the "first order conditions" corresponding to the minimization problem that defines $\hat{\theta}$ to be approximately satisfied for T sufficiently large. In the case of M–estimators where $d(m,\gamma) = m'm/2$, $p = v$, and no nuisance parameters τ and γ exist, this condition reduces to the same condition imposed by Huber (1967) to obtain his asymptotic normality results for M–estimators.

Assumption 2*(a) holds if (i) $m_t(\theta,\tau)$ is once continuously differentiable in θ on Θ_0, $\forall \tau \in \mathcal{T}$, $\forall t \geq 1$, $\forall \omega \in \Omega$ (which is implied by Assumption 2(g)), (ii) $m_t(\theta,\tau)$ and θ have the same dimension p $(= v)$, (iii) $\frac{\partial}{\partial \theta'} \bar{m}_T(\hat{\theta},\hat{\tau})$ is nonsingular with probability $\to 1$ (which is implied by Assumptions 2(g) and (h)), and (iv) Assumption 2(a) holds. Thus, Assumption 2 implies Assumption 2*(a) when $p = v$.

Assumption 2*(a) follows from conditions (i)–(iv) because the latter imply that the first order conditions for the minimization problem of (2.1) hold with probability 1. If conditions (i)–(iv) do not hold, then the second part of Assumption 2*(a) needs to be established on a case by case basis. For example, if $d(m,\gamma) = m'm/2$ and M is a square nonsingular matrix, then the second part of Assumption 2*(a) requires $\sqrt{T}\, \bar{m}_T(\hat{\theta},\hat{\tau}) \xrightarrow{p} \underset{\sim}{0}$. This can be shown by establishing that the first order conditions $\bar{m}_T(\hat{\theta},\hat{\tau}) = \underset{\sim}{0}$ hold with probability $\to 1$. See Ruppert and Carroll (1980) and Powell (1984, 1986a, b) for the verification of this condition in certain examples.

Assumption 2*(b) requires $\rho_{\Theta\times\mathcal{T}}((\hat{\theta},\hat{\tau}), (\theta_0, \tau_0)) \xrightarrow{p} 0$, where $\rho_{\Theta\times\mathcal{T}}$ is a pseudo-metric on $\Theta \times \mathcal{T}$ that is suitable for establishing the stochastic equicontinuity condition of Assumption 2*(e). For example, one could take

$$(3.32) \qquad \rho_{\Theta\times\mathcal{T}}((\theta_1, \tau_1), (\theta_2, \tau_2)) = \sup_{N\geq 1}\left[\frac{1}{N}\Sigma_1^N E\|m_t(\theta_1, \tau_1) - m_t(\theta_2, \tau_2)\|^2\right]^{1/2} .$$

Alternatively, one could take

$$(3.33) \qquad \rho_{\Theta\times\mathcal{T}}((\theta_1, \tau_1), (\theta_2, \tau_2)) = \left[\int_{\mathcal{W}}\|m(w, \theta_1, \tau_1) - m(w, \theta_2, \tau_2))\|^2 dw\right]^{1/2} ,$$

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

Assumption 2*(e) is stronger than Assumption 2(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 2*(e) can be verified using the results of ASEM:II. Its verification, however, often is somewhat less simple than that of Assumption 2(e), because the evaluation of $\bar{m}_T(\theta,\tau) - \bar{m}_T^*(\theta,\tau)$ at $\theta = \theta_0$ in Assumption 2(e) often yields a convenient simplification.

Assumption 2*(g) is weaker than Assumption 2(g) because it requires twice continuous differentiability of $Em_t(\theta,\tau)$ rather than of $m_t(\theta,\tau)$ and it does not require certain uniform WLLNs to hold. Assumption 2*(g) requires continuity of $m(\theta,\tau)$ and $M(\theta,\tau)$ with respect to some pseudo–metrics on $\Theta_0 \times \mathcal{T}$ for which $(\hat{\theta},\hat{\tau}) \xrightarrow{p} (\theta_0, \tau_0)$. The most convenient choices of pseudo–metrics are $\rho(\cdot,\cdot)$ as in (3.21) for establishing the continuity of $m(\theta,\tau)$ and

$$(3.34) \qquad \rho((\theta_1, \tau_1), (\theta_2, \tau_2)) = \overline{\lim_{N\to\infty}} \frac{1}{T}\Sigma_1^N\left\|\frac{\partial}{\partial\theta'}Em_t(\theta_1, \tau_1) - \frac{\partial}{\partial\theta'}Em_t(\theta_2, \tau_2)\right\|$$

for establishing continuity of $M(\theta,\tau)$. With these choices, continuity of $m(\theta,\tau)$ and $M(\theta,\tau)$ holds automatically and it suffices to verify that $\rho((\hat{\theta},\hat{\tau}), (\theta_0, \tau_0)) \xrightarrow{p} 0$ for each choice of $\rho(\cdot,\cdot)$.

In sum, neither Assumption 2 nor Assumption 2* is weaker. The main difference between the two is that they consider different tradeoffs between the stochastic equicontinuity assumption and the smoothness assumption on the summands $\{m_t(\theta,\tau)\}$.

*3.9. Discussion of Assumption 2***

To conclude this section, we briefly discuss Assumption 2**. This assumption applies in the common case in the literature in which τ is finite dimensional and $m_t(\theta,\tau)$ is differentiable in θ and τ. This is the case that is considered in Bierens (1981), Burguete *et al.* (1982), Gallant (1987b), Gallant and White (1988), and Andrews and Fair (1988). Assumption 2** is comparable to (i.e. no harder to verify than) the assumptions used in these papers. In particular, Assumption 2** is essentially the same as that used in Andrews and Fair (1988) to obtain asymptotic normality of parametric estimators (with finite dimensional nuisance parameters). Assumption 2(c), which is part of Assumption 2**, is less primitive than the corresponding assumption in Andrews and Fair (1988). But, Assumption 2(c) usually is so easy to verify in applications for which 2**(e) holds that it probably is not worthwhile to give more primitive conditions that imply it. (Nevertheless, if desired, such conditions can be derived quite easily when τ is finite dimensional and $m_t(\theta_0, \tau)$ is differentiable in τ.)

4. COVARIANCE MATRIX ESTIMATION

In this section we consider estimation of the covariance matrix V of the MINPIN estimators $\{\hat{\theta}\}$. We use the same definitions for $d(m,\gamma)$ and $m_t(\theta,\tau)$ in this section as were used for the asymptotic normality results of Section 3. When Assumption 2 or 2** holds, we define:

$$\text{(4.1)} \qquad \hat{D} = \frac{\partial^2}{\partial m \partial m'} d(\bar{m}_T(\hat{\theta},\hat{\tau}), \hat{\gamma}) \text{ and } \hat{M} = \frac{1}{T}\Sigma_1^T \frac{\partial}{\partial \theta'} m_t(\hat{\theta},\hat{\tau}) .$$

Under Assumption 2 or 2**, $\hat{D}$ and $\hat{M}$ are consistent for D and M respectively.

When Assumption 2* is employed, we define $\hat{D}$ as above and it is consistent for D. We define $\hat{M}$ differently in this case, howevever, since $\hat{M}$ of (4.1) does not exist when $m_t(\hat{\theta},\hat{\tau})$ is not differentiable in θ. In particular, we use a finite difference estimator of M. For a constant or rv $\epsilon_T > 0$, define the j–th column of $\hat{M}$ by

$$\hat{M}_j = \tfrac{1}{T}\Sigma_1^T(m_t(\hat{\theta} + \epsilon_T e_j, \hat{\tau}) - m_t(\hat{\theta} - \epsilon_T e_j, \hat{\tau}))/(2\epsilon_T), \tag{4.2}$$

where $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)'$ is the j–th elementary p–vector for $j = 1, \ldots, p$. The estimator $\hat{M}$ of (4.2) is consistent for M under Assumption 2* and the following assumption.

ASSUMPTION 3*: (a) $\epsilon_T \xrightarrow{p} 0$ and $\epsilon_T^{-1} = O_p(\sqrt{T})$.

(b) $Em_t(\theta,\tau)$ is differentiable in θ uniformly over $\theta \in \Theta_0$, $\tau \in \mathcal{T}$, and $t \geq 1$ (i.e.,

$$\lim_{\epsilon\to 0}\sup_{t\geq 1}\sup_{\theta\in\Theta_0,\tau\in\mathcal{T}}\left\|(Em_t(\theta + \epsilon e_j, \tau) - Em_t(\theta - \epsilon e_j, \tau))/(2\epsilon) - \frac{\partial}{\partial\theta_j}Em_t(\theta,\tau)\right\| = 0$$

$\forall j = 1, \ldots, p$).

(c) $(\hat{\theta} + \epsilon_T e_j, \hat{\tau}) \xrightarrow{p} (\theta_0, \tau_0)$ $\forall j = 1, \ldots, p$.

Note that Assumption 3*(c) is very similar to the second part of Assumption 2*(b). In consequence, verification of the former is usually a trivial extension of the verification of the latter.

In some examples where Assumption 2* is used, the matrix M simplifies and an alternative estimator to $\hat{M}$ of (4.2) is available. For example, with the WC–LAD estimator of Example 6.2 below, $M = -4E\tau_0^2(Z_t)1(X_t'\theta_0 < C_t)X_tX_t'$, estimators of $\tau_0(\cdot)$ and θ_0 are available, and $Z_t = (X_t', C_t)'$ is observed. In this case, one can estimate M using $\hat{M}$ of (4.2) or using $-4\frac{1}{T}\Sigma_1^T\hat{\tau}^2(Z_t)1(X_t'\hat{\theta} < C_t)X_tX_t'$.

Next, we discuss estimation of the matrix S. Let $\hat{S}$ be an estimator of S. If $\{m_t(\theta_0, \tau_0)\}$ is a sequence of independent or orthogonal rv's, then we can take

$$\hat{S} = \tfrac{1}{T}\Sigma_1^T m_t(\hat{\theta},\hat{\tau})m_t(\hat{\theta},\hat{\tau})'. \tag{4.3}$$

If $\{m_t(\theta_0, \tau_0)\}$ is m–dependent, then the following estimator can be used

$$(4.4) \qquad \hat{S} = \frac{1}{T}\Sigma_1^T \hat{m}_t\hat{m}_t' + \sum_{v=1}^{m} \frac{1}{T}\Sigma_{1+v}^T[\hat{m}_t\hat{m}_{t-v}' + \hat{m}_{t-v}\hat{m}_t'] ,$$

where $\hat{m}_t = m_t(\hat{\theta},\hat{\tau})$. If $\{m_t(\theta_0, \tau_0)\}$ is neither orthogonal nor m–dependent, then a more complicated estimator of S is required. In particular, heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimators that have been defined for parametric models can be used, see White (1984, pp. 147–161), Newey and West (1987), Gallant (1987b, pp. 553, 552, 573), Andrews (1991b), and Andrews and Monahan (1990). For semiparametric models these estimators can be defined in exactly the same way as for parametric models, using $\{m_t(\hat{\theta},\hat{\tau})\}$ as the underlying rv's. The consistency of such HAC estimators when τ is infinite dimensional does not follow from the results given in any of the papers above, however, because these results make use of mean value or Taylor expansions in the estimated parameters, which rely on the finite dimensional character of the parameters. Nevertheless, conditions under which various kernel HAC estimators are consistent when τ is infinite dimensional are given in Andrews (1990a).

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

ASSUMPTION 3: $\hat{S} \xrightarrow{p} S$ *(where* S *is as in Assumption* 2, 2*, *or* 2***).*

Let $\hat{\mathcal{I}} = \hat{M}'\hat{D}\hat{M}$, $\hat{\mathcal{J}} = \hat{M}'\hat{D}\hat{S}\hat{D}\hat{M}$, and $\hat{V} = \hat{\mathcal{J}}^{-1}\hat{\mathcal{I}}\hat{\mathcal{J}}^{-1}$. (Under the assumptions given below, $\hat{\mathcal{J}}$ is nonsingular with probability $\to 1$, and so $\hat{V}$ is well–defined with probability $\to 1$.)

THEOREM I.3: *Under Assumptions* 2 *and* 3, *or* 2** *and* 3, *or* 2*, 3*, *and* 3, $\hat{M} \xrightarrow{p} M$, $\hat{D} \xrightarrow{p} D$, *and* $\hat{V} \xrightarrow{p} V$, *where* $\hat{M}$ *is as defined in* (4.1) *or* (4.1) *or* (4.2) *respectively.*

COMMENT: When V simplifies, as occurs in many applications, then $\hat{V}$ simplifies or simpler estimators than $\hat{V}$ can be constructed.

5. TESTS OF NONLINEAR RESTRICTIONS

In this section, we consider tests of nonlinear restrictions of the form $H_0 : h(\theta_0) = \underset{\sim}{0}$. We start by defining the W, LM, and LR statistics to be considered and stating sufficient conditions for these statistics to be asymptotically chi–square under the null hypothesis. Then, we provide local power results.

5.1. Definition of Test Statistics and Statement of Assumptions

The R^r–valued function $h(\cdot)$ defining the restrictions is assumed to satisfy:

ASSUMPTION 4: (a) $h(\theta)$ *is continuously differentiable in a neighborhood of* θ_0 *and* $H = \frac{\partial}{\partial\theta'}h(\theta_0)$ *has full rank* $r\ (\le p)$.
(b) V *is nonsingular.*

Throughout this section we use the same definitions of $d(m,\gamma)$ and $m_t(\theta,\tau)$ as were used in the asymptotic normality results of Section 3.

The Wald statistic for testing H_0 is defined to be

$$(5.1)\qquad W_T = Th(\hat{\theta})'(\hat{H}\hat{V}\hat{H}')^{-1}h(\hat{\theta}) ,$$

where $\hat{H} = \frac{\partial}{\partial\theta'}h(\hat{\theta})$. Since $\hat{H}\hat{V}\hat{H}' \xrightarrow{p} HVH'$ and HVH' is nonsingular under Assumption 4, $\hat{H}\hat{V}\hat{H}'$ is nonsingular and $(\hat{H}\hat{V}\hat{H}')^{-1}$ is well–defined with probability $\to 1$.[10]

Two LM and two LR statistics for testing H_0 , denoted LM_{aT} , LM_{bT} , LR_{aT} , and LR_{bT} , are defined below. The LM_a and LR_a statistics are defined for any criterion function $d(\bar{m}_T(\theta,\hat{\tau}),\ \hat{\gamma})$ that satisfies Assumption 2 or 2**, although the LR_a statistic has the desired χ^2 asymptotic null distribution only under special conditions. The LM_b and LR_b statistics are defined only for criterion functions $d(\bar{m}_T(\theta,\hat{\tau}),\ \hat{\gamma})$ for which $d(m,\gamma) = m'm/2$ and $\bar{m}_T(\theta,\tau) = \frac{\partial}{\partial\theta}\bar{\rho}_T(\theta,\tau)$ for some function $\bar{\rho}_T(\theta,\tau)$. For example, these conditions hold for the WLS estimator of the PLR model and ML estimators, but not for the GMM estimator of the CMR model. For ML estimators, the LM_b and LR_b statistics correspond to the usual LM and LR test statistics.

In a given context, either the LR_a or the LR_b statistic may have the desired asymptotic χ^2 null distribution, but not both. (The only exception to this is in the very rare case that M is proportional to the identity matrix.) With the LM statistics, however, both can have the desired asymptotic χ^2 null distribution in the same context. The reason for the appearance of two alternative LM statistics is that there are two possible restricted estimators of θ_0, denoted $\tilde{\theta}_a$ and $\tilde{\theta}_b$, in those cases where LM_b and LR_b can be defined.

First we define the LM_a and LR_a statistics. They make use of the restricted MINPIN estimator $\tilde{\theta}_a$:

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

$$(5.2) \qquad d(\bar{m}_T(\tilde{\theta}_a, \hat{\tau}), \hat{\gamma}) = \inf\{d(\bar{m}_T(\theta, \hat{\tau}), \hat{\gamma}) : \theta \in \Theta, h(\theta) = \underset{\sim}{0}\}$$

with probability $\to 1$.

Suppose the null hypothesis is true. If Assumption 1 holds for the parameter space Θ, it also holds for the parameter space $\tilde{\Theta}_0 = \{\theta \in \Theta : h(\theta) = \underset{\sim}{0}\}$. Thus, Assumption 1 and Theorem I.1 imply that $\tilde{\theta}_a \xrightarrow{p} \theta_0$ under the null hypothesis. In consequence, the following assumption is straightforward to verify when θ_0 satisfies the null hypothesis:

ASSUMPTION 5a: $\tilde{\theta}_a \xrightarrow{p} \theta_0$.

The LM_a statistic uses an estimator of V that is constructed with the restricted estimator $\tilde{\theta}_a$ in place of $\hat{\theta}$. Let $\tilde{D} = \frac{\partial^2}{\partial m \partial m'} d(\bar{m}_T(\tilde{\theta}_a, \hat{\tau}), \hat{\gamma})$, $\tilde{M} = \frac{1}{T}\Sigma_1^T \frac{\partial}{\partial \theta'} m_t(\tilde{\theta}_a, \hat{\tau})$, $\tilde{J} = \tilde{M}'\tilde{D}\tilde{M}$, $\tilde{S} = \hat{S}(\tilde{\theta}_a)$, and $\tilde{H} = \frac{\partial}{\partial \theta'} h(\tilde{\theta}_a)$. Note that the estimators of the nuisance parameters (τ_0, γ_0) still are denoted $(\hat{\tau}, \hat{\gamma})$ even though they may be restricted estimators of (τ_0, γ_0). The same is true of the estimator $\hat{S}$ of S. With this notation, we do not need to alter Assumptions 2(b) or 3 when restricted estimators of (τ_0, γ_0) are used. Let $\tilde{\mathcal{I}} = \tilde{M}'\hat{D}\hat{S}\tilde{D}\tilde{M}$ and $\tilde{V} = \tilde{\mathcal{J}}^{-1}\tilde{\mathcal{I}}\tilde{\mathcal{J}}^{-1}$. As with $\hat{V}$, the estimator $\tilde{V}$ can be simplified when V simplifies, as often occurs in applications of interest.

The LM_a statistic for testing H_0 is defined to be

$$\text{(5.3)} \qquad LM_{aT} = T\frac{\partial}{\partial\theta'}d(\bar{m}_T(\tilde{\theta}_a, \hat{\tau}), \hat{\gamma})\tilde{\mathcal{J}}^{-1}\tilde{H}'(\tilde{H}\tilde{V}\tilde{H}')^{-1}\tilde{H}\tilde{\mathcal{J}}^{-1}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a, \hat{\tau}), \hat{\gamma}) .$$

As shown below, this statistic often simplifies considerably.

If $d(m,\gamma)$ is a quadratic form in m, as is usually the case, then the LM_a statistic is a quadratic form in the v–vector $\bar{m}_T(\tilde{\theta}_a, \hat{\tau})$. For example, with the WLS estimator of the PLR model, $\bar{m}_T(\tilde{\theta}_a, \hat{\tau})$ is just the vector of normal equations for the estimator evaluated at the restricted estimator $\tilde{\theta}_a$. Similarly, with the GMM estimator of the CMR model, $\bar{m}_T(\tilde{\theta}_a, \hat{\tau})$ is just the sample average of the weighted cross–products between the instrumental variables and the model evaluated at the restricted estimator $\tilde{\theta}_a$. With a maximum likelihood (ML) estimator, $\bar{m}_T(\tilde{\theta}_a, \hat{\tau})$ is the average of the score functions for different observations evaluated at the restricted estimator $\tilde{\theta}_a$. In each case, $\bar{m}_T(\tilde{\theta}_a, \hat{\tau})$ is a random variable that is roughly centered at $\underset{\sim}{0}$ under the null hypothesis, since $\bar{m}_T^*(\theta_0, \tau_0) = \underset{\sim}{0}$.

Next, the LR_a statistic is defined by

$$\text{(5.4)} \qquad LR_{aT} = 2T(d(\bar{m}_T(\tilde{\theta}_a, \hat{\tau}), \hat{\gamma}) - d(\bar{m}_T(\hat{\theta},\hat{\tau}), \hat{\gamma}))/\hat{b} ,$$

where $\hat{b}$ is a scalar rv defined in Assumption 6a below. The preliminary estimators $(\hat{\tau},\hat{\gamma})$ used in LR_{aT} may be restricted or unrestricted estimators of (τ_0, γ_0). For the asymptotic results given below they must be the same in both criterion functions used to calculate LR_{aT}, however, and they must be such that both $\hat{\theta}$ and $\tilde{\theta}_a$ are consistent under the null hypothesis.

The LR_a statistic has the desired asymptotic χ^2 null distribution only under the following assumption:

ASSUMPTION 6a: $\mathcal{I} = b\mathcal{J}$ *for some scalar constant* $b \neq 0$ *and* $\hat{b} \xrightarrow{p} b$ *for some sequence of non-zero rv's* $\{\hat{b}\}$.

Assumption 6a is satisfied with $b = \hat{b} = 1$ by the GMM estimator of the CMR model when $\{\psi(Z_t, \theta_0) : t \geq 1\}$ is an uncorrelated sequence conditional on $\{X_t : t \geq 1\}$. In this case, LR_a is given by the difference between a criterion function restricted and unrestricted:

(5.5) $$LR_a = T\bar{m}_T(\tilde{\theta}_a, \hat{\tau})'\hat{\gamma}\bar{m}_T(\tilde{\theta}_a, \hat{\tau}) - T\bar{m}_T(\hat{\theta},\hat{\tau})'\hat{\gamma}\bar{m}_T(\hat{\theta},\hat{\tau}) .$$

Assumption 6a also is satisfied by two and three stage LS estimators of nonlinear simultaneous equations models when the errors are uncorrelated and homoskedastic or when these estimators use preliminary estimators that transform the observations to achieve uncorrelated and homoskedastic errors. Assumption 6a is not satisfied by the WLS estimator in the PLR model since $\mathcal{J} = S^2$ and $\mathcal{I} = S^3$. Nor is it satisfied by ML estimators. For these estimators, the LR_b statistic (defined below) must be used.

If $\hat{\mathcal{I}} = \hat{b}\hat{\mathcal{J}}$ for some scalar rv $\hat{b} \neq 0$, as usually occurs when Assumption 6a holds, then $\hat{V}$ and W_T simplify:

(5.6) $$\hat{V} = \hat{b}\hat{\mathcal{J}}^{-1} \text{ and } W_T = Th(\hat{\theta})'(\hat{H}\hat{\mathcal{J}}^{-1}\hat{H}')^{-1}h(\hat{\theta})/\hat{b} .$$

Similarly, if $\tilde{\mathcal{I}} = \hat{b}\tilde{\mathcal{J}}$ for some scalar $\hat{b} \neq 0$, as usually occurs when Assumption 6a holds, then $\tilde{V}$ and LM_{aT} simplify:

(5.7) $$\tilde{V} = \hat{b}\tilde{\mathcal{J}}^{-1} \text{ and } LM_{aT} \doteq T\frac{\partial}{\partial\theta'}d(\bar{m}_T(\tilde{\theta}_a, \hat{\tau}), \hat{\gamma})\tilde{\mathcal{J}}^{-1}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a, \hat{\tau}), \hat{\gamma})/\hat{b} ,$$

where $\doteq$ denotes equality that holds with probability $\to 1$. The simplification of LM_{aT} holds because $\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a, \hat{\tau}), \hat{\gamma}) \doteq -\tilde{H}'\tilde{\lambda}$ for some vector $\tilde{\lambda}$ of Lagrange multipliers.

The simplifications of W_T and LM_{aT} given in (5.6) and (5.7) hold when using the GMM estimator of the CMR model provided $\psi(Z_t, \theta_0)$ is conditionally uncorrelated across time periods. These simplifications do not hold, however, for the W_T statistic based on the WLS estimator of the PLR model. A different simplification is given below that is applicable in the latter example.

Next, we consider the context in which the LM_b and LR_b statistics are defined:

ASSUMPTION 6b: (i) $d(m,\gamma) = m'm/2$. *There exist functions* $\{\rho_{Tt}(W_{Tt},\theta,\tau)\}$ *such that* $\frac{\partial}{\partial\theta}\rho_{Tt}(W_{Tt},\theta,\tau) = m_{Tt}(W_{Tt}, \theta, \tau)$ $\forall t, \forall T$ or $\frac{\partial}{\partial\theta}\rho_{Tt}(W_{Tt},\theta,\tau) = -m_{Tt}(W_{Tt},\theta,\tau)$ $\forall T, \forall T$. *With probability* $\to 1$, $\hat{\theta}$ *solves* $\bar{\rho}_T(\hat{\theta},\hat{\tau}) = \inf\{\bar{\rho}_T(\theta,\hat{\tau}) : \theta \in \Theta\}$, *where* $\bar{\rho}_T(\theta,\hat{\tau}) = \frac{1}{T}\Sigma_1^T\rho_{Tt}(W_{Tt}, \theta, \hat{\tau})$.

(ii) $S = cM$ *for some scalar* $c \neq 0$ *and* $\hat{c} \xrightarrow{p} c$ *for some sequence of non-zero rv's* $\{\hat{c}\}$.

Assumption 6b is satisfied by ML estimators for finite dimensional parametric models. Assumption 6b(i) is satisfied by the LS estimator, feasible GLS estimators, and many M-estimators for nonlinear regression models and PLR models, among others. Assumption 6b(ii) is satisfied with these estimators only when the errors are uncorrelated and homoskedastic or when these estimators use preliminary estimators that transform the observations to achieve uncorrelated and homoskedastic errors. The WLS estimator of the PLR model is an example of such an estimator. Specifically, for this estimator, we have

$$(5.8) \qquad \rho_{Tt}(W_{Tt}, \theta, \tau) = \tfrac{1}{2}[Y_t - \tau_1(Z_t) - (X_t - \tau_2(Z_t))'\theta]^2/\tau_3(Z_t) \text{ and } S = -M .$$

Thus, Assumption 6b holds in this case with $c = \hat{c} = -1$. Assumption 6b also holds for ML estimators with $c = \hat{c} = -1$. Assumption 6b(i) does not hold for the GMM estimator of the CMR model. In consequence, no LM_b or LR_b statistic is defined for that example.

Note that Assumption 6b(i) is compatible with the definition of $\hat{\theta}$ given in (2.1) because an estimator $\hat{\theta}$ that minimizes $\bar{\rho}_T(\theta,\hat{\tau})$ is in the interior of Θ with probability $\to 1$ under Assumption 2 or 2**, and hence, also minimizes $d(\bar{m}_T(\theta,\hat{\tau}),\hat{\gamma})$ with probability $\to 1$.

When Assumption 6b(i) holds the LM_b and LR_b statistics are defined. The LR_b statistic has the desired asymptotic χ^2 null distribution, however, only when Assumption

6b(ii) also holds. Both the LM_b and the LR_b statistics make use of the restricted MINPIN estimator $\tilde{\theta}_b$:

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

$$\bar{\rho}_T(\tilde{\theta}_b, \hat{\tau}) = \inf\{\bar{\rho}_T(\theta,\hat{\tau}) : \theta \in \Theta, h(\theta) = \underset{\sim}{0}\} \tag{5.9}$$

with probability $\to 1$.

Note that $\tilde{\theta}_a$ and $\tilde{\theta}_b$ differ in general, because $\tilde{\theta}_b$ minimizes the minimand $\bar{\rho}_T(\theta,\hat{\tau})$ subject to the restrictions $h(\theta) = \underset{\sim}{0}$, whereas $\tilde{\theta}_a$ minimizes the inner product of the first order conditions of this minimization problem, i.e. $\frac{\partial}{\partial\theta'}\bar{\rho}_T(\theta,\hat{\tau})\frac{\partial}{\partial\theta}\bar{\rho}_T(\theta,\hat{\tau})$, subject to the restrictions.

As with $\{\tilde{\theta}_a\}$, the consistency of $\{\tilde{\theta}_b\}$ can be established using Theorem I.1, so the following assumption is straightforward to verify when θ_0 satisfies the null hypothesis:

ASSUMPTION 5b: $\tilde{\theta}_b \xrightarrow{p} \theta_0$.

By definition,

$$\begin{aligned} LM_{bT} &= T\frac{\partial}{\partial\theta'}\bar{\rho}_T(\tilde{\theta}_b, \hat{\tau})\tilde{M}^{-1}\tilde{H}'(\tilde{H}\tilde{M}^{-1}\hat{S}\tilde{M}^{-1}\tilde{H}')^{-1}\tilde{H}\tilde{M}^{-1}\frac{\partial}{\partial\theta}\bar{\rho}_T(\tilde{\theta}_b, \hat{\tau}) \text{ and} \\ LR_{bT} &= 2T(\bar{\rho}_T(\tilde{\theta}_b, \hat{\tau}) - \bar{\rho}_T(\hat{\theta},\hat{\tau}))/|\hat{c}| , \end{aligned} \tag{5.10}$$

where $\tilde{M}$, $\tilde{H}$, and $\hat{S}$ are as defined above but with $\tilde{\theta}_b$ in place of $\tilde{\theta}_a$. As with LR_{aT} above, the preliminary estimator $\hat{\tau}$ must be the same in both criterion functions used to calculate LR_{bT} and used to define $\tilde{\theta}_b$ and $\hat{\theta}$.

With the WLS estimator of the PLR model, LM_{bT} is a quadratic form in the vector

$$\frac{1}{T}\Sigma_1^T[Y_t - \hat{\tau}_1(Z_t) - (X_t - \hat{\tau}_2(Z_t))'\tilde{\theta}_b][X_t - \hat{\tau}_2(Z_t)]/\hat{\tau}_3(Z_t) . \tag{5.11}$$

In addition, LR_{bT} is given simply by the difference between a restricted and an unrestricted criterion function:

$$\text{(5.12)}\qquad \begin{aligned} LR_{bT} = \Sigma_1^T[Y_t - \hat{\tau}_1(Z_t) - (X_t - \hat{\tau}_2(Z_t))'\tilde{\theta}_b]^2/\hat{\tau}_3(Z_t) \\ - \Sigma_1^T[Y_t - \hat{\tau}_1(Z_t) - (X_t - \hat{\tau}_2(Z_t))'\hat{\theta}]^2/\hat{\tau}_3(Z_t)\,. \end{aligned}$$

With the ML estimator, LR_{bT} is the standard LR test statistic (i.e., minus two times the log of the likelihood ratio).

When Assumptions 2 and 6b(i) (or 2** and 6b(i)) hold, W_T and LM_{aT} simplify. In this case, $D = I_p$, $M = \lim_{T\to\infty} \frac{1}{T}\Sigma_1^T E \frac{\partial^2}{\partial\theta\partial\theta'}\rho_{Tt}(W_t, \theta_0, \tau_0)$, $\mathcal{J} = M^2$, $\mathcal{I} = MSM$, $\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a, \hat{\tau}), \hat{\gamma}) = \tilde{M}\frac{\partial}{\partial\theta}\bar{\rho}_T(\tilde{\theta}_a, \hat{\tau})$, M is nonsingular by Assumption 2(h), and W_T and LM_{aT} are given by

$$\text{(5.13)}\qquad \begin{aligned} W_T &\doteq Th(\hat{\theta})'(\hat{H}\hat{M}^{-1}\hat{S}\hat{M}^{-1}\hat{H}')^{-1}h(\hat{\theta}) \quad \text{and} \\ LM_{aT} &\doteq T\frac{\partial}{\partial\theta'}\bar{\rho}_T(\tilde{\theta}_a, \hat{\tau})\tilde{M}^{-1}\tilde{H}'(\tilde{H}\tilde{M}^{-1}\hat{S}\tilde{M}^{-1}\tilde{H}')^{-1}\tilde{H}\tilde{M}^{-1}\frac{\partial}{\partial\theta}\bar{\rho}_T(\tilde{\theta}_a, \hat{\tau})\,. \end{aligned}$$

Note that in this case LM_{aT} is the same as LM_{bT} with $\tilde{\theta}_b$ replaced by $\tilde{\theta}_a$.

If, in addition to Assumptions 2 and 6b(i) (or 2** and 6b(i)), we have (i) $\hat{S} = \hat{c}\hat{M}$ or (ii) $\hat{S} = \hat{c}\tilde{M}$ for some scalar rv $\hat{c} \neq 0$, as usually occurs if Assumption 6b(ii) holds, then (i) W_T or (ii) LM_{aT} and LM_{bT}, respectively, simplify:

$$\text{(5.14)}\qquad \begin{aligned} W_T &\doteq Th(\hat{\theta})'(\hat{H}\hat{M}^{-1}\hat{H}')^{-1}h(\hat{\theta})/\hat{c}\,, \\ LM_{aT} &\doteq T\frac{\partial}{\partial\theta'}\bar{\rho}_T(\tilde{\theta}_a, \hat{\tau})\tilde{M}^{-1}\frac{\partial}{\partial\theta}\bar{\rho}_T(\tilde{\theta}_a, \hat{\tau})/\hat{c}\,, \quad \text{and} \\ LM_{bT} &\doteq T\frac{\partial}{\partial\theta'}\bar{\rho}_T(\tilde{\theta}_b, \hat{\tau})\tilde{M}^{-1}\frac{\partial}{\partial\theta}\bar{\rho}_T(\tilde{\theta}_b, \hat{\tau})/\hat{c}\,. \end{aligned}$$

The simplification of LM_{aT} uses the fact that $\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a,\hat{\tau}), \hat{\gamma})$ $(= \tilde{M}\frac{\partial}{\partial\theta}\bar{\rho}_T(\tilde{\theta}_a,\hat{\tau}))$ $= -\tilde{H}'\tilde{\lambda}$ for some vector $\tilde{\lambda}$ of Lagrange multipliers. The simplification of LM_{bT} uses the fact that $\frac{\partial}{\partial\theta}\bar{\rho}_T(\tilde{\theta}_b, \hat{\tau}) \doteq -\tilde{H}'\tilde{\eta}$ for some vector $\tilde{\eta}$ of Lagrange multipliers.

The simplifications of W_T , LM_{aT} , and LM_{bT} given in (5.14) apply to the test statistics based on the WLS estimator of the PLR model, since it satisfies Assumption 6b. The simplifications also apply to tests based on ML estimators, since the latter also satisfy

Assumption 6b. In particular, when $\bar{\rho}_T(\theta,\hat{\tau})$ corresponds to the ML estimator, the simplified version of LM_{bT} given in (5.14) is the standard LM test statistic.

5.2. Asymptotic Results under the Null Hypothesis

We now state the conditions under which the test statistics defined above have the desired asymptotic χ^2 null distribution:

THEOREM I.4: *Suppose Assumptions 2–4 or 2**, 3, and 4 hold under* P *and* θ_0 *satisfies the null hypothesis. Then the following results hold:*

(a) $W_T \xrightarrow{d} \chi^2_r$ *and Assumption 2 or 2** can be replaced by 2* and 3* if need be,*

(b) $LM_{aT} \xrightarrow{d} \chi^2_r$ *provided Assumption 5a also holds,*

(c) $LR_{aT} \xrightarrow{d} \chi^2_r$ *provided Assumptions 5a and 6a hold in place of 3,*

(d) $LM_{bT} \xrightarrow{d} \chi^2_r$ *provided Assumptions 5b and 6b(i) also hold, and*

(e) $LR_{bT} \xrightarrow{d} \chi^2_r$ *provided Assumptions 5b and 6b hold in place of 3,*

where r *is the number of restrictions and* χ^2_r *denotes the chi–square distribution with* r *degrees of freedom.*

COMMENTS: 1. The estimation of the infinite dimensional nuisance parameter τ_0 has no effect on the asymptotic distribution of the test statistics considered here or on the chosen form of the test statistics. This is analogous to the above estimation results for $\{\hat{\theta}\}$.

2. The results of Theorem I.4 can be extended to incorporate "Chow–type" tests of structural change in semiparametric models. The extension is analogous to the results of Andrews and Fair (1988) for parametric models with known breakpoints and to the results of Andrews (1990b) for parametric models with unknown breakpoints.

5.3. Local Power Results

Next, we present asymptotic local power (ℓp) results for the W, LM, and LR tests. These results can be used to approximate the power functions of the tests. We impose one of the three following assumptions:

ASSUMPTION 2–ℓp: *The distribution* P *of the triangular array* $\{W_{Tt} : t = 1, \ldots, T; T \geq 1\}$ *is such that Assumption* 2 *holds under* P *with* θ_0 *replaced by* $\theta_T = \theta_0 + \eta/\sqrt{T}$ *in part* 2(c) *and in the definition of* $\nu_T(\cdot)$ *for some* $\eta \in R^p$.

ASSUMPTION 2*–ℓp: *The distribution* P *of the triangular array* $\{W_{Tt}\}$ *is such that Assumption* 2* *holds under* P *with* θ_0 *replaced by* $\theta_T = \theta_0 + \eta/\sqrt{T}$ *in parts* 2*(c) *and* (e) *and in the definition of* $\nu_T(\tau_0)$ *in part* 2*(d) *for some* $\eta \in R^p$.

ASSUMPTION 2**–ℓp: *Assumption 2–ℓp holds with "2" replaced by "2**" everywhere it appears.*

The definition of the rv's $\{W_{Tt}\}$ under sequences of local alternatives usually is quite easy to determine in applications. For example, in the PLR model, the sequence of models is

$$(5.15) \qquad Y_{Tt} = X_t'\theta_T + g(Z_t) + U_t \ \ \forall t = 1, \ldots, T, \ T \geq 1$$

and W_{Tt} is defined by (Y_{Tt}, X_t', Z_t') $\forall t = 1, \ldots, T$, $T \geq 1$. In this case, Assumption 2(c) is satisfied with θ_0 replaced by θ_T under the same conditions on $\hat{\tau}$ as are used above to verify Assumption 2(c) for the WLS estimator of the original PLR model (2.6), viz., Assumptions WLS/PLR 1(e) and 2(c). This follows because with Y_{Tt} as defined above, we have

$$(5.16) \quad \|\sqrt{T}\bar{m}_T^*(\theta_T,\hat{\tau})\| = \|\sqrt{T}\int(\tau_{10}-\hat{\tau}_1)(\tau_{20}-\hat{\tau}_2)/\hat{\tau}_3 dP + \sqrt{T}\int(\hat{\tau}_2-\tau_{20})'\theta_T(\tau_{20}-\hat{\tau}_2)/\hat{\tau}_3 dP\|$$

and the right–hand side of (5.16) is bounded above by the same expression as in (3.31), where $\hat{\tau}_j$, τ_{j0}, and dP abbreviate $\hat{\tau}_j(z)$, $\tau_{j0}(z)$, and dP(z), respectively, for $j = 1, 2, 3$.

In the GMM/CMR example, the sequence of CMR models are defined under local alternatives to be such that θ_T satisfies

$$(5.17) \qquad E(\psi(Z_{Tt}, \theta_T)|X_t) = \underset{\sim}{0} \text{ a.s. } \forall t = 1, \ldots, T, \ T \geq 1$$

and W_{Tt} is defined by $(Z_{Tt}', X_t')'$. In this case, Assumption 2(c) with θ_0 replaced by θ_T holds trivially, as in (3.26), because

$$(5.18) \qquad \sqrt{T}\, \bar{m}_T^*(\theta_T, \tau) = \frac{1}{\sqrt{T}}\Sigma_1^T E\tau(X_t)\psi(Z_{Tt}, \theta_T) = \underset{\sim}{0} \ \forall\tau, \ \forall T \geq 1 .$$

Local power results are given in the following theorem.

THEOREM I.5: *Suppose Assumptions 2–ℓp, 3, and 4 or 2**–ℓp, 3, and 4 hold under* P *and* θ_0 *satisfies the null hypothesis. Then,*

(a) $W_T \xrightarrow{d} \chi_r^2(\delta^2)$ *and Assumption 2–ℓp or 2**–ℓp can be replaced by 2*–ℓp and 3* if need be,*

(b) $LM_{aT} \xrightarrow{d} \chi_r^2(\delta^2)$ *provided Assumption 5a holds,*

(c) $LR_{aT} \xrightarrow{d} \chi_r^2(\delta^2)$ *provided Assumptions 5a and 6a hold in place of 3,*

(d) $LM_{bT} \xrightarrow{d} \chi_r^2(\delta^2)$ *provided Assumptions 5b and 6b(i) also hold, and*

(e) $LR_{bT} \xrightarrow{d} \chi_r^2(\delta^2)$ *provided Assumptions 5b and 6b hold in place of 3,*

where r *is the number of restrictions,* $\delta^2 = \eta' H'(HVH')^{-1}H\eta$, *and* $\chi_r^2(\delta^2)$ *denotes the noncentral chi–square distribution with noncentrality parameter* δ^2 *and* r *degrees of freedom.*

COMMENTS: 1. Since $\sqrt{T}\, h(\theta_T) \to H\eta$, power approximations can be based on a $\chi_r^2(\delta_T^2)$ distribution, where $\delta_T^2 = Th(\theta_T)'(HVH')^{-1}h(\theta_T)$. In particular, to approximate the power of a test against an alternative θ when the sample size is T, we set $\theta = \theta_T$ and take $\delta_T^2 = Th(\theta)'(HVH')^{-1}h(\theta)$.

2. Due to the local nature of the alternatives in Theorem I.5, the approximations described in Comment 1 usually are more accurate for close alternatives to the null hypothesis than for distant alternatives.

3. Local power results, such as those of Theorem I.5, can be used to gauge and summarize the evidence provided by an hypothesis test in those cases where the test fails to reject the null hypothesis, see Andrews (1989c).

6. EXAMPLES

In this section, we illustrate how the results of Sections 2–5 can be applied in several examples. We do not, however, use the results above to give complete proofs of the asymptotic normality of the estimators considered. Rather, the purpose of this section is to show what the definitions and assumptions introduced above reduce to in particular examples. For brevity, we concentrate our discussion on two of the key assumptions used above to obtain asymptotic normality of MINPIN estimators, viz., the consistency of $\hat{\tau}$ (the second part of Assumption 2(b) or 2*(b)) and the asymptotic orthogonality condition (Assumption 2(c)). Sufficient conditions for the first part of Assumption 2(b) and Assumption 2(e) are discussed in ASEM:II for some of the examples.

The examples discussed in this section are examples (3)–(7) listed in the Introduction. These examples have been chosen because each illustrates a different feature of the results given in Sections 2–5. In addition, most of the examples concern estimators for which no proof of $\sqrt{T}$–consistency and asymptotic normality is currently available in the literature. None of the LM and LR tests discussed in the examples are considered elsewhere in the literature.

6.1. Regression with Unobserved Risk Variables

In contrast to the GMM/CMR and WLS/PLR examples discussed above, the model considered here is inherently a time series model. This model and the estimation procedure considered for it were first analyzed by Pagan and Ullah (1988). Pagan and Ullah did not, however, provide a proof of the $\sqrt{T}$–consistency and asymptotic normality of their estimator.

The model under consideration is a regression model with an unobserved regressor σ_t^2. σ_t^2 is a risk variable that equals the conditional variance of an observed rv ψ_t given an information set $\mathcal{F}_t$ (i.e. $\sigma_t^2 = E((\psi_t - E(\psi_t|\mathcal{F}_t))^2|\mathcal{F}_t))$. The model is

$$(6.1.1) \qquad Y_t = X_t'\theta_{10} + \sigma_t^2\theta_{20} + U_t \text{ for } t = 1, \ldots, T,$$

where Y_t and X_t are observed dependent and regressor variables respectively and U_t is an unobserved error. It is assumed that $E(U_t|\mathcal{F}_t) = 0$ a.s., X_t is $\mathcal{F}_t$–measurable, $E(\psi_t|\mathcal{F}_t) = E(\psi_t|V_{1t})$ a.s. for some $\mathcal{F}_t$–measurable random vector V_{1t}, and the σ–fields $\{\mathcal{F}_t\}$ are non–decreasing in t. For notational simplicity we take σ_t^2 to be a scalar and the remainder of the regression function to be linear in θ_{10}. Neither is necessary. See the reference above for descriptions of and references to various applications of this model.

The model (6.1.1) can be re–written as

$$(6.1.2) \qquad Y_t = X_t'\theta_{10} + \phi_t^2\theta_{20} + \tilde{U}_t = \tilde{X}_t'\theta_0 + \tilde{U}_t,$$

where $\phi_t = \psi_t - E(\psi_t|V_{1t})$, $\tilde{U}_t = U_t + (\sigma_t^2 - \phi_t^2)\theta_{20}$, $\tilde{X}_t = (X_t', \phi_t^2)'$, and $\theta_0 = (\theta_{10}', \theta_{20})'$. Note that $E(\tilde{U}_t|\mathcal{F}_t) = 0$ a.s., but $E\phi_t^2\tilde{U}_t$ does not equal zero in general.

Let $\hat{\tau}_1(\cdot)$ and $\hat{\tau}_2(\cdot)$ denote nonparametric estimators of

$$(6.1.3) \qquad \tau_{10}(\cdot) = E(\psi_t|V_{1t} = \cdot) \text{ and } \tau_{20}(\cdot) = E((\psi_t - E(\psi_t|V_{2t} = \cdot))^2|V_{2t} = \cdot),$$

respectively, where V_{2t} is some $\mathcal{F}_t$–measurable random vector. Pagan and Ullah suggest estimating θ_0 by applying an instrumental variables (IV) procedure to (6.1.2) in which the unobserved rv ϕ_t^2 is replaced by the proxy $\hat{\phi}_t^2 = (\psi_t - \hat{\tau}_1(V_{1t}))^2$ and the IV is given by $\hat{Z}_t = (X_t', \hat{\tau}_2(V_{2t}))'$. That is,

$$(6.1.4) \qquad \hat{\theta} = \left[\Sigma_1^T \hat{Z}_t \hat{X}_t'\right]^{-1} \Sigma_1^T \hat{Z}_t Y_t,$$

where $\hat{X}_t = (X_t', \hat{\phi}_t^2)'$.

The estimator $\hat{\theta}$ is a MINPIN estimator with $W_t = (Y_t, X_t', \psi_t, V_{1t}', V_{2t}')'$,

$$d(m,\gamma) = m'\gamma m/2\ ,\ m_t(\theta,\tau) = (Y_t - X_t'\theta_1 - (\psi_t - \tau_1(V_{1t}))^2\theta_2)\begin{bmatrix} X_t \\ \tau_2(V_{2t}) \end{bmatrix},$$

$$\hat{\gamma} = \tfrac{1}{T}\Sigma_1^T\begin{bmatrix} X_t \\ \hat{\tau}_2(V_{2t}) \end{bmatrix}\begin{bmatrix} X_t \\ \hat{\tau}_2(V_{2t}) \end{bmatrix}', \tag{6.1.5}$$

$\tau = (\tau_1, \tau_2)'$, and $\theta = (\theta_1', \theta_2)'$.[11] Theorem I.2 above can be used to establish the $\sqrt{T}$–consistency and asymptotic normality of $\hat{\theta}$ by verifying Assumption 2.

When Assumption 2 holds, the asymptotic covariance matrix V of $\hat{\theta}$ is

$$V = M^{-1}S(M^{-1})'\ ,\ \text{where}$$

$$S = \lim_{T\to\infty}\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} E\tilde{U}_s\tilde{U}_t\tilde{Z}_s\tilde{Z}_t'\ ,\ M = \lim_{T\to\infty}\tfrac{1}{T}\Sigma_1^T E\tilde{Z}_t\tilde{X}_t'\ ,\ \text{and} \tag{6.1.6}$$

$$\tilde{Z}_t = (X_t', \tau_{20}(V_{2t}))' .$$

Assumption 2(c) is satisfied in this model only if $\hat{\tau}_1$ is consistent for τ_{10} at rate $T^{1/4}$. We have

$$\|\sqrt{T}\,\bar{m}_T^*(\theta_0, \hat{\tau})\|$$

$$= \left\|\tfrac{1}{T}\Sigma_1^T\int[T^{1/4}(\hat{\tau}_1(v_1) - \tau_{10}(v_1))]^2(x', \hat{\tau}_2(v_2))'dP_t(x, v_1, v_2)\right\| \tag{6.1.7}$$

$$\le \tfrac{1}{T}\Sigma_1^T\int[T^{1/4}(\hat{\tau}_1(v_1) - \tau_{10}(v_1))]^2\|(x', \hat{\tau}_2(v_2))\|dP_t(x, v_1, v_2)\ ,$$

where $P_t(\cdot,\cdot,\cdot)$ denotes the distribution of $(X_t', V_{1t}', V_{2t}')'$. If the right–hand side of (6.1.7) converges in probability to zero, then Assumption 2(c) holds. For example, if X_t is bounded, $\hat{\tau}_2(V_{2t})$ is bounded with probability $\to$ 1, and $\{V_{1t} : t \ge 1\}$ are identically distributed, then Assumption 2(c) holds if

$$T^{1/4}\left[\int(\hat{\tau}_1(v_1) - \tau_{10}(v_1))^2dP(v_1)\right]^{1/2} \xrightarrow{p} 0 \ \text{as}\ T\to\infty\ , \tag{6.1.8}$$

where $P(\cdot)$ denotes the distribution of V_{1t} .

When Assumption 2(e) is verified using Theorem II.7 of ASEM:II, which allows for strong mixing and near–epoch dependent rv's, the metric $\rho_T(\cdot,\cdot)$ is such that the second part of Assumption 2(b) holds provided

$$\int_{\mathcal{V}_1}(\hat{\tau}_1(v_1)-\tau_{10}(v_1))^2 dv_1 \xrightarrow{p} 0 \text{ and}$$

(6.1.9)

$$\int_{\mathcal{V}_2}(\hat{\tau}_2(v_2)-\tau_{20}(v_2))^2 dv_2 \xrightarrow{p} 0 ,$$

where V_{1t} and V_{2t} are bounded rv's that take values in $\mathcal{V}_1$ and $\mathcal{V}_2$ respectively.

The W, LM, and LR tests for testing the restrictions $H_0 : h(\theta_0) = \underset{\sim}{0}$ are applicable in this example. For the LR test, however, the following additional conditions are needed:

(6.1.10)

(i) $E(\tilde{U}_t^2 | \tilde{Z}_t) = \sigma^2$ a.s. for some constant $\sigma^2 > 0$ and

(ii) $\{\tilde{U}_t \tilde{Z}_t : t \geq 1\}$ is uncorrelated.

Under these conditions, Assumption 6a holds with $b = \sigma^2$ and the LR test statistic is given by

(6.1.11) $$LR_{aT} = T[\bar{m}_T(\tilde{\theta}_a, \hat{\tau})' \hat{\gamma} \bar{m}_T(\tilde{\theta}_a, \hat{\tau}) - \bar{m}_T(\hat{\theta},\hat{\tau})' \hat{\gamma} \bar{m}_T(\hat{\theta},\hat{\tau})]/\hat{\sigma}^2 ,$$

where $\tilde{\theta}_a$ is a restricted estimator of θ_0 that minimizes $d(\bar{m}_T(\theta,\hat{\tau}), \hat{\gamma})$ over Θ subject to $h(\theta) = \underset{\sim}{0}$ and $\hat{\sigma}^2$ is some estimator of σ^2 that is consistent under the null.

6.2. Weighted Censored LAD Estimation of the Censored Regression Model under Conditional Median Restrictions

This example makes use of Assumption 2* rather than Assumption 2, because the criterion function considered is not everywhere differentiable in θ.

The $\sqrt{T}$–consistency and asymptotic normality of the weighted censored least absolute deviations (WC–LAD) estimator considered here has not been established in the literature. Such results have been established, however, for a one–step version of the WC–LAD estimator, see Newey and Powell (1990). The one–step estimator is also a

MINPIN estimator, but for brevity we do not discuss it here. We note that the method of this paper allows one to establish asymptotic normality of the WC–LAD estimator and one-step versions of it without sample splitting and for the case of temporally dependent rv's.

The model is

$$(6.2.1) \qquad Y_t = \min\{X_t'\theta_0 + U_t, C_t\},\ t = 1, 2, \ldots, T,$$

where Y_t is an observed dependent variable, X_t is an observed p–vector of regressors, C_t is an observed censoring point, and U_t is an unobserved error with conditional median zero given $Z_t = (X_t', C_t)'$. For simplicity, $\{(Y_t, X_t, U_t, C_t) : t = 1, 2, \ldots\}$ is assumed to be identically distributed.

Suppose U_t has an absolutely continuous distribution function in a neighborhood of 0 conditional on Z_t. Denote its value at 0 by $\tau_0(Z_t)$. Let $\hat{\tau}(Z_t)$ be an estimator of $\tau_0(Z_t)$. A weighted censored LAD estimator $\hat{\theta}$ of θ_0 minimizes

$$(6.2.2) \qquad \Sigma_1^T 2\hat{\tau}(Z_t)\,|Y_t - \min\{X_t'\theta, C_t\}|$$

over $\theta \in \Theta \subset R^p$. Under suitable assumptions, it also solves the first order conditions

$$(6.2.3) \qquad \underset{\sim}{0} = \Sigma_1^T 2\hat{\tau}(Z_t)1(X_t'\hat{\theta} < C_t)\mathrm{sgn}(Y_t - X_t'\hat{\theta})X_t$$

with probability $\to 1$. Thus, $\hat{\theta}$ is a MINPIN estimator with

$$(6.2.4) \qquad \begin{aligned} &d(m,\gamma) = m'm/2 \quad \text{and} \\ &m_t(\theta,\tau) = 2\tau(Z_t)1(X_t'\theta < C_t)\mathrm{sgn}(Y_t - X_t'\theta)X_t . \end{aligned}$$

Note that $m_t(\theta,\tau)$ is not everywhere differentiable in θ.

When Assumption 2* holds, the asymptotic covariance matrix V of $\hat{\theta}$ is

$$(6.2.5) \qquad \begin{aligned} &V = M^{-1}SM^{-1}, \text{ where } M = -4E\tau_0^2(Z_t)1(X_t'\theta_0 < C_t)X_tX_t' \text{ and} \\ &S = \lim_{T\to\infty} \frac{4}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} \tau_0(Z_s)1(X_s'\theta_0 < C_s)\mathrm{sgn}(U_s)\tau_0(Z_t)1(X_t'\theta_0 < C_t)\mathrm{sgn}(U_t)X_sX_t' . \end{aligned}$$

When the observations are independent or the error U_t has conditional median zero given

$(X_t, C_t, X_s, C_s, U_s)$ $\forall s \neq t$, the asymptotic covariance matrix of $\hat{\theta}$ simplifies to

$$V = \left[E4\tau_0^2(X_t)1(X_t'\theta_0 < C_t)X_tX_t'\right]^{-1} . \tag{6.2.6}$$

In consequence, $\hat{\theta}$ attains the semiparametric asymptotic efficiency bound for the model (6.2.1), see Newey and Powell (1990), when the observations are iid.

In this example, the second part of Assumption 2*(a) holds if (6.2.3) holds. Assumption 2*(c) holds trivially for any estimator $\hat{\tau}$ of τ_0 by (3.26)–(3.27), because

$$\sqrt{T}\,\bar{m}_T^*(\theta_0, \tau) = \sqrt{T}\,E\tau(Z_t)1(X_t'\theta_0 < C_t)\mathrm{sgn}(U_t)X_t = \underset{\sim}{0} , \tag{6.2.7}$$

since $E(\mathrm{sgn}(U_t)\,|\,Z_t) = 0$ a.s. by the conditional median zero assumption on U_t .

When pseudo–metric of (3.32) is used in verifying Assumption 2*(e), we have

$$\begin{aligned} &\rho^2_{\Theta\times\mathcal{T}}((\theta,\tau), (\theta_0, \tau_0)) = E\|m_t(\theta,\tau) - m_t(\theta_0, \tau_0)\|^2 \\ &= 4\int[\tau(x,c)1(x'\theta < c)\mathrm{sgn}(u-x'(\theta-\theta_0))-\tau_0(x,c)1(x'\theta_0<c)\mathrm{sgn}(u)]^2\|x\|^2dP(x,c,u) , \end{aligned} \tag{6.2.8}$$

where $P(\cdot,\cdot,\cdot)$ denotes the distribution of (X_t, C_t, U_t) . With this pseudo-metric, $(\hat{\theta},\hat{\tau}) \xrightarrow{p} (\theta_0, \tau_0)$ (i.e., the second part of Assumption 2*(b) holds), if

$$\begin{aligned} &\text{(a)}\ \int(\hat{\tau}(x,c) - \tau_0(x,c))^2\|x\|^2dP(x,c) \xrightarrow{p} 0 , \\ &\text{(b)}\ \hat{\theta} - \theta_0 \xrightarrow{p} \underset{\sim}{0} ,\ \text{and} \\ &\text{(c)}\ P(X_t'\theta_0 = C_t) = 0 , \end{aligned} \tag{6.2.9}$$

where $P(\cdot,\cdot)$ denotes the distribution of (X_t, C_t) . Part (b) just requires the standard sort of consistency for a finite dimensional estimator and can be established by Theorem I.1 above.

The LM and LR tests do not apply in this example because neither Assumption 2 nor 2** holds. The Wald test, however, does apply.

This example can be extended to the case where the errors in the model (6.2.1) have some constant conditional quantile other than the median. For brevity, we do not discuss this extension.

6.3. MAD–DUC Estimation of Index Regression Models

This example exhibits the case where the preliminary infinite dimensional nuisance parameter estimator $\hat{\tau}$ is a function of the finite dimensional parameter θ. This complicates the way in which MAD–DUC estimators are defined as MINPIN estimators, but still allows the results of Sections 2–5 above to be applied to them.

Examples of MAD–DUC estimators include Ichimura's (1985), Ichimura and Lee's (1990), and Klein and Spady's (1987) estimators of single index, multiple index, and binary choice models respectively. Proofs of the $\sqrt{T}$–consistency and asymptotic normality of these estimators are already available in the literature (for the case of independent observations). Other estimators, for which $\sqrt{T}$–consistency and asymptotic normality results are not available in the literature, are also included in the MAD–DUC class described below. The LM and LR test statistics discussed below have not been considered elsewhere in the literature.

The model is

$$(6.3.1) \qquad Y_t = \varphi(h(X_t, \theta_0)) + U_t \ \text{ for } \ t = 1, \ldots, T,$$

where $E(U_t | X_t) = 0$ a.s., $\{W_t\} = \{(Y_t, X_t')' : t \geq 1\}$ are identically distributed, the multiple index function $h(\cdot,\cdot) : R^L \times R^p \to R^{k_a}$ is known, the transformation $\varphi(\cdot) : R^{k_a} \to R$ is unknown, and the distribution of U_t is unspecified. Additional assumptions are needed to identify the parameter θ_0. These assumptions vary from one index model to the next, so we do not specify such assumptions here. Numerous econometric models with latent errors of unspecified distribution are of the index regression form including censored and truncated regression and qualitative choice models (although there is a loss of information in many such models if they are viewed solely as index regression models). In the binary choice model, for example, the function $\varphi(\cdot)$ is given by $\varphi(v) = P(Y_t = 1 | h(X_t, \theta_0) = v)$.

MAD–DUC estimators of θ_0 Minimize the Average "Distance" between the Dependent variable Unconditional and Conditional on the index: $\hat{\theta}$ is defined to minimize

(6.3.2) $\qquad \frac{1}{T}\Sigma_1^T \eta(Y_t, \hat{\tau}(\theta, h(X_t, \theta)))$

over $\theta \in \Theta$. Here, $\hat{\tau}(\cdot,\cdot)$ is an estimator of $\tau_0(\cdot,\cdot)$, $\tau_0(\cdot,\cdot)$ is defined by

(6.3.3) $\qquad \tau_0(\theta,v) = E(Y_t \mid h(X_t, \theta) = v)$ for $v \in R^{k_a}$,

and $\eta(\cdot,\cdot)$ is some "distance" function. Note that $\tau_0(\theta_0, \cdot) = \varphi(\cdot)$. Ichimura's (1985) and Ichimura and Lee's (1990) estimators take

(6.3.4) $\qquad \eta(Y_t, \tau) = (Y_t - \tau)^2/2$.

Klein and Spady's (1987) estimator takes

(6.3.5) $\qquad \eta(Y_t, \tau) = -Y_t \ln \tau - (1 - Y_t)\ln(1-\tau)$

and applies in the binary choice model in which Y_t equals 0 or 1 and $k_a = 1$. Numerous other functions $\eta(\cdot,\cdot)$ can be considered.

The estimator $\hat{\tau}(\cdot,\cdot)$ consists of a family of univariate nonparametric regression estimators of Y_t on $h(X_t, \theta)$ – one for each value of $\theta \in \Theta \subset R^p$. In each of the papers referred to above, kernel estimators are used. Other nonparametric methods also could be used.

MAD–DUC estimators are MINPIN estimators with

(6.3.6) $\qquad d(m,\gamma) = m'm/2$ and $m_t(\theta,\tau) = \eta'(Y_t, \tau(\theta, h(X_t, \theta)))\frac{\partial}{\partial\theta}\tau(\theta, h(X_t, \theta))$,

where $\eta'(\cdot,\cdot)$ denotes the derivative of $\eta(\cdot,\cdot)$ with respect to its second argument.

When Assumption 2 holds, the asymptotic covariance matrix V of a MAD–DUC estimator is given by

$$
(6.3.7) \quad
\begin{aligned}
&V = M^{-1}SM^{-1}\ , \text{ where} \\
&S = \lim_{T\to\infty} \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} E\eta'(Y_s, \varphi_s)\eta'(Y_t, \varphi_t)H_sH_t'\ , \\
&M = E\eta''(Y_t, \varphi_t)H_tH_t',\ \varphi_t = \varphi(h(X_t, \theta_0))\ , \\
&H_t = \left[\frac{\partial}{\partial\theta'}h(X_t, \theta_0) - E\left[\frac{\partial}{\partial\theta'}h(X_t, \theta_0) \,|\, h(X_t, \theta_0)\right]\right]' \frac{\partial}{\partial h}\varphi(h(X_t, \theta_0))\ ,
\end{aligned}
$$

and $\eta''(y,\varphi)$ denotes $d^2\eta(y,\varphi)/d\varphi^2$. The expression for M uses the assumption that $E(\eta'(Y_t, \varphi_t)|X_t) = 0$ a.s., which usually follows from the conditions on $\eta(\cdot,\cdot)$ needed for consistency of $\hat{\theta}$. This assumption holds for $\eta(\cdot,\cdot)$ as defined in (6.3.4) and (6.3.5). The vector H_t arises in the expressions above because it can be shown to equal $\frac{\partial}{\partial\theta}\tau_0(\theta_0, h(X_t, \theta_0))$. When estimating V , the appropriate sample analogue of H_t to use is $\hat{\tau}^{(1)}(\hat{\theta}, h(X_t, \hat{\theta})) + \left[\frac{\partial}{\partial\theta'}h(X_t, \hat{\theta})\right]'\hat{\tau}^{(2)}(\hat{\theta}, h(X_t, \hat{\theta}))$, where $\hat{\tau}^{(j)}$ denotes the derivative of $\hat{\tau}$ with respect to its j–th argument.

If $\{W_t\}$ is an independent sequence of rv's and

$$
(6.3.8) \qquad E(\eta'(Y_t, \varphi_t)^2|h(X_t, \theta_0)) = cE(\eta''(Y_t, \varphi_t)|h(X_t, \theta_0)) \text{ a.s.},
$$

for some scalar $c \neq 0$, then Assumption 6b holds with $\rho_{Tt}(W_t, \theta, \tau) = \eta(Y_t, \tau(\theta, h(X_t, \theta)))$ and the LR_b statistic for testing $H_0 : h(\theta_0) = \underset{\sim}{0}$ is asymptotically chi–square under the null hypothesis. In addition, the W , LM_a , and LM_b statistics simplify as in (5.14). With Ichimura's and Ichimura and Lee's estimators, (6.3.8) holds with $c = -\sigma^2$ if $E(U_t^2|h(X_t, \theta_0)) = \sigma^2$ a.s. The latter does not hold, however, in most index models. In the binary choice model, (6.3.8) holds for Klein and Spady's estimator with $c = 1$. In this case, the LR_b statistic is given by

$$
(6.3.9) \quad
\begin{aligned}
LR_{bT} = &-2\Sigma_1^T\left[Y_t \ln \hat{\tau}(\tilde{\theta}_b, h(X_t, \tilde{\theta}_b)) + (1 - Y_t)\ln(1 - \hat{\tau}(\tilde{\theta}_b, h(X_t, \tilde{\theta}_b)))\right] \\
&+ 2\Sigma_1^T\left[Y_t \ln \hat{\tau}(\hat{\theta}, h(X_t, \hat{\theta})) + (1 - Y_t)\ln(1 - \hat{\tau}(\hat{\theta}, h(X_t, \hat{\theta})))\right]\ ,
\end{aligned}
$$

where $\tilde{\theta}_b$ is the restricted MINPIN estimator defined in Section 5. Note also that Klein and Spady's estimator attains the semiparametric asymptotic efficiency bound when $\{W_t\}$ are independent.

The asymptotic orthogonality condition, Assumption 2(c), holds for MAD–DUC estimators if

$$\begin{aligned}
&\left\|\sqrt{T}\,\frac{\partial}{\partial m}d(\bar{m}_T^*(\theta_0,\hat{\tau}),\hat{\gamma})\right\| \\
&= \left\|\sqrt{T}\,E\eta'(\varphi_t,\tau(\theta_0,h(X_t,\theta_0)))\left[\frac{\partial}{\partial\theta}\tau(\theta_0,h(X_t,\theta_0))-\frac{\partial}{\partial\theta}\tau_0(\theta_0,h(X_t,\theta_0))\right]\right\|\Bigg|_{\tau=\hat{\tau}} \\
&\le T^{\delta_1}\left[\int\eta'(\varphi(v),\hat{\tau}(\theta_0,v))^r dP(v)\right]^{1/r}\left[T^{\delta_2}\left[\int\|\hat{\tau}^{(1)}(\theta_0,v)-\tau_0^{(1)}(\theta_0,v)\|^s dP(v)\right]^{1/s}\right. \\
&\qquad\left.+\,T^{\delta_2}\left[\int\|z\|^s\|\hat{\tau}^{(2)}(\theta_0,v)-\tau_0^{(2)}(\theta_0,v)\|^s d\tilde{P}(v,z)\right]^{1/s}\right] \\
&\xrightarrow{p} 0\,,
\end{aligned} \tag{6.3.10}$$

where $\delta_1+\delta_2=1/2$, $r\ge 1$, $s\ge 1$, $1/r+1/s=1$, $P(\cdot)$ denotes the distribution of $h(X_t,\theta_0)$, $\tilde{P}(\cdot,\cdot)$ denotes the distribution of $\left[h(X_t,\theta_0),\frac{\partial}{\partial\theta'}h(X_t,\theta_0)\right]$ and $\hat{\tau}^{(j)}$ and $\tau_0^{(j)}$ denote the derivatives of $\hat{\tau}$ and τ_0 with respect to their j–th arguments. The equality in (6.3.10) holds provided $E\eta'(Y_t,\tau(\theta_0,h(X_t,\theta_0))|X_t) = E\eta'(\varphi_t,\tau(\theta_0,h(X_t,\theta_0))|X_t)$ a.s., using the fact that $E\left[\frac{\partial}{\partial\theta}\tau_0(\theta_0,h(X_t,\theta_0))\,|\,h(X_t,\theta_0)=v\right]=E(H_t\,|\,h(X_t,\theta_0)=v)=\underset{\sim}{0}\ \ \forall v$. The former condition holds, for example, with Ichimura's, Ichimura and Lee's, and Klein and Spady's estimators.

For Ichimura's and Ichimura and Lee's estimators, sufficient conditions for (6.3.10) and Assumption 2(c) are

$$\begin{aligned}
&T^{\delta_1}\left[\int(\hat{\tau}(\theta_0,v)-\tau_0(\theta_0,v))^2 dP(v)\right]^{1/2}\xrightarrow{p} 0\,, \\
&T^{\delta_2}\left[\int\|\hat{\tau}^{(1)}(\theta_0,v)-\tau_0^{(1)}(\theta_0,v)\|^2 dP(v)\right]^{1/2}\xrightarrow{p} 0\,, \\
&T^{\delta_2}\left[\int\|\hat{\tau}^{(2)}(\theta_0,v)-\tau_0^{(2)}(\theta_0,v)\|^4 dP(v)\right]^{1/4}\xrightarrow{p} 0\,,\ \text{and}\ E\left\|\frac{\partial}{\partial\theta'}h(X_t,\theta_0)\right\|^4<\infty\,.
\end{aligned} \tag{6.3.11}$$

For Klein and Spady's estimator, sufficient conditions for (6.3.10) and Assumption 2(c) are

(6.3.11) plus

$$\inf_{v} \min\left\{|\hat{\tau}(\theta_0, v)|, |1-\hat{\tau}(\theta_0, v)|\right\} \geq \epsilon \text{ with probability} \to 1 \text{ for some } \epsilon > 0 , \tag{6.3.12}$$

where the inf is taken over all v in the support of $h(X_t, \theta_0)$.

For ρ_T as in (3.23), the second part of Assumption 2(b) holds if

$$\begin{aligned}
\rho_T(\hat{\tau},\tau_0) \leq & \left[\int \eta'(u + \varphi(v), \hat{\tau}(\theta_0,v))^2 \|\hat{\tau}^{(1)}(\theta_0,v) - \tau_0^{(1)}(\theta_0,v)\|^2 dP(u,v)\right]^{1/2} \\
& + \left[\int \eta'(u + \varphi(v), \hat{\tau}(\theta_0,v))^2 \|z\|^2 \|\hat{\tau}^{(2)}(\theta_0,v) - \tau_0^{(2)}(\theta_0,v)\|^2 dP(u,v,z)\right]^{1/2} \\
& + \left[\int [\eta'(u + \varphi(v), \hat{\tau}(\theta_0,v)) - \eta'(u + \varphi(v), \tau_0(\theta_0,v))]^2 \|(\tau_0^{(1)}(\theta_0,v) \right. \\
& \left. + z'\tau_0^{(2)}(\theta_0,v)\|^2 dP(u,v,z)\right]^{1/2} \\
& \xrightarrow{P} 0,
\end{aligned} \tag{6.3.13}$$

where $P(\cdot,\cdot)$ denotes the distribution of $(U_t, h(X_t, \theta_0))$ and $P(\cdot,\cdot,\cdot)$ denotes the distribution of $\left[U_t, h(X_t, \theta_0), \frac{\partial}{\partial\theta'}h(X_t, \theta_0)\right]$.

For Ichimura's and Ichimura and Lee's estimators, sufficient conditions for (6.3.13) are

$$\begin{aligned}
& \int [\hat{\tau}(\theta_0, v) - \tau_0(\theta_0, v)]^4 dP(v) \xrightarrow{P} 0 , \quad \int \|\hat{\tau}^{(1)}(\theta_0, v) - \tau_0^{(1)}(\theta_0, v)\|^4 dP(v) \xrightarrow{P} 0 , \\
& \int \|\hat{\tau}^{(2)}(\theta_0, v) - \tau_0^{(2)}(\theta_0, v)\|^8 dP(v) \xrightarrow{P} 0 , \quad EU_t^4 < \infty , \\
& E\|\tau_0^{(1)}(\theta_0,h(X_t,\theta_0))\|^4 < \infty, \ E\|\tau_0^{(2)}(\theta_0,h(X_t,\theta_0))\|^8 < \infty, \text{ and } E\left\|\frac{\partial}{\partial\theta'}h(X_t,\theta_0)\right\|^8 < \infty.
\end{aligned} \tag{6.3.14}$$

The same conditions plus (6.3.12) suffice for Klein and Spady's estimator. In this case, $EU_t^4 < \infty$ holds trivially, since $|U_t|$ is a bounded rv.

6.4. Three-step Estimation of Sample Selection Models

This example illustrates the case where a preliminary finite dimensional nuisance parameter estimator appears that is not asymptotically orthogonal to the estimator of interest. In this case, Assumption 2(c) fails. It is shown how the definition of the MINPIN estimator can be adjusted to circumvent this problem and allow the results of Sections 2–5 to be applied.

The three–step sample selection estimator considered here has not been considered elsewhere in the literature. It is quite similar, however, to the two–step estimators of Powell (1987) and Newey (1988). The difference is that it uses a WLS estimator that takes account of the heteroskedasticity that necessarily arises in the equation that is estimated, whereas the two–step estimators of Powell and Newey use the ordinary LS estimator. Some efficiency gains should result from the use of the WLS estimator.

The model is given by

$$\tag{6.4.1} \begin{aligned} &\tilde{Y}_t = \tilde{X}_t'\beta_0 + U_t \text{ and } D_t = 1(h(Z_t, \alpha_0) + \epsilon_t > 0)\,, \text{ where} \\ &(Y_t, D_t, X_t, Z_t) = (\tilde{Y}_t D_t, D_t, \tilde{X}_t D_t, Z_t) \text{ are observed for } t = 1, \ldots, T\,, \end{aligned}$$

$\tilde{Y}_t$ 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 identically distributed, and (U_t, ϵ_t) is independent of $(\tilde{X}_t, Z_t)$ and has unknown distribution. Additional assumptions must be imposed to identify α_0 and β_0, e.g., see Powell (1987) and Newey (1988). Such assumptions allow $\tilde{X}_t$ and Z_t to have elements in common. The first equation of model (6.4.1) multiplied by D_t can be re–written as

$$\tag{6.4.2} \begin{aligned} &Y_t = X_t'\beta_0 + D_t g(h(Z_t, \alpha_0)) + \mu_t\,, \text{ where} \\ &g(v) = E(U_t \mid \epsilon_t > -v),\ \mu_t = D_t(U_t - g(h(Z_t, \alpha_0)))\,, \text{ and} \\ &E(\mu_t \mid D_t = 1, X_t, Z_t) = 0 \text{ a.s.} \end{aligned}$$

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

The three–step estimator $\hat{\beta}$ of β_0 is defined as follows. First, one purges Y_t and X_t of their correlation with $h(Z_t, \alpha_0)$ by subtracting from Y_t and X_t nonparametric estimates of their conditional expectations given $h(X_t, \alpha_0)$ and $D_t = 1$, where a preliminary estimator $\hat{\alpha}$ of α_0 is used in computing the nonparametric estimates. Second, one weights the transformed Y_t and X_t variables using a nonparametric estimate of the conditional variance of μ_t given $h(Z_t, \alpha_0)$ and $D_t = 1$. Third, one regresses the weighted and transformed Y_t variables on the weighted and transformed X_t variables to obtain a WLS estimator $\hat{\beta}$ of β_0.

As it happens, the asymptotic distribution of the three-step estimator $\hat{\beta}$ depends on the choice of preliminary estimator $\hat{\alpha}$. Thus, if one takes θ_0 to equal β_0 and one takes α_0 to be part of the nuisance parameter τ_0, then the orthogonality condition 2(c) fails to hold. This problem can be circumvented by defining θ_0 to equal $(\alpha_0', \beta_0')'$ and $\hat{\theta}$ to equal $(\hat{\alpha}', \hat{\beta}')'$, i.e. by treating the nuisance parameter α_0 as though it is part of the parameter of interest. This same method of avoiding failures of the orthogonality condition can be applied in other examples as well.

To compute the three-step estimator $\hat{\beta}$, one needs estimates of $\tau_{j0}(\alpha_0, h(Z_t, \alpha_0))$ for $t = 1, \ldots, T$ and $j = 1, 2, 3$, where

$$\begin{aligned} &\tau_{10}(\alpha,v) = E(Y_t \mid h(Z_t, \alpha) = v, D_t = 1), \\ (6.4.3) \qquad &\tau_{20}(\alpha,v) = E(X_t \mid h(Z_t, \alpha) = v, D_t = 1), \text{ and} \\ &\tau_{30}(\alpha,v) = \mathrm{Var}(Y_t - X_t'\beta_0 \mid h(Z_t,\alpha) = v, D_t = 1). \end{aligned}$$

$\tau_{30}(\alpha,v)$ is defined such that $\tau_{30}(\alpha_0, v) = E(\mu_t^2 \mid h(Z_t, \alpha_0) = v, D_t = 1)$. As an alternative to the definition of (6.4.3), $\tau_{30}(\alpha,v)$ could be defined to be $E(\mu_t^2 \mid h(Z_t, \alpha) = v, D_t = 1)$. The definition of $\tau_{30}(\alpha,v)$ in (6.4.3), however, has the advantage that $\tau_{30}(\alpha,v)$ can be estimated via a nonparametric regression with variables that are constructed using preliminary estimators of *finite* dimensional parameters rather than *infinite* dimensional parameters.

Let $\hat{\tau}_j(\cdot,\cdot)$ denote an estimator of $\tau_{j0}(\cdot,\cdot)$ for $j = 1, 2, 3$. In practice, one only needs to compute $\hat{\tau}_j(\hat{\alpha},\cdot)$ for $j = 1, 2, 3$. $\hat{\tau}_1(\hat{\alpha},\cdot)$ and $\hat{\tau}_2(\hat{\alpha},\cdot)$ are obtained by nonparametric regressions of Y_t and X_t on $h(Z_t, \hat{\alpha})$ using the observations for which $D_t = 1$. Note that $\tau_{30}(\alpha,v) = E((Y_t - X_t'\beta_0)^2 \mid h(Z_t,\alpha) = v, D_t = 1) - (\tau_{10}(\alpha,v) - \tau_{20}(\alpha,v)'\beta_0)^2$.

Thus, one can take $\hat{\tau}_3(\hat{\alpha},\cdot) = \hat{\tau}_{3a}(\hat{\alpha},\cdot) - (\hat{\tau}_1(\hat{\alpha},\cdot) - \hat{\tau}_2(\hat{\alpha},\cdot)'\tilde{\beta})^2$, where $\hat{\tau}_{3a}(\hat{\alpha},\cdot)$ is obtained by a nonparametric regression of $(Y_t - X_t'\tilde{\beta})^2$ on $h(Z_t, \hat{\alpha})$ using the observations for which $D_t = 1$ and $\tilde{\beta}$ is some consistent estimator of β_0. The most convenient choice for $\tilde{\beta}$ is just the two–step estimator of β_0, which is the LS estimator of β_0 from the regression of $Y_t - \hat{\tau}_1(\hat{\alpha}, h(Z_t, \hat{\alpha}))$ on $X_t - \hat{\tau}_2(\hat{\alpha}, h(Z_t, \hat{\alpha}))$.

Let the estimator $\hat{\alpha}$ used above be a semiparametric MINPIN estimator of α_0 based on the second equation of (6.4.1). Suppose $\hat{\alpha}$ satisfies Assumption 2 and has defining functions $d(m,\gamma) = m'm/2$ and $m_t(\theta,\tau) = m_{1t}(\alpha, \tau_4)$, where τ_4 denotes some nuisance parameter that enters in the estimation of α_0. For example, $\hat{\alpha}$ could be Klein and Spady's (1987) or Ichimura's (1985) semiparametric estimator for the binary choice model (see Example 6.3).

The three–step estimator $\hat{\beta}$ of β_0 is obtained by regressing

$$(6.4.4)\qquad \begin{aligned} &(Y_t - \hat{\tau}_1(\hat{\alpha}, h(Z_t, \hat{\alpha})))/\hat{\tau}_3^{1/2}(\hat{\alpha}, h(Z_t, \hat{\alpha})) \text{ on} \\ &(X_t - \hat{\tau}_2(\hat{\alpha}, h(Z_t, \hat{\alpha})))/\hat{\tau}_3^{1/2}(\hat{\alpha}, h(Z_t, \hat{\alpha})) \end{aligned}$$

using the observations where $D_t = 1$. Letting $\hat{\theta} = (\hat{\alpha}',\hat{\beta}')'$, we see that $\hat{\theta}$ is a MINPIN estimator of $\theta_0 = (\alpha_0',\beta_0')'$ with $W_t = (Y_t,D_t,X_t',Z_t')'$, $\tau = (\tau_1,\tau_2',\tau_3,\tau_4')'$, $d(m,\gamma) = m'm/2$, and

$$(6.4.5)\qquad \begin{aligned} m_t(\theta,\tau) &= \begin{bmatrix} m_{1t}(\alpha, \tau_4) \\ m_{2t}(\theta,\tau) \end{bmatrix}, \text{ where} \\ m_{2t}(\theta,\tau) &= D_t[Y_t - \tau_1(\alpha,h(Z_t, \alpha)) - (X_t - \tau_2(\alpha, h(Z_t,\alpha)))'\beta] \\ &\qquad \times [X_t - \tau_2(\alpha, h(Z_t,\alpha))]/\tau_3(\alpha, h(Z_t, \alpha)) . \end{aligned}$$

When Assumption 2 holds and the observations $\{W_t : t \geq 1\}$ are temporally independent, the asymptotic covariance matrix V of $\hat{\theta}$ is given by

$$V = M^{-1}S(M^{-1})' = \begin{bmatrix} M_1^{-1}S_1M_1^{-1} & -M_1^{-1}S_1M_1^{-1}M_3'M_4^{-1} \\ -M_4^{-1}M_3M_1^{-1}S_1M_1^{-1} & M_4^{-1}(S_2 + M_3M_1^{-1}S_1M_1^{-1}M_3')M_4^{-1} \end{bmatrix},$$

where $S = \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}$, $M = \begin{bmatrix} M_1 & 0 \\ M_3 & M_4 \end{bmatrix}$,

(6.4.6)
$$\begin{aligned} S_1 &= Em_{1t}(\alpha_0, \tau_{40})m_{1t}(\alpha_0, \tau_{40})' , \\ S_2 &= ED_t[X_t - \tau_{20}(\alpha_0, h(Z_t,\alpha_0))][X_t - \tau_{20}(\alpha_0, h(Z_t,\alpha_0))]/E(\mu_t^2 | h(Z_t,\alpha_0), D_t{=}1) , \\ M_1 &= E\frac{\partial}{\partial\alpha'}m_{1t}(\alpha_0, \tau_{40}) , \\ M_3 &= E\,\frac{\partial}{\partial\alpha'}m_{2t}(\theta_0,\tau_0) = ED_t[X_t - \tau_{20}(\alpha_0,h(Z_t,\alpha_0))][\beta_0'\tau_{20}^{(2)}(\alpha_0, h(Z_t,\alpha_0)) \\ &\quad - \tau_{10}^{(2)}(\alpha_0, h(Z_t,\alpha_0))]\frac{\partial}{\partial\alpha'}h(Z_t,\alpha_0)/E(\mu_t^2 | h(Z_t, \alpha_0), D_t = 1) , \text{ and} \\ M_4 &= E\frac{\partial}{\partial\beta'}m_{2t}(\theta_0,\tau_0) = -S_2 . \end{aligned}$$

Here, $\tau_{10}^{(2)}(\cdot,\cdot)$ and $\tau_{20}^{(2)}(\cdot,\cdot)$ denote the derivatives of $\tau_{10}(\cdot,\cdot)$ and $\tau_{20}(\cdot,\cdot)$, respectively, with respect to their (scalar) second argument. The expression given for S_2 uses the fact that $E(\mu_t^2 | h(Z_t, \alpha_0), X_t, D_t = 1) = E(\mu_t^2 | h(Z_t, \alpha_0), D_t = 1)$ a.s. An estimator of V can be constructed by taking sample analogues of the quantities above.

For the case of temporally dependent observations, the asymptotic covariance matrix V of $\hat{\theta}$ is as in (6.4.6) except that S is given by $\lim_{T\to\infty} \mathrm{Var}\left[\frac{1}{\sqrt{T}}\Sigma_1^T m_t(\theta_0, \tau_0)\right]$.

Note that Powell's (1987) and Newey's (1988) two–step estimators equal the three-step estimator $\hat{\beta}$ when the latter is defined with $\hat{\tau}_3(\hat{\alpha}, h(Z_t, \hat{\alpha})) = 1$. Powell uses (higher order bias reducing) kernel estimators to estimate τ_{10} and τ_{20}, whereas Newey uses series estimators.

Neither Assumption 6a nor 6b need hold in this example. Thus, only the W and LM_a statistics are available for testing parametric restrictions.

To verify Assumption 2(c) for the three–step estimator, we treat $m_{1t}(\alpha, \tau_4)$ and $m_{2t}(\theta,\tau)$ separately. The former is assumed to satisfy Assumption 2(c) by choice of $\hat{\alpha}$.

For the latter, we have

$$\sqrt{T}\, Em_{2t}(\theta_0, \tau)\Big|_{\tau=\hat{\tau}} = \sqrt{T}\int 1(D = 1))[\tau_{10}(\alpha_0, v) - \hat{\tau}_1(\alpha_0, v)]$$

$$(6.4.7) \qquad \times [\tau_{20}(\alpha_0, v) - \hat{\tau}_2(\alpha_0, v)]/\hat{\tau}_3(\alpha_0, v)dP(D,v)$$

$$+ \sqrt{T}\int 1(D = 1)(\tau_{20}(\alpha_0, v) - \hat{\tau}_2(\alpha_0, v))'\beta_0$$

$$\times [\tau_{20}(\alpha_0, v) - \hat{\tau}_2(\alpha_0, v)]/\hat{\tau}_3(\alpha_0, v)dP(D,v)\,,$$

where $P(\cdot,\cdot)$ denotes the distribution of $(D_t, h(Z_t, \alpha_0))$. Hence, Assumption 2(c) holds if

$$T^{1/4}\left[\int 1(D = 1)(\hat{\tau}_1(\alpha_0,v) - \tau_{10}(\alpha_0,v))^2 dP(D,v)\right]^{1/2} \xrightarrow{p} 0\,,$$

$$(6.4.8) \qquad T^{1/4}\left[\int 1(D = 1)\|\hat{\tau}_2(\alpha_0,v) - \tau_{20}(\alpha_0,v)\|^2 dP(D,v)\right]^{1/2} \xrightarrow{p} 0\,, \text{ and}$$

$$\inf_{v\in\mathcal{V}}|\hat{\tau}_3(\alpha_0, v)| \ge \epsilon \text{ with probability} \to 1 \text{ for some } \epsilon > 0\,,$$

where $\mathcal{V}$ denotes the support of $h(Z_t, \alpha_0)$.

As stated above, Assumption 2(c) fails when one tries to establish the asymptotic normality of $\hat{\beta}$ with $\hat{\alpha}$ defined as part of the nuisance parameter $\hat{\tau}$ rather than as part of $\hat{\theta}$. The difference in Assumption 2(c) between these cases is that in the former $\hat{\alpha}$ enters the expression $Em_{2t}(\theta_0, \tau)\big|_{\tau=\hat{\tau}}$, whereas in the latter α_0 enters this expression.

Next, we consider the second part of Assumption 2(b). With ρ_T defined by (3.23), we have

$$\rho_T(\hat{\tau}, \tau_0) \le \left[E\|m_{1t}(\alpha_0, \tau_4) - m_{1t}(\alpha_0, \tau_{40})\|^2\Big|_{\tau_4=\hat{\tau}_4}\right]^{1/2}$$

$$(6.4.9) \qquad + \left[E\|m_{2t}(\theta_0, \tau) - m_{2t}(\theta_0, \tau_0)\|^2\Big|_{\tau=\hat{\tau}}\right]^{1/2}.$$

By choice of $\hat{\alpha}$, the first summand on the right–hand side of of (6.4.9) is assumed to have probability limit zero. The second summand of (6.4.9) squared equals

$$\int 1(D=1)\|[\tau_{10}(v)-\hat{\tau}_1(v)+(\hat{\tau}_2(v)-\tau_{20}(v))'\beta_0] \tag{6.4.10}$$
$$\times [x-\hat{\tau}_2(v)]/\hat{\tau}_3(v)+\mu[\tau_{20}(v)-\hat{\tau}_2(v)]/\hat{\tau}_3(v)$$
$$+\mu[x-\tau_{20}(v)][\tau_{30}(v)-\hat{\tau}_3(v)]/(\hat{\tau}_3(v)\tau_{30}(v))\|^2 dP(\mu,D,x,v)\ ,$$

where $\tau_{j0}(v)$ and $\hat{\tau}_j(v)$ abbreviate $\tau_{j0}(\alpha_0, v)$ and $\hat{\tau}_j(\alpha_0, v)$ for $j = 1, 2, 3$ and $P(\cdot,\cdot,\cdot,\cdot)$ denotes the distribution of $(\mu_t, D_t, X_t, h(Z_t, \alpha_0))$. Thus, sufficient conditions for the second summand of (6.4.9) to have probability limit zero are

$$\int 1(D=1)\|\hat{\tau}_j(\alpha_0,v)-\tau_{j0}(\alpha_0,v)\|^4 dP(D,v) \xrightarrow{p} 0 \ \text{ for } j=1,2,3\ ,$$
$$\inf_{v\in\mathcal{V}}|\tau_{30}(\alpha_0,v)|>0,\ \inf_{v\in\mathcal{V}}|\hat{\tau}_3(\alpha_0,v)|\geq\epsilon \text{ with probability} \to 1 \text{ for some } \epsilon>0, \tag{6.4.11}$$
$$ED_tU_t^8<\infty\ , \text{ and } ED_t\|X_t\|^8<\infty\ ,$$

where $P(\cdot,\cdot)$ denotes the distribution of $(D_t, h(Z_t, \alpha_0))$.

6.5. Adaptive Linear Regression Estimation with Asymmetric Errors

This example illustrates how the MINPIN results given above can be applied to a model in which some parameters are adaptively estimable, while others are not. The MINPIN results can be used to obtain the $\sqrt{T}$–consistency and asymptotic normality of estimators of the adaptively estimable parameters. The model considered is inherently a time series model. The estimator considered has not been considered elsewhere in the literature (to the best of my knowledge). In contrast to most of the previous examples, the estimator utilizes a preliminary nonparametric density estimator rather than a nonparametric regression estimator.

The model is

$$Y_t = X_t'\theta_0 + U_t \ \text{ for } t=1,\ \ldots,\ T\ , \tag{6.5.1}$$

where (i) $\{(X_t, U_t) : t \geq 1\}$ is a stationary asymptotically weakly dependent (e.g., near-epoch dependent) sequence of square integrable rv's in R^p and R respectively, (ii) X_t and U_t are independent $\forall t$, (iii) $E(X_t - EX_t)(X_t - EX_t)'$ is nonsingular,

(iv) $EX_t = \underset{\sim}{0}$ and (v) for some integer $r \geq 0$, $\{U_t : t \geq 1\}$ is an r–th order Markov process with density $\tau_0(u_t | u_{t-1}, \ldots, u_{t-r})$ of U_t given $(U_{t-1}, \ldots, U_{t-r})$ (with respect to some measure μ). If $r = 0$, the errors $\{U_t\}$ are independent and $\tau_0(u_t | u_{t-1}, \ldots, u_{t-r})$ denotes $\tau_0(u_t)$, the unconditional density of U_t. This case is considered by Bickel (1982), although Bickel considers a different estimator from the one discussed below. Bickel's estimator is a one–step version of the one considered below.

Under assumptions (i)–(iii), the location of U_t is not specified, and hence, assumption (iv) can always be made to hold by redefining X_t and U_t. Assumption (iii) implies that X_t does not contain an intercept. The intercept is not deemed important here and is incorporated in the error U_t. Since the error need not be symmetric about some point, there is no unambiguous definition of the intercept in this example.

We consider an estimator $\hat{\theta}$ of the slope parameters θ_0 that is adaptive in the sense of being asymptotically efficient for any distribution of the errors $\{U_t\}$ within a given nonparametric class of distributions. Let $\hat{\tau}$ be an estimator of the density τ_0 given $U_{t-1}, \ldots, U_{t-r}$. For example, for $r \geq 1$, $\hat{\tau}$ could be the ratio of nonparametric estimators of the densities of $(U_t, \ldots, U_{t-r})$ and $(U_{t-1}, \ldots, U_{t-r})$ based on the LS residuals $\{\hat{U}_t\}$. An estimator $\hat{\theta}$ of the regression slope parameters θ_0 is defined to minimize

$$(6.5.2) \qquad -\Sigma_1^T \log \hat{\tau}(U_t(\theta) | U_{t-1}(\theta), \ldots, U_{t-r}(\theta))$$

over $\theta \in \Theta \subset R^p$, where $U_t(\theta) = Y_t - X_t'\theta$. The estimator $\hat{\theta}$ is a MINPIN estimator with $W_t = (Y_t, X_t')'$,

$$(6.5.3) \quad d(m,\gamma) = m'm/2 \text{ , and } m_t(\theta,\tau) = \sum_{j=0}^{r} \frac{\tau^{(j)}(U_t(\theta) | U_{t-1}(\theta), \ldots, U_{t-r}(\theta))}{\tau(U_t(\theta) | U_{t-1}(\theta), \ldots, U_{t-r}(\theta))} X_{t-j} \text{ ,}$$

where $\tau^{(j)}$ denotes the derivative of τ with respect to its j–th argument.

When Assumption 2 holds, the asymptotic covariance matrix V of $\hat{\theta}$ is

$$V = S^{-1} = (-M)^{-1} = \left[\sum_{j=0}^{r} \sum_{k=0}^{r} E \frac{\tau_0^{(j)} \tau_0^{(k)}}{\tau_0^2} E X_{t-j} X'_{t-k} \right]^{-1} , \tag{6.5.4}$$

where $\tau_0^{(j)}$ and τ_0 abbreviate $\tau_0^{(j)}(U_t | U_{t-1}, \ldots, U_{t-r})$ and $\tau_0(U_t | U_{t-1}, \ldots, U_{t-r})$ respectively. This is the same covariance matrix as that of the maximum likelihood estimator when the latter uses the true conditional density τ_0. Thus, $\hat{\theta}$ is adaptive and asymptotically efficient.

Assumption 6b holds in this example with

$$\rho_{Tt}(W_t, \theta, \tau) = -\log \tau(U_t(\theta) | U_{t-1}(\theta), \ldots, U_{t-r}(\theta)) \tag{6.5.5}$$

and $c = \hat{c} = -1$. Thus, the LR_b statistic for testing $H_0 : h(\theta_0) = \underset{\sim}{0}$ is given by

$$\begin{aligned} LR_{bT} = -2\Big[& \Sigma_1^T \log \hat{\tau}(U_t(\tilde{\theta}_b) | U_{t-1}(\tilde{\theta}_b), \ldots, U_{t-r}(\tilde{\theta}_b)) \\ & - \Sigma_1^T \log \hat{\tau}(U_t(\hat{\theta}) | U_{t-1}(\hat{\theta}), \ldots, U_{t-r}(\hat{\theta})) \Big] , \end{aligned} \tag{6.5.6}$$

where $\tilde{\theta}_b$ is the restricted estimator that minimizes (6.5.2) over $\theta \in \Theta$ subject to the restrictions. The W, LM_a, and LM_b statistics simplify as in (5.14).

Assumption 2(c) holds trivially in this example by (3.26)–(3.27), because

$$\sqrt{T}\, \bar{m}_T^*(\theta_0, \tau) = \sqrt{T} \sum_{j=0}^{r} E \frac{\tau^{(j)}(U_t | U_{t-1}, \ldots, U_{t-r})}{\tau(U_t | U_{t-1}, \ldots, U_{t-r})} X_{t-j} = \underset{\sim}{0} , \ \forall \tau , \ \forall T \geq 1 , \tag{6.5.7}$$

provided $E|\tau^{(j)}(U_t | (U_{t-1}, \ldots, U_{t-r}) / \tau(U_t | U_{t-1}, \ldots, U_{t-r})| < \infty \quad \forall j = 1, \ldots, r$, using the independence of U_t and X_t and the mean zero property of X_t.

If Assumption 2(e) is verified using the pseudo–metric of (3.24), as is convenient for the case of dependent errors $(r \geq 1)$, see Theorem II.7 of ASEM:II, then X_t and U_t must be bounded (or the observations where X_t and the residuals $\hat{U}_t$ lie outside a given bounded region must be trimmed out) and the second part of Assumption 2(b) holds if

(6.5.8)
$$\int \cdots \int_{\mathcal{U}} \left[\frac{\hat{\tau}^{(j)}(u_t \mid u_{t-1}, \ldots, u_{t-r})}{\hat{\tau}(u_t \mid u_{t-1}, \ldots, u_{t-r})} - \frac{\tau_0^{(j)}(u_t \mid u_{t-1}, \ldots, u_{t-r})}{\tau(u_t \mid u_{t-1}, \ldots, u_{t-r})} \right]^2 \times du_t \times \cdots \times du_{t-r} \stackrel{p}{\to} 0$$

for all $j = 0, \ldots, r$, where $\mathcal{U}$ ($\subset R^{r+1}$) denotes the support of $(U_t, \ldots, U_{t-r})$.

Alternatively, if Assumption 2(e) is verified using the pseudo–metric of (3.23), as is convenient for the case of independent errors $(r = 0)$, see Theorems II.4 and II.5 of ASEM:II, then the second part of Assumption 2(b) holds if

(6.5.9)
$$\rho_T(\hat{\tau}, \tau_0)^2 / E\|X_1\|^2 = \int \left[\frac{\hat{\tau}'(u)}{\hat{\tau}(u)} - \frac{\tau_0'(u)}{\tau_0(u)} \right]^2 dP(u) \stackrel{p}{\to} 0 ,$$

where $P(\cdot)$ denotes the distribution of U_t.

The model discussed above has errors that are homoskedastic but are not necessarily symmetric. Analogous results can be obtained for (linear and nonlinear) regression models that have heteroskedastic symmetric errors. See Manski (1984) for the analysis of a one-step estimator for such models.

APPENDIX

For notational simplicity, we let $\bar{m}_T(\theta)$ abbreviate $\bar{m}_T(\theta,\hat{\tau})$ and $m(\theta)$ abbreviate $m(\theta, \tau_0)$ throughout the Appendix except in those places where the dependence on $\hat{\tau}$ or τ_0 must be made explicit for reasons of clarity.

The proof of Theorem I.1 uses the following lemma, which is similar to numerous results in the literature. The lemma appears in this form in Pötscher and Prucha (1989, Lemma 3.1) (with a different proof than that given below) and perhaps elsewhere in the literature.

LEMMA A–1: *Suppose* $\hat{\theta}$ *minimizes a random real function* $Q_T(\theta)$ *over* $\theta \in \Theta$ *with probability* $\to 1$, *where* Θ *is a pseudo–metric space. If*

(a) $\sup_{\theta \in \Theta} |Q_T(\theta) - Q(\theta)| \xrightarrow{p} 0$ *for some real function* Q *on* Θ *and*

(b) *for every neighborhood* Θ_0 *of* θ_0, $\inf_{\theta \in \Theta/\Theta_0} Q(\theta) > Q(\theta_0)$,

then $\hat{\theta} \xrightarrow{p} \theta_0$.

PROOF OF LEMMA A–1: By Assumption (b), given any neighborhood Θ_0 of θ_0, there exists a constant $\delta > 0$ such that $\inf_{\theta \in \Theta/\Theta_0} Q(\theta) > Q(\theta_0) + \delta$. Thus,

$$P(\hat{\theta} \in \Theta/\Theta_0) \leq P(Q(\hat{\theta}) - Q(\theta_0) > \delta) \to 0 , \tag{A.1}$$

where $\to 0$ holds provided $Q(\hat{\theta}) \xrightarrow{p} Q(\theta_0)$. Using Assumptions (a) and (b), the latter follows from

$$\begin{aligned} 0 &\leq Q(\hat{\theta}) - Q(\theta_0) = Q(\hat{\theta}) - Q_T(\hat{\theta}) + Q_T(\hat{\theta}) - Q(\theta_0) \\ &\leq Q(\hat{\theta}) - Q_T(\hat{\theta}) + Q_T(\theta_0) - Q(\theta_0) + o_p(1) \leq 2 \sup_{\theta \in \Theta} |Q_T(\theta) - Q(\theta)| + o_p(1) \xrightarrow{p} 0 . \ \square \end{aligned} \tag{A.2}$$

PROOF OF THEOREM I.1: We show that Assumption 1 implies that conditions (a) and (b) of Lemma A–1 hold with $Q_T(\theta) = d(\bar{m}_T(\theta,\hat{\tau}), \hat{\gamma})$ and $Q(\theta) = d(m(\theta, \tau_0), \gamma_0)$. Condition (b) holds by Assumption 1(d). Condition (a) follows from

$$
\begin{aligned}
&\sup_{\theta\in\Theta} |d(\bar{m}_T(\theta,\hat{\tau}), \hat{\gamma}) - d(m(\theta, \tau_0), \gamma_0)| \\
&\le \sup_{\theta\in\Theta} |d(\bar{m}_T(\theta,\hat{\tau}), \hat{\gamma}) - d(m(\theta,\hat{\tau}),\hat{\gamma})| + \sup_{\theta\in\Theta} |d(m(\theta,\hat{\tau}), \hat{\gamma}) - d(m(\theta, \tau_0), \gamma_0)| \\
&\le \sup_{\substack{\theta\in\Theta,\, \tau\in\mathcal{T}, \\ \gamma\in\Gamma_0}} |d(\bar{m}_T(\theta,\tau), \gamma) - d(m(\theta,\tau), \gamma)| + o_p(1) \\
\text{(A.3)}\qquad &\qquad + \sup_{\theta\in\Theta} |d(m(\theta,\hat{\tau}), \hat{\gamma}) - d(m(\theta, \tau_0), \gamma_0)| \\
&\xrightarrow{p} 0 ,
\end{aligned}
$$

where " $\xrightarrow{p} 0$ " holds using Assumptions 1(a)–(c). □

PROOF OF THEOREM I..2: First we suppose Assumption 2 holds. Element by element mean value expansions of $\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\hat{\theta}), \hat{\gamma})$ about θ_0 give: $\forall j = 1, \ldots, p$,

$$
\begin{aligned}
o_p(1) &= \sqrt{T}\frac{\partial}{\partial\theta_j}d(\bar{m}_T(\hat{\theta}), \hat{\gamma}) \\
\text{(A.4)}\qquad &= \sqrt{T}\frac{\partial}{\partial\theta_j}d(\bar{m}_T(\theta_0), \hat{\gamma}) + \frac{\partial^2}{\partial\theta'\partial\theta_j}d(\bar{m}_T(\theta^*),\hat{\gamma})\sqrt{T}(\hat{\theta} - \theta_0) ,
\end{aligned}
$$

where θ^* is a rv that depends on j and lies on the line segment joining $\hat{\theta}$ and θ_0, and hence, $\theta^* \xrightarrow{p} \theta_0$. (See Jennrich (1969) Lemma 3 for the mean value theorem for random functions.) The first equality holds because $\hat{\theta}$ minimizes $d(\bar{m}_T(\theta), \hat{\gamma})$ and $\hat{\theta}$ is in the interior of Θ with probability $\to 1$ by Assumption 2(a). The second equality actually only holds with probability $\to 1$, since the mean value expansions require $\hat{\theta} \in \Theta_0$.

Below we show that

$$
\text{(A.5)}\qquad \frac{\partial^2}{\partial\theta'\partial\theta_j}d(\bar{m}_T(\theta^*), \hat{\gamma}) = \frac{\partial^2}{\partial\theta'\partial\theta_j}d(m(\theta_0), \gamma_0) + o_p(1),
$$

where $\frac{\partial^2}{\partial\theta\partial\theta'}d(m(\theta_0), \gamma_0) = M'DM$, and

$$\sqrt{T}\,\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_0), \hat{\gamma}) \xrightarrow{d} N(\underset{\sim}{0}, M'DSDM)\ . \tag{A.6}$$

These results, equation (A.4), and the nonsingularity of $M'DM$ give

$$\sqrt{T}(\hat{\theta}-\theta_0) = -(M'DM)^{-1}\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_0),\hat{\gamma}) + o_p(1) \xrightarrow{d} N(\underset{\sim}{0},V)\ . \tag{A.7}$$

To show (A.5), we proceed as follows:

$$\begin{aligned}\frac{\partial^2}{\partial\theta_j\partial\theta_\ell}d(\bar{m}_T(\theta^*),\hat{\gamma}) &= \frac{\partial^2}{\partial\theta_j\partial\theta_\ell}\bar{m}_T(\theta^*)'\frac{\partial}{\partial m}d(\bar{m}_T(\theta^*), \hat{\gamma}) \\ &\quad + \frac{\partial}{\partial\theta_j}\bar{m}_T(\theta^*)'\frac{\partial^2}{\partial m\partial m'}d(\bar{m}_T(\theta^*),\hat{\gamma})\frac{\partial}{\partial\theta_\ell}\bar{m}_T(\theta^*)\ .\end{aligned} \tag{A.8}$$

By Assumptions 2(a), (b), and (g),

$$\begin{aligned}\|\bar{m}_T(\theta^*) - m(\theta_0)\| &\le \|\bar{m}_T(\theta^*,\hat{\tau}) - \bar{m}_T^*(\theta^*,\hat{\tau})\| \\ &\quad + \|\bar{m}_T^*(\theta^*,\hat{\tau}) - m(\theta^*,\hat{\tau})\| + \|m(\theta^*,\hat{\tau}) - m(\theta_0,\tau_0)\| \xrightarrow{p} 0\ .\end{aligned} \tag{A.9}$$

Using this result, the continuity of $\frac{\partial}{\partial m}d(m,\gamma)$ at $(m(\theta_0),\gamma_0)$ (Assumption 2(f)), the Assumption 2(b) that $\hat{\gamma} \xrightarrow{p} \gamma_0$, and the continuous mapping theorem, we get

$$\frac{\partial}{\partial m}d(\bar{m}_T(\theta^*), \hat{\gamma}) \xrightarrow{p} \frac{\partial}{\partial m}d(m(\theta_0), \gamma_0) = \underset{\sim}{0}\ , \tag{A.10}$$

where the equality holds by Assumptions 2(b), (c), (f), and (g). Using Assumption 2(g) and Markov's inequality, it is straightforward to show that $\frac{\partial^2}{\partial\theta_j\partial\theta_\ell}\bar{m}_T(\theta^*) = O_p(1)$. This result and (A.10) imply that the first term of (A.8) is $o_p(1)$.

Similarly, the continuity of $\frac{\partial^2}{\partial m\partial m'}d(m,\gamma)$ at $(m(\theta_0), \gamma_0)$ (Assumption 2(f)), equation (A.9), $\hat{\gamma} \xrightarrow{p} \gamma_0$, and the continuous mapping theorem give

$$\frac{\partial^2}{\partial m\partial m'}d(\bar{m}_T(\theta^*),\hat{\gamma}) \xrightarrow{p} \frac{\partial^2}{\partial m\partial m'}d(m(\theta_0),\gamma_0) \equiv D\ . \tag{A.11}$$

It follows from Assumptions 2(a), (b), and (g) that $M = M(\theta_0,\tau_0)$ and

$$\text{(A.12)} \qquad \left\|\frac{\partial}{\partial\theta'}\bar{m}_T(\theta^*) - M\right\| \le \left\|\frac{\partial}{\partial\theta'}\bar{m}_T(\theta^*) - M(\theta^*,\hat{\tau})\right\| + \|M(\theta^*,\hat{\tau}) - M(\theta_0,\tau_0)\| \xrightarrow{p} 0 .$$

Equations (A.11) and (A.12) imply that the second term of (A.8) equals $[M'DM]_{j\ell} + o_p(1)$, and hence, (A.5) is established.

To establish equation (A.6), we write

$$\text{(A.13)} \qquad \begin{aligned} \sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_0), \hat{\gamma}) &= \sqrt{T}\left[\frac{\partial}{\partial\theta'}\bar{m}_T(\theta_0)\right]'\frac{\partial}{\partial m}d(\bar{m}_T(\theta_0), \hat{\gamma}) \\ &= M'\sqrt{T}\frac{\partial}{\partial m}d(\bar{m}_T(\theta_0), \hat{\gamma}) + o_p(1) \end{aligned}$$

using 2(g) provided $\sqrt{T}\frac{\partial}{\partial m}d(\bar{m}_T(\theta_0), \hat{\gamma}) = O_p(1)$, as we now demonstrate.

By the mean value theorem, the j–th element of $\sqrt{T}\frac{\partial}{\partial m}d(\bar{m}_T(\theta_0,\hat{\tau}), \hat{\gamma})$ can be expanded about $\bar{m}^*_T(\theta_0,\hat{\tau})$ to get:

$$\text{(A.14)} \qquad \begin{aligned} &\sqrt{T}\frac{\partial}{\partial m_j}d(\bar{m}_T(\theta_0,\hat{\tau}), \hat{\gamma}) \\ &= \sqrt{T}\frac{\partial}{\partial m_j}d(\bar{m}^*_T(\theta_0, \hat{\tau}), \hat{\gamma}) + \frac{\partial^2}{\partial m'\partial m_j}d(m^*,\hat{\gamma})\sqrt{T}(\bar{m}_T(\theta_0, \hat{\tau}) - \bar{m}^*_T(\theta_0, \hat{\tau})) , \end{aligned}$$

where m^* is on the line segment joining $\bar{m}_T(\theta_0, \hat{\tau})$ and $\bar{m}^*_T(\theta_0, \hat{\tau})$, and hence, $m^* \xrightarrow{p} m(\theta_0)$. (More precisely, (A.14) holds with probability $\to$ 1.)

The first term of the right–hand side of (A.14) is $o_p(1)$ by Assumption 2(c). Also, using Assumption 2(f), $\frac{\partial^2}{\partial m'\partial m_j}d(m^*,\hat{\gamma}) = [D]_j + o_p(1)$, where $[D]_j$ denotes the j–th row of D. Hence, if $\sqrt{T}(\bar{m}_T(\theta_0,\hat{\tau}) - \bar{m}^*_T(\theta_0,\hat{\tau})) = O_p(1)$, the above results and (A.14) yield

$$\text{(A.15)} \qquad \sqrt{T}\frac{\partial}{\partial m}d(\bar{m}_T(\theta_0,\hat{\tau}), \hat{\gamma}) = D\sqrt{T}(\bar{m}_T(\theta_0, \hat{\tau}) - \bar{m}^*_T(\theta_0, \hat{\tau})) + o_p(1) .$$

The proof based on Assumption 2 is complete once we show that

$$\text{(A.16)} \qquad \nu_T(\hat{\tau}) = \sqrt{T}(\bar{m}_T(\theta_0, \hat{\tau}) - \bar{m}^*_T(\theta_0, \hat{\tau})) \xrightarrow{d} N(\underset{\sim}{0},S) ,$$

since this implies that (A.15) and (A.13) hold, which establishes (A.6).

Using Assumption 2(d), one sees that (A.16) holds if $\nu_T(\hat{\tau}) - \nu_T(\tau_0) \xrightarrow{p} \underset{\sim}{0}$. The latter follows from Assumptions 2(b) and (e) by (3.12).

Now, we prove the second part of Theorem I.2. Suppose Assumption 2* holds. Element by element mean value expansions of $\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T^*(\theta_0), \hat{\gamma})$ about $\hat{\theta}$ give: $\forall j = 1, \ldots, p$,

$$\text{(A.17)}\quad o_p(1) = \sqrt{T}\frac{\partial}{\partial\theta_j}d(\bar{m}_T^*(\theta_0), \hat{\gamma}) = \sqrt{T}\frac{\partial}{\partial\theta_j}d(\bar{m}_T^*(\hat{\theta}), \hat{\gamma}) - \frac{\partial^2}{\partial\theta'\partial\theta_j}d(\bar{m}_T^*(\theta^*), \hat{\gamma})\sqrt{T}(\hat{\theta} - \theta_0),$$

where θ^* is a rv on the line segment joining $\hat{\theta}$ and θ_0, and hence, $\theta^* \xrightarrow{p} \theta_0$. The first equality holds by Assumption 2*(c), because

$$\text{(A.18)}\qquad \sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T^*(\theta_0), \hat{\gamma}) = \left[\frac{\partial}{\partial\theta'}\bar{m}_T^*(\theta_0, \hat{\tau})\right]'\sqrt{T}\frac{\partial}{\partial m}d(\bar{m}_T^*(\theta_0, \hat{\tau}), \hat{\gamma})$$

and $\frac{\partial}{\partial\theta'}\bar{m}_T^*(\theta_0, \hat{\tau}) = M + o_p(1)$ by Assumptions 2*(b) and (g).

By an argument analogous to that used to establish (A.5) above, we have

$$\text{(A.19)}\qquad \frac{\partial^2}{\partial\theta'\partial\theta_j}d(\bar{m}_T^*(\theta^*), \hat{\gamma}) = [M'DM]_j + o_p(1)\,.$$

This argument uses the fact that Assumptions 2*(b), (c), (f), and (g) imply that $\frac{\partial}{\partial m}d(m(\theta_0), \gamma_0) = \underset{\sim}{0}$.

Next we show that

$$\text{(A.20)}\qquad \sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T^*(\hat{\theta}), \hat{\gamma}) = \left[\frac{\partial}{\partial\theta'}\bar{m}_T^*(\hat{\theta})\right]'\sqrt{T}\frac{\partial}{\partial m}d(\bar{m}_T^*(\hat{\theta}), \hat{\gamma}) \xrightarrow{d} N(\underset{\sim}{0}, M'DSDM)\,.$$

Equations (A.17), (A.19), and (A.20) combine to give the desired result.

To show (A.20), first note that $\frac{\partial}{\partial\theta'}\bar{m}_T^*(\hat{\theta}) = M + o_p(1)$ by Assumptions 2*(a), (b), and (g). Next, expand $\sqrt{T}\frac{\partial}{\partial m_j}d(\bar{m}_T^*(\hat{\theta}), \hat{\gamma})$ about $\bar{m}_T(\hat{\theta},\hat{\tau})$:

$$\text{(A.21)}\ \sqrt{T}\frac{\partial}{\partial m_j}d(\bar{m}_T^*(\hat{\theta}),\hat{\gamma}) = \sqrt{T}\frac{\partial}{\partial m_j}d(\bar{m}_T(\hat{\theta},\hat{\tau}),\hat{\gamma}) - \frac{\partial^2}{\partial m'\partial m_j}d(m^*,\hat{\gamma})\sqrt{T}(\bar{m}_T(\hat{\theta},\hat{\tau}) - \bar{m}_T^*(\hat{\theta},\hat{\tau})),$$

where m^* is on the line segment joining $\bar{m}_T(\hat{\theta},\hat{\tau})$ and $\bar{m}_T^*(\hat{\theta},\hat{\tau})$, and hence, $m^* \xrightarrow{p} m(\theta_0, \tau_0)$. The first term on the right–hand side of (A.21) stacked for $j = 1, \ldots, v$ to form a vector and premultiplied by $\left[\frac{\partial}{\partial\theta'}\bar{m}_T^*(\hat{\theta})\right]'$ is $o_p(1)$ by Assumption 2*(a). The matrix $\frac{\partial^2}{\partial m'\partial m_j}d(m^*,\hat{\gamma})$ equals $[D]_j + o_p(1)$ by Assumptions 2*(b) and (f). In addition,

$\nu_T(\hat{\theta},\hat{\tau}) = \sqrt{T}(\bar{m}_T(\hat{\theta},\hat{\tau}) - \bar{m}_T^*(\hat{\theta},\hat{\tau}))$ converges in distribution to $N(\underset{\sim}{0},S)$ by Assumptions 2*(a), (b), (d), and (e) (by an analogous argument to that of (3.12)). This gives (A.20) and the proof is complete.

Last, suppose Assumption 2** holds. The difference between Assumptions 2 and 2** is that the latter specifies $\mathcal{T}$ to be finite dimensional and it replaces the stochastic equicontinuity of $\{\nu_T(\cdot)\}$ at τ_0 by Assumption 2**(e). Stochastic equicontinuity of $\{\nu_T(\cdot)\}$ is used in the proof above only to show that $\nu_T(\hat{\tau}) - \nu_T(\tau_0) \xrightarrow{p} \underset{\sim}{0}$. Thus, it suffices to show that Assumption 2**(e) implies $\nu_T(\hat{\tau}) - \nu_T(\tau_0) \xrightarrow{p} \underset{\sim}{0}$.

For each $j = 1, \ldots, v$, a mean value expansion yields

$$\text{(A.22)} \qquad \nu_{Tj}(\hat{\tau}) - \nu_{Tj}(\tau_0) = \left[\frac{1}{\sqrt{T}}\frac{\partial}{\partial\tau'}\nu_{Tj}(\tau^*)\right]\sqrt{T}(\hat{\tau} - \tau_0),$$

where τ^* lies on the line segment joining $\hat{\tau}$ and τ_0. The right–hand side above is $o_p(1)$ under Assumption 2**(e), since $\sqrt{T}(\hat{\tau} - \tau_0) = O_p(1)$ and

$$\begin{aligned}\left\|\frac{1}{\sqrt{T}}\frac{\partial}{\partial\tau'}\nu_{Tj}(\tau^*)\right\| &= \left\|\frac{1}{T}\Sigma_1^T\left[\frac{\partial}{\partial\tau'}m_{tj}(\theta_0, \tau^*) - E\frac{\partial}{\partial\tau'}m_{tj}(\theta_0, \tau)\Big|_{\tau=\tau^*}\right]\right\| \\ \text{(A.23)} \qquad &\le \sup_{\tau\in\mathcal{T}}\left\|\frac{1}{T}\Sigma_1^T\left[\frac{\partial}{\partial\tau'}m_{tj}(\theta_0, \tau) - E\frac{\partial}{\partial\tau'}m_{tj}(\theta_0, \tau)\right]\right\| \\ &\xrightarrow{p} 0,\end{aligned}$$

where $\frac{\partial}{\partial\tau'}Em_{tj}(\theta_0, \tau)\Big|_{\tau=\tau^*} = E\frac{\partial}{\partial\tau'}m_{tj}(\theta_0, \tau)\Big|_{\tau=\tau^*}$ (which is used in the equality above) holds by the moment condition in Assumption 2**(e). □

PROOF OF THEOREM I.3: Under Assumptions 2 or 2**, $\hat{D} \xrightarrow{p} D$ and $\hat{M} \xrightarrow{p} M$ by the arguments used in equations (A.9), (A.11), and (A.12). Thus, under these assumptions plus Assumption 3, $\hat{\mathcal{I}} \xrightarrow{p} \mathcal{I}$, $\hat{\mathcal{J}}^{-1} \xrightarrow{p} \mathcal{J}^{-1}$ (since $\mathcal{J}$ is nonsingular by Assumption 2(h)), and $\hat{V} \xrightarrow{p} V$.

Under Assumption 2*, $\bar{m}_T(\hat{\theta},\hat{\tau}) \xrightarrow{p} m(\theta_0, \tau_0)$ by 2*(b), (e), and (g), $\hat{\gamma} \xrightarrow{p} \gamma_0$ by 2*(b), and $\frac{\partial^2}{\partial m\partial m'}d(m,\gamma)$ is continuous at $(m(\theta_0, \tau_0), \gamma_0)$ by 2*(f). Combining these results gives $\hat{D} \xrightarrow{p} D$ when Assumption 2* holds.

Next, suppose Assumptions 2* and 3* hold. We show that $\hat{M} \xrightarrow{p} M$ as follows. Let M_j and $M_j(\theta,\tau)$ denote the j–th columns of M and $M(\theta,\tau)$ respectively. With some abuse of notation, let $Em_t(\hat{\theta},\hat{\tau})$ denote $Em_t(\theta,\tau)$ evaluated at $(\theta,\tau) = (\hat{\theta},\hat{\tau})$. We have

$$
\begin{aligned}
\|\hat{M}_j - M_j\| \le & \left\|\hat{M}_j - \tfrac{1}{T}\Sigma_1^T(Em_t(\hat{\theta} + \epsilon_T e_j, \hat{\tau}) - Em_t(\hat{\theta} - \epsilon_T e_j, \hat{\tau}))/(2\epsilon_T)\right\| \\
& + \left\|\tfrac{1}{T}\Sigma_1^T(Em_t(\hat{\theta} + \epsilon_T e_j, \hat{\tau}) - Em_t(\hat{\theta} - \epsilon_T e_j, \hat{\tau}))/(2\epsilon_T) - \tfrac{1}{T}\Sigma_1^T \tfrac{\partial}{\partial\theta_j}Em_t(\hat{\theta},\hat{\tau})\right\| \\
& + \left\|\tfrac{1}{T}\Sigma_1^T \tfrac{\partial}{\partial\theta_j}Em_t(\hat{\theta},\hat{\tau}) - M_j(\hat{\theta},\hat{\tau})\right\| + \|M_j(\hat{\theta},\hat{\tau}) - M_j\| \qquad \text{(A.24)} \\
= & A_{1T} + A_{2T} + A_{3T} + A_{4T} \text{ (say).}
\end{aligned}
$$

Since

$$\text{(A.25)} \qquad A_{1T} = \frac{1}{2\epsilon_T\sqrt{T}}(\nu_T(\hat{\theta}+\epsilon_T e_j,\hat{\tau}) - \nu_T(\theta_0,\tau_0)) - \frac{1}{2\epsilon_T\sqrt{T}}(\nu_T(\hat{\theta}-\epsilon_T e_j,\hat{\tau}) - \nu_T(\theta_0,\tau_0)),$$

Assumptions 2*(e), 3*(a), and 3*(c) combine to yield $A_{1T} \xrightarrow{p} 0$. Assumptions 2*(a), 2*(b), 3*(a), and 3*(b) imply $A_{2T} \xrightarrow{p} 0$. Assumption 2*(g) implies $A_{3T} \xrightarrow{p} 0$ and $A_{4T} \xrightarrow{p} 0$. Hence, $\hat{M} \xrightarrow{p} M$. In turn, $\hat{V} \xrightarrow{p} V$ as above. □

PROOF OF THEOREM I.4: Under the assumptions, part (a) follows from Theorems I.2 and I.3 and the continuous mapping theorem.

Next we establish part (b). Standard arguments give

$$\text{(A.26)} \qquad \tilde{\mathcal{J}} \xrightarrow{p} \mathcal{J}, \ \tilde{H} \xrightarrow{p} H, \text{ and } \tilde{V} \xrightarrow{p} V.$$

Mean value expansions about θ_0 yield: $\forall j = 1, \ldots, p$, $\forall s = 1, \ldots, r$,

$$\text{(A.27)} \qquad \sqrt{T}\frac{\partial}{\partial\theta_j}d(\bar{m}_T(\tilde{\theta}_a),\hat{\gamma}) = \sqrt{T}\frac{\partial}{\partial\theta_j}d(\bar{m}_T(\theta_0),\hat{\gamma}) + \frac{\partial^2}{\partial\theta'\partial\theta_j}d(\bar{m}_T(\theta_1),\hat{\gamma})\sqrt{T}(\tilde{\theta}_a - \theta_0),$$

$$\text{(A.28)} \qquad \sqrt{T}\, h_s(\tilde{\theta}_a) = \sqrt{T}\, h_s(\theta_0) + \frac{\partial}{\partial\theta'}h_s(\theta_2)\sqrt{T}(\tilde{\theta}_a - \theta_0),$$

where θ_1 and θ_2 depend on j and s, respectively, and lie on the line segment joining $\tilde{\theta}_a$ and θ_0, and hence, satisfy $\theta_1 \xrightarrow{p} \theta_0$ and $\theta_2 \xrightarrow{p} \theta_0$. We stack equations (A.27) and (A.28) for $j = 1, \ldots, p$ and $s = 1, \ldots, r$ and write them as

(A.29) $\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a), \hat{\gamma}) = \sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_0), \hat{\gamma}) + \mathcal{J}_1\sqrt{T}(\tilde{\theta}_a - \theta_0)$ and

(A.30) $\underset{\sim}{0} = H_2\sqrt{T}(\tilde{\theta}_a - \theta_0)$

using the fact that $h(\tilde{\theta}_a) = h(\theta_0) = \underset{\sim}{0}$.

By equation (A.6), $\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_0), \hat{\gamma}) \xrightarrow{d} N(\underset{\sim}{0},\mathcal{I})$. By standard arguments, $\mathcal{J}_1 \xrightarrow{p} \mathcal{J}$. Hence, using the nonsingularity of $\mathcal{J}$, we get $\mathcal{J}_1^{-}\mathcal{J}_1 \doteq \mathcal{J}_1^{-1}\mathcal{J}_1 = I_p$, where $(\cdot)^{-}$ denotes some g–inverse and $\doteq$ denotes equality that holds with probability $\to 1$. By Assumptions 4 and 5a, $H_2 \xrightarrow{p} H$. Pre–multiplication of (A.29) by $H_2\mathcal{J}_1^{-}$ now gives

$$\begin{aligned}H_2\mathcal{J}_1^{-}\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a),\hat{\gamma}) &= H_2\mathcal{J}_1^{-}\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_0),\hat{\gamma}) + H_2\mathcal{J}_1^{-}\mathcal{J}_1\sqrt{T}(\tilde{\theta}_a - \theta_0)\\ &\doteq H_2\mathcal{J}_1^{-}\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_0),\hat{\gamma}) \xrightarrow{d} N(\underset{\sim}{0}, H\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}H')\,.\end{aligned} \tag{A.31}$$

Below we show that

(A.32) $\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a), \hat{\gamma}) = O_p(1)$.

Equations (A.26), (A.31), and (A.32) yield

(A.33) $\tilde{H}\tilde{\mathcal{J}}^{-1}\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a), \hat{\gamma}) \xrightarrow{d} N(\underset{\sim}{0}, HVH')$.

The desired result now follows from equations (A.26) and (A.33) and the continuous mapping theorem.

For part (b), it remains to show (A.32). With probability $\to 1$, $\tilde{\theta}_a$ is in the interior of Θ and there exists a rv $\tilde{\lambda}$ of Lagrange multipliers such that

(A.34) $\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a),\hat{\gamma}) + \tilde{H}'\tilde{\lambda} = \underset{\sim}{0}$,

where $\tilde{H} = \frac{\partial}{\partial\theta'}h(\tilde{\theta}_a)$. Equations (A.31) and (A.34) combine to give

(A.35) $-H_2\mathcal{J}_1^{-}\tilde{H}'\sqrt{T}\tilde{\lambda} \doteq H_2\mathcal{J}_1^{-}\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a),\hat{\gamma}) = O_p(1)$.

Since $H_2\mathcal{J}_1^{-}\tilde{H}' \xrightarrow{p} H\mathcal{J}^{-1}H'$ and $H\mathcal{J}^{-1}H'$ is nonsingular, equations (A.34) and (A.35) imply that $\sqrt{T}\tilde{\lambda} = O_p(1)$ and that (A.32) holds.

We now prove part (c). Suppose that Assumption 6a holds. A two–term Taylor expansion of $d(\bar{m}_T(\tilde{\theta}_a), \hat{\gamma})$ about $\hat{\theta}$ gives

$$
\begin{aligned}
LR_{aT} &= 2T\left[d(\bar{m}_T(\tilde{\theta}_a),\hat{\gamma}) - d(\bar{m}_T(\hat{\theta}),\hat{\gamma})\right]/\hat{b} \\
&= 2T\frac{\partial}{\partial\theta'}d(\bar{m}_T(\hat{\theta}),\hat{\gamma})(\tilde{\theta}_a - \hat{\theta})/\hat{b} + T(\tilde{\theta}_a - \hat{\theta})'\frac{\partial^2}{\partial\theta\partial\theta'}d(\bar{m}_T(\theta_3),\hat{\gamma})(\tilde{\theta}_a - \hat{\theta})/\hat{b} \qquad \text{(A.36)} \\
&\doteq T(\tilde{\theta}_a - \hat{\theta})'\mathcal{J}_3(\tilde{\theta}_a - \hat{\theta})/\hat{b} \,,
\end{aligned}
$$

where θ_3 lies on the line segment joining $\tilde{\theta}_a$ and $\hat{\theta}$, and hence, $\theta_3 \xrightarrow{p} \theta_0$, $\mathcal{J}_3$ is defined implicitly, and " $\doteq$ " holds by the first order conditions for the estimator $\hat{\theta}$.

Applying the mean value theorem element by element and stacking the equations yields

$$
\begin{aligned}
\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a), \hat{\gamma}) &= \sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\hat{\theta}), \hat{\gamma}) + \mathcal{J}_4\sqrt{T}(\tilde{\theta}_a - \hat{\theta}) \qquad \text{(A.37)} \\
&\doteq \mathcal{J}_4\sqrt{T}(\tilde{\theta}_a - \hat{\theta})
\end{aligned}
$$

for a matrix $\mathcal{J}_4$ that satisfies $\mathcal{J}_4 \xrightarrow{p} \mathcal{J}$. Pre–multiplying (A.37) by $\mathcal{J}_3\mathcal{J}_4^-$ and substituting the result in (A.36) gives

$$
\begin{aligned}
LR_{aT} &\doteq T\frac{\partial}{\partial\theta'}d(\bar{m}_T(\tilde{\theta}_a), \hat{\gamma})(\mathcal{J}_3\mathcal{J}_4^-)'\mathcal{J}_3^-\mathcal{J}_3\mathcal{J}_4^-\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a), \hat{\gamma})/\hat{b} \qquad \text{(A.38)} \\
&= T\frac{\partial}{\partial\theta'}d(\bar{m}_T(\tilde{\theta}_a), \hat{\gamma})\tilde{\mathcal{J}}^{-1}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a), \hat{\gamma})/\hat{b} + o_p(1) \,.
\end{aligned}
$$

This follows because (A.5), (A.26), and (A.32) imply that $\mathcal{J}_4^-\mathcal{J}_4 \doteq I_p$, $\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a), \hat{\gamma}) = O_p(1)$, $\mathcal{J}_3\mathcal{J}_4^- \xrightarrow{p} I_p$, and $\tilde{\mathcal{J}} - \mathcal{J}_3 \xrightarrow{p} \underset{\sim}{0}$.

Since $\mathcal{I} = b\mathcal{J}$ and $\hat{b} \xrightarrow{p} b$ by Assumption 6a, $\tilde{V} = \hat{b}\tilde{\mathcal{J}}^{-1} + o_p(1)$. In this case, LM_{aT} simplifies to

$$
LM_{aT} = T\frac{\partial}{\partial\theta'}d(\bar{m}_T(\tilde{\theta}_a),\hat{\gamma})\tilde{\mathcal{J}}^{-1}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a), \hat{\gamma})/\hat{b} + o_p(1) = LR_{aT} + o_p(1) \qquad \text{(A.39)}
$$

using $\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a), \hat{\gamma}) \doteq -\tilde{H}'\tilde{\lambda}$, as above. The desired result now follows from part (b) of the Theorem.

The proof of part (d) is the same as that of part (b) given in equations (A.26) to (A.35) with the following changes: (i) $\tilde{\mathcal{J}}$, $\mathcal{J}$, $d(\bar{m}_T(\cdot), \hat{\gamma})$, $\tilde{\theta}_a$, $\mathcal{J}_1$, and $\mathcal{I}$ are replaced by $\tilde{M}$, M , $\bar{\rho}_T(\cdot,\hat{\tau})$, $\tilde{\theta}_b$, M_1 , and S respectively, everywhere they appear, where M_1 has j–th row equal to $\frac{\partial^2}{\partial\theta'\partial\theta_j}\bar{\rho}_T(\theta_1, \hat{\tau})$ (as in (A.27)), (ii) $\tilde{V}$ and V are defined by $\tilde{V} = \tilde{\mathcal{J}}^{-1}\tilde{\mathcal{I}}\tilde{\mathcal{J}}^{-1} = \tilde{M}^{-1}\hat{S}\tilde{M}^{-1}$ and $V = \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1} = M^{-1}SM^{-1}$ respectively, and (iii) the assertion following (A.30) that $\sqrt{T}\frac{\partial}{\partial\theta}\bar{\rho}_T(\theta_0, \hat{\tau}) \xrightarrow{d} N(\underset{\sim}{0},S)$ is verified by noting that

$$\text{(A.40)}\qquad \begin{aligned}\sqrt{T}\frac{\partial}{\partial\theta}\bar{\rho}_T(\theta_0, \hat{\tau}) &= \sqrt{T}(\bar{m}_T(\theta_0, \hat{\tau}) - \bar{m}_T^*(\theta_0, \hat{\tau})) + \sqrt{T}\frac{\partial}{\partial m}d(\bar{m}_T^*(\theta_0, \hat{\tau}), \hat{\gamma}) \\ &\xrightarrow{d} N(\underset{\sim}{0},S) ,\end{aligned}$$

by (A.16) and Assumption 2(c).

The proof of part (e) is the same as that of part (c) given in equations (A.36) to (A.39) with the following changes. First, consider the case where $\frac{\partial}{\partial\theta}\rho_{Tt}(W_{Tt}, \theta, \gamma) = m_{Tt}(W_{Tt}, \theta, \tau)$ $\forall t \leq T$, $T \geq 1$ in Assumption 6b. (In this case, the constant c in Assumption 6b(ii) is positive.) The changes are: (i) $d(\bar{m}_T(\cdot), \hat{\gamma})$, $\tilde{\theta}_a$, LR_{aT} , $\hat{b}$, $\mathcal{J}_3$, $\mathcal{J}_4$, $\tilde{\mathcal{J}}$, $\mathcal{I}$, $\mathcal{J}$, b , and LM_{aT} are replaced by $\bar{\rho}_T(\cdot,\hat{\tau})$, $\tilde{\theta}_b$, LR_{bT} , $\hat{c}$, M_3 , M_4 , $\tilde{M}$, S , M , c , and LM_{bT} respectively, everywhere they appear, where M_3 and M_4 are defined implicitly in (A.36) and (A.37), respectively, and satisfy $M_3 \xrightarrow{p} M$ and $M_4 \xrightarrow{p} M$, (ii) $\tilde{V}$ is defined as $\tilde{V} = \tilde{M}^{-1}\hat{S}\tilde{M}^{-1}$ and $\tilde{V}$ satisfies $\tilde{V} = \hat{c}\tilde{M}^{-1} + o_p(1)$, and (iii) the references to Assumption 6a and Theorem I.4(b) are replaced by references to Assumption 6b and Theorem I.4(d). Next, consider the case where $\rho_{Tt}(W_{Tt}, \theta, \tau) = -m_{Tt}(W_{Tt}, \theta, \tau)$. (In this case, the constant c in Assumption 6b(ii) is negative.) For this case, the same changes are made as above except that $|\hat{c}|$, $-M_3$, $-M_4$, $-\tilde{M}$, $-M$, and c , rather than $\hat{c}$, M_3 , M_4 , $\tilde{M}$, M , and c , are used to replace $\hat{b}$, $\mathcal{J}_3$, $\mathcal{J}_4$, $\tilde{\mathcal{J}}$, $\mathcal{J}$, and b . (This explains the appearance of $|\hat{c}|$ rather than $\hat{c}$ in the denominator of the LR_{bT} statistic.) □

PROOF OF THEOREM I.5: First we prove part (a). The proof of Theorem I.3 shows that $\hat{D} \xrightarrow{p} D$ and $\hat{M} \xrightarrow{p} M$, since $\hat{\theta} \xrightarrow{p} \theta_0$, $\hat{\tau} \xrightarrow{p} \tau_0$, $\hat{\gamma} \xrightarrow{p} \gamma_0$, and Assumption 2(g) holds. We have: HVH' is nonsingular, $\hat{S} \xrightarrow{p} S$, and $\hat{H} \xrightarrow{p} H$, by Assumptions 4, 3, and 2–ℓp and 4, respectively. Thus, $(\hat{H}\hat{V}\hat{H}')^{-1} \xrightarrow{p} (HVH')^{-1}$.

Mean value expansions of $h_s(\hat{\theta})$ about $h_s(\theta_T)$, stacked for $s = 1,, \ldots, r$, yield

(A.41) $$\sqrt{T}h(\hat{\theta}) = \sqrt{T}h(\theta_T) + H^*\sqrt{T}(\hat{\theta} - \theta_T)$$

for an $r \times p$ matrix H^* that satisfies $H^* \xrightarrow{p} H$. Assumption 4 and element by element mean value expansions give $\sqrt{T}h(\theta_T) \to H\eta$. Part (a) now follows by the continuous mapping theorem once we show that

(A.42) $$\sqrt{T}(\hat{\theta} - \theta_T) \xrightarrow{d} N(\underset{\sim}{0}, V) .$$

This follows using Assumption 2–ℓp by the proof of Theorem I.2 with θ_0 replaced by θ_T everywhere except in the paragraphs (or parts of the paragraphs) that contain (A.5), (A.8)–(A.12), (A.19), and (A.21) and except in those cases where θ_0 appears in an expression that is the limit as $T \to \infty$ of some sequence.

To prove part (b), note that under Assumptions 2–ℓp, 3, 4, and 5a the proof of Theorem I.4(b) above goes through with the following changes: The parameter θ_0 is replaced by θ_T in equations (A.27)–(A.29) and equations (A.30), (A.31), and (A.33) are replaced by

(A.43) $$\underset{\sim}{0} = \sqrt{T}h(\theta_T) + H_2\sqrt{T}(\tilde{\theta}_a - \theta_T) ,$$

(A.44) $$H_2\mathcal{J}_1\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a), \hat{\gamma}) \doteq H_2\mathcal{J}_1\sqrt{T}\frac{\partial}{\partial\theta}d(\bar{m}_T(\theta_T), \hat{\gamma}) - \sqrt{T}h(\theta_T)$$
$$\xrightarrow{d} N(-H\eta, HVH') \text{ and}$$

(A.45) $$\tilde{H}\tilde{\mathcal{J}}^{-1}\frac{\partial}{\partial\theta}d(\bar{m}_T(\tilde{\theta}_a), \hat{\gamma}) \xrightarrow{d} N(-H\eta, HVH')$$

respectively.

Part (c) is proved using the proof of Theorem I.4(c). The latter goes through under Assumptions 2–ℓp, 4, 5a, and 6a with the only changes being that references to Theorem I.4(b) are replaced by references to Theorem I.5(b).

Part (d) is proved using the proof of Theorem I.4(b) with the changes to that proof that are described in the proofs of Theorems I.4(d) and I.5(b), with θ_0 replaced by θ_T in (A.40), and with the changes described in the proof of Theorem I.4(d) also applied to (A.43)–(A.45).

Part (e) is proved using the proof of Theorem I.4(c) with the changes to that proof outlined in the proof of Theorem I.4(e), but with references to Theorem I.4(d) in the latter replaced by references to Theorem I.5(d). □

FOOTNOTES

[1]I thank Whitney K. Newey, Adrian Pagan, David Pollard, B. M. Pötscher, James L. Powell, Yoon–Jae Whang, two referees, and the co–editor Lars Hansen for helpful comments. I gratefully acknowledge research support from the Alfred P. Sloan Foundation and the National Science Foundation through a Research Fellowship and grant nos. SES–8618617 and SES–8821021 respectively.

[2]This is not to suggest that existing methods are incapable of being extended in such directions.

[3]Manski's maximum score estimator fails a full rank condition, Assumption 2(h), introduced below when one takes $d(m,\gamma) = m$. Horowitz's smoothed maximum score estimator fails a CLT condition, Assumption 2(d), introduced below.

[4]The criterion function $d(\overline{m}_T(\theta,\hat{\tau}), \hat{\gamma})$ is allowed to depend on *two* preliminary estimators $\hat{\tau}$ and $\hat{\gamma}$. In contrast, Burguete *et al.* (1982), Gallant (1987b), and Andrews and Fair (1988) allow the criterion function to depend on *one* preliminary estimator. In the latter papers there is no loss in generality from doing so. In the present paper, however, the use of two preliminary estimators allows one to simplify and weaken the assumptions. In particular, with two preliminary estimators, $\nu_T(\cdot)$ only has to be indexed by τ in Assumption 2(e) rather than by (τ,γ) and similarly in Assumption 2*(e).

The estimator $\hat{\theta}$ is required to solve (2.1) only with probability $\to 1$ to enable one to define the same estimator using different $m_t(\theta,\tau)$ and $d(m,\gamma)$ functions for the purposes of (i) consistency and (ii) asymptotic normality and testing. See footnote 7 below.

The infinite dimensional estimator $\hat{\tau}$ is only required to lie in $\mathcal{T}$ with probability $\to 1$, because $\mathcal{T}$ is taken below to contain elements that satisfy certain properties, e.g., smoothness properties. In many cases, not all realizations of $\hat{\tau}$ satisfy these properties, but the realizations in a set whose probability $\to 1$ do satisfy them.

In most examples the function $m_{Tt}(\cdot,\cdot,\cdot)$ does not depend on T . In some examples, however, such as spectral regression estimators and feasible GLS estimators for models with autocorrelation of unknown form, it must depend on T . For notational simplicity the possible dependence of $m_t(\theta,\tau)$ on T is not shown explicitly.

In cross–sectional applications the dimension k_{Tt} of W_{Tt} usually is finite and does not depend on T or t . In time series and panel data applications, on the other hand, it is sometimes convenient to allow k_{Tt} to be infinite or to depend on T or t .

[5]In contrast, the one–step GMM estimator considered by Newey (1987) is defined by

$$\hat{\theta} = \theta^* - \left[\tfrac{1}{T}\Sigma_1^T \hat{\Delta}(X_t)'\hat{\Omega}^{-1}(X_t)\tfrac{\partial}{\partial\theta'}\psi(Z_t, \theta^*)\right]^{-1} \tfrac{1}{T}\Sigma_1^T \hat{\Delta}(X_t)'\hat{\Omega}^{-1}(X_t)\psi(Z_t, \theta^*) ,$$

where θ^* is some preliminary $\sqrt{T}$–consistent estimator of θ_0. This estimator is a MINPIN estimator with $d(m,\gamma) = m'm/2$, $m_t(\theta,\tau) = \tau_1 - \tau_2\tau_3(X_t)\psi(Z_t, \tau_1) - \theta$ for $\tau = (\tau_1, \tau_2, \tau_3(\cdot))$, and $\hat{\tau} = (\hat{\tau}_1, \hat{\tau}_2, \hat{\tau}_3(\cdot))$
$= \left[\theta^*, \left[\frac{1}{T}\Sigma_1^T \hat{\Delta}(X_t)'\hat{\Omega}^{-1}(X_t)\frac{\partial}{\partial\theta'}\psi(Z_t, \theta^*)\right]^{-1}, \hat{\Delta}(\cdot)'\hat{\Omega}^{-1}(\cdot)\right]$. In this case,
$\tau_0 = (\tau_{10}, \tau_{20}, \tau_{30}(\cdot))$, $\tau_{10} = \theta_0 \in R^p$, $\tau_{20} = (E\Delta_0(X_t)'\Omega_0^{-1}(X_t)\Delta_0(X_t))^{-1} \in R^{p\times p}$, and $\tau_{30}(\cdot) = \Delta_0(\cdot)'\Omega_0^{-1}(\cdot)$. For brevity, we do not discuss the one–step GMM estimator further.

[6]If desired, $m_t(\theta,\tau)$ in (2.10) can be replaced by $m_t(\theta,\tau) = (Y_t - \tau_1(Z_t) - (X_t - \tau_2(Z_t))'\theta)^2/\tau_3(Z_t) - U_t^2/\tau_3(Z_t)$. This yields numerically the same estimator $\hat{\theta}$ as (2.10), but it allows a slight weakening of assumptions. For example, for Assumption 1 below to hold, it does not require $\lim_{T\to\infty} \frac{1}{T}\Sigma_1^T EU_t^2/\tau_3(X_t)$ to exist whereas (2.10) does. In other scenarios, this trick of subtracting off a term from $m_t(\theta,\tau)$ that does not depend on θ often can be used to weaken the requisite moment assumptions on the underlying rv's.

[7]A number of estimators, including LS estimators, M–estimators, and ML estimators, can be defined in terms of an underlying minimization problem (UMP) or in terms of the first order conditions (FOC) of this problem. Both definitions can be accommodated by the definition of an extremum estimator given in (2.1). Equations (2.10) and (2.11) exemplify this for the WLS estimator of PLR model. The choice between the UMP and FOC definitions depends primarily on Assumption 1(d). If Assumption 1(d) is satisfied using the FOC definition (i.e., (2.11) for the WLS estimator), then this is the most convenient definition because it must be used in any event for the asymptotic normality and testing results.

On the other hand, the limiting FOC may have multiple solutions even though the function $d(m(\theta,\tau_0), \gamma_0)$ that corresponds to the UMP definition has a unique solution. In this case, one needs to use the UMP definition to establish consistency and then redefine $m_t(\theta,\tau)$ and $d(m,\gamma)$ and use the FOC definition to establish asymptotic normality and testing results. If θ_0 is in the interior of Θ and $m_t(\theta,\tau)$ is differentiable in θ using the UMP definition, then an estimator that solves (2.1) using the UMP definition also solves (2.1) using the FOC definition with probability $\to 1$.

[8]Here and below, pseudo–metrics $\rho(\cdot,\cdot)$ are defined using a dummy variable N (rather than T) to avoid confusion when we consider objects such as $\text{plim}_{T\to\infty} \rho(\hat{\tau}, \tau_0)$. Note that the pseudo–metrics are assumed to be independent of the sample size T.

[9]Here and below, consistency of $\hat{\tau}$ for τ_0 "at rate $T^{1/4}$" or convergence of $\hat{\tau}$ to τ_0 "at rate $T^{1/4}$" means that $T^{1/4}(\hat{\tau} - \tau_0)$ converges in probability to zero in some sense (such as in L^Q).

[10]If necessary, the nonsingularity of HVH' can be avoided by using asymptotic distributional results for quadratic forms with g–inverted weighting matrices and singular limiting weight matrix – see Andrews (1987a).

[11]The estimator $\hat{\theta}$ also can be defined as a MINPIN estimator with $\hat{\gamma} = I$. With this definition, however, the likelihood ratio–like statistic of (6.1.11) does not have a chi–square asymptotic null distribution.

REFERENCES

Andrews, D. W. K. (1987a): "Asymptotic Results for Generalized Wald Tests," *Econometric Theory*, 3, 348–358.

_______ (1987b): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465–1472.

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

_______ (1989a): "Asymptotics for Semiparametric Econometric Models: II. Stochastic Equicontinuity," Cowles Foundation Discussion Paper No. 909R, Yale University.

_______ (1989b): "Generic Uniform Convergence," Cowles Foundation Discussion Paper No. 940, Yale University.

_______ (1989c): "Power in Econometric Applications," *Econometrica*, 57, 1059–1090.

_______ (1990a): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimators for Semiparametric Models," manuscript in preparation, Cowles Foundation, Yale University.

_______ (1990b): "Tests for Parameter Instability and Structural Change with Unknown Change Point," Cowles Foundation Discussion Paper No. 943, Yale University.

_______ (1991a): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59, forthcoming.

_______ (1991b): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, forthcoming.

Andrews, D. W. K. and R. C. Fair (1988): "Inference in Nonlinear Econometric Models with Structural Change," *Review of Economic Studies*, 55, 615–640.

Andrews, D. W. K. and J. C. Monahan (1990): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," Cowles Foundation Discussion Paper No. 942, Yale University.

Bates, C. E. and H. White (1988): "Determination of Estimators with Minimum Asymptotic Covariance Matrices," Department of Economics, University of California–San Diego.

Begun, J., W. Hall, W. Huang, and J. Wellner (1983): "Information and Asymptotic Efficiency in Parametric–Nonparametric Models," *Annals of Statistics*, 11, 432–452.

Bickel, P. J. (1982): "On Adaptive Estimation," *Annals of Statistics*, 10, 647–671.

Bickel, P. J., C. A. J. Klaasen, Y. Ritov, and J. A. Wellner (1988): *Efficient and Adaptive Inference in Semiparametric Models.* Baltimore: Johns Hopkins University Press, forthcoming.

Bierens, H. J. (1981): *Robust Methods and Asymptotic Theory.* Lecture Notes in Economics and Mathematical Systems No. 192. Berlin: Springer.

_______ (1987): "Kernel Estimators of Regression Functions," in *Advances in Econometrics: Fifth World Congress*, Vol. I, ed. by T. F. Bewley. New York: Cambridge University Press.

Burguete, J. F., A. R. Gallant, and G. Souza (1982): "On Unification of the Asymptotic Theory of Nonlinear Econometric Models," *Econometric Reviews*, 1, 151–190.

Carroll, R. J. (1982): "Adapting for Heteroscedasticity in Linear Models," *Annals of Statistics*, 10, 1224–1233.

Chamberlain, G. (1986): "Notes on Semiparametric Regression," unpublished manuscript, Department of Economics, Harvard University.

_______ (1987a): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–324.

_______ (1987b): "Efficiency Bounds for Semiparametric Regression," unpublished manuscript, Department of Economics, Harvard University.

Cosslett, S. R. (1983): "Distribution–free Maximum Likelihood Estimator of the Binary Choice Model," *Econometrica*, 51, 765–782.

Cox, D. D. (1988): "Approximation of Least Squares Regression on Nested Subspaces," *Annals of Statistics*, 16, 713–732.

Cox, D. R. (1975): "Partial Likelihood," *Biometrika*, 62, 269–276.

Delgado, M. A. (1988a): "Adaptive Minimum Distance Estimation under Heteroscedasticity of Unknown Nature," unpublished manuscript, London School of Economics.

_______ (1988b): "An Alternative Efficient Semiparametric Generalised Least Squares Estimator," unpublished manuscript, London School of Economics.

Devroye, L. and L. Györfi (1985): *Nonparametric Density Estimation: The* L_1 *View.* New York: Wiley.

Devroye, L. P. and T. J. Wagner (1980): "Distribution–free Consistency Results in Nonparametric Discrimination and Regression Function Estimation," *Annals of Statistics*, 8, 231–239.

Domowitz, I. and H. White (1982): "Misspecified Models with Dependent Observations," *Journal of Econometrics, 20, 35–58.*

Eubank, R. L. (1988): *Spline Smoothing and Nonparametric Regression.* New York: Marcel Dekker.

Gallant, A. R. (1987a): "Identification and Consistency in Seminonparametric Regression," in *Advances in Econometrics*, Fifth World Congress of the Econometric Society, Volume I, ed. by T. F. Bewley. Cambridge: Cambridge University Press.

_______ (1987b): *Nonlinear Statistical Models.* New York: Wiley.

Gallant, A. R. and H. White (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* New York: Basil Blackwell.

Grenander, U. (1981): *Abstract Inference.* New York: Wiley.

Györfi, L., W. Härdle, P. Sarda, and P. Vieu (1989): *Nonparametric Curve Estimation from Time Series.* New York: Springer–Verlag.

Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory and its Applications.* New York: Academic Press.

Han, A. K. (1987): "Non–parametric Analysis of a Generalized Regression Model," *Journal of Econometrics*, 35, 303–316.

Herrndorf, N. (1984): "An Invariance Principle for Weakly Dependent Sequences of Random Variables," *Annals of Probability*, 12, 141–153.

Horowitz, J. L. (1988): "Semiparametric M–estimation of Censored Linear Regression Models," *Advances in Econometrics*, 7, 45–83.

________ (1989): "A Smoothed Maximum Score Estimator for the Binary Response Model," unpublished manuscript, Department of Economics, University of Iowa.

Huber, P. J. (1967): "The Behaviour of Maximum Likelihood Estimates under Nonstandard Conditions," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 221–233.

________ (1973): "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *Annals of Statistics*, 1, 799–821.

Ichimura, H. (1985): "Semiparametric Least Squares Estimation of Single Index Models," unpublished manuscript, Department of Economics, University of Minnesota.

Ichimura, H. and L.–F. Lee (1990): "Semiparametric Estimation of Multiple Index Models," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, ed. by W. A. Barnett, J. L. Powell, and G. Tauchen. New York: Cambridge University Press, forthcoming.

Jennrich, R. (1969): "Asymptotic Properties of Nonlinear Least Squares Estimators," *Annals of Mathematical Statistics*, 40, 633–643.

Kim, J. and D. Pollard (1990): "Cube Root Asymptotics," *Annals of Statistics*, 18, 191–219.

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. and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Kreiss, J.–P. (1987): "On Adaptive Estimation in Stationary ARMA Processes," *Annals of Statistics*, 15, 112–133.

Laroque, G. and B. Salanie (1989): "Estimation of Multi–market Fix–price Models: An Application of Pseudo Maximum Likelihood Methods," *Econometrica*, 57, 831–860.

Lee, L.–F. (1989): "Semiparametric Maximum Profile Likelihood Estimation of Polytomous and Sequential Choice Models," Discussion Paper No. 253, Department of Economics, University of Minnesota.

Manski, C. F. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205–228.

_______ (1984): "Adaptive Estimation of Non–linear Regression Models," *Econometric Reviews*, 3, 145–194.

_______ (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 313–333.

McFadden, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995–1026.

McLeish, D. L. (1975a): "A Maximal Inequality and Dependent Strong Laws," *Annals of Probability*, 3, 826–836.

_______ (1975b): "Invariance Principles for Dependent Variables," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebeite*, 32, 165–178.

_______ (1977): "On the Invariance Principle for Nonstationary Mixingales," *Annals of Probability*, 5, 616–621.

Newey, W. K. (1987): "Efficient Estimation of Models with Conditional Moment Restrictions," unpublished manuscript, Department of Economics, Princeton University.

_______ (1988): "Two Step Estimation of Sample Selection Models," unpublished manuscript, Department of Economics, Princeton University.

_______ (1989a): "The Asymptotic Variance of Semiparametric Estimators," unpublished manuscript, Department of Economics, Princeton University.

_______ (1989b): "Uniform Convergence in Probability and Stochastic Equicontinuity," unpublished manuscript, Department of Economics, Princeton University.

_______ (1990a): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.

_______ (1990b): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99–135.

Newey, W. K. and J. L. Powell (1990): "Efficient Estimation of Linear and Type I Censored Regression Models under Conditional Quantile Restrictions," *Econometric Theory*, 6, forthcoming.

Newey, W. K. and K. D. West (1987): "A Simple Positive Semi–Definite, Hetero–skedasticty and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Pagan, A. and A. Ullah (1988): "The Econometric Analysis of Models with Risk Terms," *Journal of Applied Econometrics*, 3, 87–105.

Pakes, A. and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

Pollard, D. (1984): *Convergence of Stochastic Processes.* New York: Springer Verlag.

_______ (1985): "New Ways to Prove Central Limit Theorems," *Econometric Theory*, 1, 295–314.

Pötscher, B. M. and I. R. Prucha (1989): "A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes," *Econometrica*, 57, 675–683.

_______ (1990): "Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometric Models," unpublished manuscript, University of Maryland, College Park.

Powell, J. L. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325.

_______ (1986a): "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155.

_______ (1986b): "Symmetrically Trimmed Least Squares Estimators for Tobit Models," *Econometrica*, 54, 1435–1460.

_______ (1987): "Semiparametric Estimation of Bivariate Latent Variable Models," SSRI Working Paper No. 8704, University of Wisconsin, Madison.

_______ (1990): "Estimation of Monotonic Regression Models under Quantile Restrictions," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, ed. by W. A. Barnett, J. L. Powell, and G. Tauchen. New York: Cambridge University Press, forthcoming.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.

Prakasa Rao, B. L. S. (1983): *Nonparametric Functional Estimation.* Orlando, FL: Academic Press.

Robinson, P. M. (1987): "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 55, 875–891.

_______ (1988): "Root–N–Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

_______ (1989): "Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Review of Economic Studies*, 56, 511–534.

Ruppert, D. and R. J. Carroll (1980): "Trimmed Least Squares Estimation in the Linear Model," *Journal of the American Statistical Association*, 75, 828–838.

Severini, T. A. and W. H. Wong (1987a): "Convergence Rates of Maximum Likelihood and Related Estimates in General Parameter Spaces," unpublished manuscript, Department of Statistics, University of Chicago.

_______ (1987b): "Profile Likelihood and Semiparametric Models," unpublished manuscript, Department of Statistics, University of Chicago.

Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis.* New York: Chapman and Hall.

Stone, C. J. (1977): "Consistent Nonparametric Regression," *Annals of Statistics*, 5, 595–620.

_______ (1982): "Optimal Global Rates of Convergence for Nonparametric Regression," *Annals of Statistics*, 10, 1040–1053.

Wahba, G. (1990): *Spline Models in Statistics.* Regional Conference Series in Applied Mathematics. Philadelphia: Society for Industrial and Applied Mathematics.

Whang, Y.–J. and D. W. K. Andrews (1990): "Tests of Model Specification for Parametric and Semiparametric Models," manuscript in preparation, Cowles Foundation, Yale University.

White, H. (1984): *Asymptotic Theory for Econometricians.* Orlando, FL: Academic Press.

Wooldridge, J. M. and H. White (1988a): "Central Limit Theorems for Dependent, Heterogeneous Processes with Trending Moments," Department of Economics, M.I.T.

_______ (1988b): "Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes," *Econometric Theory*, 4, 210–230.

_______ (1990): "Some Results on Sieve Estimators with Dependent Observations," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, ed. by W. A. Barnett, J. L. Powell, and G. Tauchen. New York: Cambridge University Press, forthcoming.