GENERALIZED METHOD OF MOMENTS ESTIMATION WHEN A PARAMETER IS ON A BOUNDARY

BY

DONALD W. K. ANDREWS

COWLES FOUNDATION PAPER NO. 1131

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2006

http://cowles.econ.yale.edu/

Generalized Method of Moments Estimation When a Parameter Is on a Boundary

Donald W. K. ANDREWS
Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520-8281
(*donald.andrews@yale.edu*)

This article establishes the asymptotic distributions of generalized method of moments (GMM) estimators when the true parameter lies on the boundary of the parameter space. The conditions allow the estimator objective function to be nonsmooth and to depend on preliminary estimators. The boundary of the parameter space may be curved and/or kinked. The article discusses three examples: (1) instrumental variables (IV) estimation of a regression model with nonlinear equality and/or inequality restrictions on the parameters; (2) method of simulated moments estimation of a multinomial discrete response model with some random coefficient variances equal to 0, some random effect variances equal to 0, or some measurement error variances equal to 0; and (3) semiparametric least squares estimation of a partially linear regression model with nonlinear equality and/or inequality restrictions on the parameters.

KEY WORDS: Asymptotic distribution; Boundary; Equality restriction; Inequality restriction; Partially linear model; Semiparametric estimator.

1. INTRODUCTION

Hansen's (1982) seminal article on generalized method of moments (GMM) estimation provides sufficient conditions for the asymptotic normality of GMM estimators. A key assumption is that the parameter lies in the interior of the parameter space. Pakes and Pollard (1989) extended the results of Hansen to allow for cases in which the sample moment conditions are not smooth functions of θ. In addition, they allowed the weight matrix to depend on θ and hence cover estimators now referred to as *continuously updated* GMM estimators (CUE). Like Hansen (1982), Pakes and Pollard (1989) assumed that the true parameter lies in the interior of the parameter space.

This article extends the results of Hansen (1982) and Pakes and Pollard (1989) to cover cases in which the true parameter is not in the interior of the parameter space. In such cases, the asymptotic distribution of the GMM estimator is no longer normal. The results given here allow for nonsmooth sample moments, the CUE, and sample moments that depend on preliminary finite- or infinite-dimensional preliminary nuisance parameters.

The results of this article are obtained by applying more general results for extremum estimators given by Andrews (1999). The latter rely on high-level assumptions. This article provides a number of (more) primitive sufficient conditions for the case of GMM estimators. These sufficient conditions are used to determine the asymptotic distributions of GMM estimators under primitive conditions in three examples.

The approach used by Andrews (1999) is to approximate the estimator objective function by a quadratic function rather than rely on first-order conditions. This approach was used by Chernoff (1954) to establish the asymptotic distribution of the likelihood ratio test in iid models with smooth likelihoods when the true parameter may be on a boundary. It has also been used by various authors to obtain the asymptotic properties of estimators when the true parameter is in the interior of the parameter space (see, e.g., Le Cam 1960; Jeganathan 1982; Pollard 1985; Pakes and Pollard 1989).

A number of papers in the literature also consider the asymptotic properties of estimators and tests when the true parameter lies on the boundary of the parameter space (Chernoff 1954; Aitchison and Silvey 1958; Moran 1971; Chant 1974; Self and Liang 1987; Gourieroux and Monfort 1989, chap. 21; Geyer 1994; Wang 1996; Andrews 1999, 2001). Andrews (1999, 2001) has given brief discussions of these papers. The results here are most similar to those of Geyer (1994), who considered estimators that minimize a sample average depending on a parameter θ when the true parameter θ_0 lies on a boundary of the parameter space.

The method of establishing the asymptotic distribution of the GMM estimator when the true parameter is on a boundary is now outlined. Let $\hat{\theta}$ be the GMM estimator that minimizes the GMM criterion function $L_T(\theta)$ over the parameter space $\Theta \subset R^s$. First, one shows that the GMM criterion function equals a quadratic function,

$$q_T(T^{1/2}(\theta-\theta_0)) = (T^{1/2}(\theta-\theta_0)-Z_T)'\mathcal{J}(T^{1/2}(\theta-\theta_0)-Z_T), \quad (1)$$

plus a term that does not depend on θ and another term that depends on θ but is sufficiently small so that it does not affect the asymptotics. In (1), $\mathcal{J}$ is a nonsingular matrix and Z_T is an asymptotically normal random vector. In particular, suppose that $G_T(\theta)$ is the vector of sample moments that defines the GMM estimator, M is the limit of the GMM weight matrix, and $\Gamma = \text{p}\lim_{T\to\infty}(\partial/\partial\theta')G_T(\theta_0)$. Then

$$\mathcal{J} = \Gamma' M \Gamma \qquad \text{and} \qquad Z_T = \mathcal{J}^{-1}\Gamma' M T^{1/2} G_T(\theta_0). \quad (2)$$

Journal of Business & Economic Statistics
October 2002, Vol. 20, No. 4
DOI 10.1198/073500102288618667

Under suitable conditions, a central limit theorem (CLT) implies that

$$T^{1/2}G_T(\theta_0) \to_d N(0, \mathcal{V}) \quad \text{and} \quad Z_T \to_d Z \sim N(0, \mathcal{J}^{-1}\Gamma' M\mathcal{V}M\Gamma\mathcal{J}^{-1}) \text{ as } T \to \infty. \tag{3}$$

Combining (1) and (3) gives

$$q_T(\lambda) \to_d q(\lambda) \text{ for each } \lambda \in R^s, \quad \text{where} \quad q(\lambda) = (\lambda - Z)'\mathcal{J}(\lambda - Z). \tag{4}$$

If the GMM estimator $\hat{\theta}$ is $T^{1/2}$ consistent, then the only part of the parameter space Θ that effects the asymptotic distribution of $\hat{\theta}$ is the part of Θ near the true parameter θ_0. Equivalently, the only part of the shifted parameter space $\Theta - \theta_0$ that is relevant is the part near the origin. The case is considered where the shifted parameter space $\Theta - \theta_0$ can be approximated near the origin by a convex cone Λ. The approximation concept used is due to Chernoff (1954).

Now, by definition, $\hat{\theta}$ minimizes $L_T(\theta)$ over Θ. Hence the normalized estimator $\hat{\lambda}_T = T^{1/2}(\hat{\theta} - \theta_0)$ minimizes $L_T(\theta_0 + \lambda/T^{1/2})$ over $T^{1/2}(\Theta - \theta_0)$. Asymptotically, $L_T(\theta_0 + \lambda/T^{1/2})$ behaves like $q_T(\lambda)$ by (1), $q_T(\lambda)$ behaves like $q(\lambda)$ by (4), and $T^{1/2}(\Theta - \theta_0)$ behaves like Λ for θ near θ_0. Consequently, $\hat{\lambda}_T = T^{1/2}(\hat{\theta} - \theta_0)$ can be shown to behave asymptotically like the minimizer $\hat{\lambda}$ of $q(\lambda)$ over $\lambda \in \Lambda$. That is,

$$T^{1/2}(\hat{\theta} - \theta_0) \to_d \hat{\lambda}, \quad \text{where} \quad \hat{\lambda} \text{ minimizes } q(\lambda) = (\lambda - Z)'\mathcal{J}(\lambda - Z) \text{ over } \Lambda. \tag{5}$$

Typically, the convex cone Λ is defined by linear equality and/or inequality constraints, and $\hat{\lambda}$ is the solution to a quadratic programming problem. Consequently, the asymptotic distribution of $T^{1/2}(\hat{\theta} - \theta_0)$ can be simulated using standard GAUSS or Matlab programs quite quickly and easily.

If only one element of θ_0 is on a boundary and it is constrained from above or below, then the corresponding element of $\hat{\theta}$ has a half-normal asymptotic distribution. In this case the asymptotic distributions of the other parameters are affected and are nonnormal unless a block diagonality condition is satisfied.

Note that if θ_0 lies in the interior of Θ, then

$$\Lambda = R^s, \quad \hat{\lambda} = Z, \quad \text{and} \quad T^{1/2}(\hat{\theta} - \theta_0) \to_d Z \sim N(0, \mathcal{J}^{-1}\Gamma' M\mathcal{V}M\Gamma\mathcal{J}^{-1}), \tag{6}$$

which is the result of Hansen (1982) and Pakes and Pollard (1989).

Note that the assumptions used here are such that one often can use existing results in the literature (that are designed for the case where the true parameter is an interior point) to help verify the assumptions. This is particularly useful for semiparametric estimators. One does not need to reprove results regarding the effect of preliminary nonparametric estimators on the properties of the estimator objective function. This is illustrated in an example.

The results given here cover both minimum distance (MD) and GMM estimators. Details are given later.

The results of this article cover models with deterministic time trends via the method of Andrews and McDermott (1995). But the results do not cover models with stochastic trends. (See Andrews 1999 for results that do cover such cases.)

Three examples are considered in this article. The first example is a linear regression model estimated by instrumental variables (IV) with nonlinear equality and/or inequality restrictions on the regression parameters. The errors, regressors, and IVs are assumed to be iid. This example exhibits curved and kinked boundaries to the parameter space. Nonlinear inequality restrictions arise in utility, cost, and profit function estimation when convexity, quasi-convexity, concavity, or quasi-concavity is imposed at some point(s) in the sample (see Gallant and Golub 1984).

The second example is a multinomial discrete response model estimated via a method of simulated moments (MSM) estimator as done by McFadden (1989) and Pakes and Pollard (1989). Considered here is the case where the model includes random coefficients (as in Hausman and Wise 1978), random effects (as in McFadden 1989), or measurement error (as in McFadden 1989), and the variances of some of these random terms are 0. This example illustrates the case of an estimator objective function that is discontinuous. A related class of GMM estimators of discrete response models with random coefficients used in the industrial organization literature is that of Berry (1994) and Berry, Levinsohn, and Pakes (1995). The results given in this article could also be applied to determine the asymptotic distribution of these estimators when some of the random coefficient variances are 0.

The third example is a partially linear model estimated by the semiparametric least squares (LS) estimator of Robinson (1988), but subject to nonlinear equality and/or inequality constraints. This example illustrates the application of the general results to a semiparametric estimator and to an estimator that depends on a preliminary estimator. The example shows that one can derive the limit distribution for the estimator when the parameter is on the boundary with very little additional work beyond that which is needed to establish its distribution when the true parameter is in the interior of the parameter space. In particular, the hard parts of the verification of the assumptions follow directly from the results of Robinson (1988) with no additional work.

Note that the results of the article apply to parametric two-step estimators, although none of the examples herein are of this form. For example, consider the Heckman two-step estimator of a sample selection model. If the correlation between the errors in the two equations of the model is generated by a common random effect, then the coefficient on the selection bias correction term in the main equation is necessarily nonnegative. In this case the regression parameter of the main equation is on the boundary of the parameter space when true random effect variance is 0, which corresponds to the case where there is no selectivity bias. The results herein apply to this case.

The rest of the article is organized as follows. Section 2 introduces the three examples. Section 3 presents the general results. Section 4 applies the general results to the IV regression example. Section 5 applies the general results to the multinomial response example. Section 6 does likewise for the partially linear regression example. An Appendix provides proofs of the results.

All limits are taken "as $T \to \infty$" unless stated otherwise. Let "for all $\gamma_T \to 0$" abbreviate "for all sequences of positive scalar constants $\{\gamma_T : T \geq 1\}$ for which $\gamma_T \to 0$." Let $\to_p$ and $\to_d$ denote convergence in probability and distribution. Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and largest eigenvalues of a matrix A. Let $\mathrm{cl}(\Lambda)$ denote the closure of a set Λ. Let $S(\theta, \varepsilon)$ denote an open sphere centered at θ with radius ε. Let $C(\theta, \varepsilon)$ denote an open cube centered at θ with sides of length 2ε.

2. EXAMPLES

This section introduces three examples. These examples cover parameter spaces with linear and nonlinear boundaries. One example has a GMM criterion function that is not smooth; another has a GMM criterion function that depends on a preliminary infinite-dimensional nuisance parameter.

2.1 Instrumental Variables Regression With Restricted Parameters

The first example is an iid linear regression model estimated by IVs with equality and/or inequality restrictions on the regression parameters. Both linear and nonlinear restrictions are considered. The asymptotic distribution of the IV estimator (subject to the restrictions) when some of the inequality restrictions are satisfied as equalities is determined. In this case the true regression parameter is on a linear or nonlinear boundary of the parameter space.

The model is

$$Y_t = X_t'\theta + \varepsilon_t, \tag{7}$$

where $\{(Y_t, X_t, Z_t) : t \leq T\}$ are observed iid dependent, regressor, and instrumental variables; $\{\varepsilon_t : t \leq T\}$ are iid unobserved errors; $X_t, \theta \in R^s$; $Y_t, \varepsilon_t \in R$; and $Z_t \in R^k$. The regressors may be correlated with the errors.

The parameter θ is assumed to satisfy the restrictions

$$m_a(\theta) = \mathbf{0}, \qquad m_b(\theta) \leq \mathbf{0}, \qquad \text{and} \qquad h(\theta) \leq \mathbf{0}, \tag{8}$$

where $m_j(\cdot) \in R^{c_j}$ for $j = a, b$ and $h(\cdot)$ is vector-valued.

Suppose that the true parameter θ_0 satisfies the restrictions of (8) with

$$m_a(\theta_0) = \mathbf{0}, \qquad m_b(\theta_0) = \mathbf{0}, \qquad \text{and} \qquad h(\theta_0) < \mathbf{0}. \tag{9}$$

Thus θ_0 is on the part of the boundary of the parameter space, Θ, that is determined by $m_a(\cdot)$ and $m_b(\cdot)$. The estimator considered here is an IV estimator that satisfies the restrictions.

2.2 Multinomial Response Model

This example is a multinomial response model estimated by the method of simulated moments (MSM). The notation is the same as used by McFadden (1989) and Pakes and Pollard (1989). The tth individual has m alternatives to choose between. The ℓth choice is associated with a utility (or profit) of $Z_{t\ell}'h(\eta_t, \theta)$ for $\ell = 1, \ldots, m$, where $Z_{t\ell}$ is a b-vector of covariates for the ℓth choice and the tth individual, $\theta \in R^s$ is an unknown parameter, $\eta_t \in R^r$ is a vector of errors with known distribution, and $h(\cdot, \cdot)$ is a known R^b-valued function. The tth individual chooses the alternative with the greatest utility. Thus the response vector $d_t \in \{0, 1\}^m$ can be written as

$$d_t = D(Z_t h(\eta_t, \theta)), \qquad \text{where } Z_t = [Z_{t1} \cdots Z_{tm}]' \in R^{m \times b} \tag{10}$$

and $D(\cdot) : R^m \to \{0, 1\}^m$ puts a 1 in the position of the largest component and a 0 elsewhere. The choice is indicated by the component with a 1. That there is zero probability of a tie is assumed. The random variables $\{(Z_t, \eta_t) : t = 1, \ldots, T\}$ are assumed to be iid.

By specifying different covariate vectors $Z_{t\ell}$, error vectors η_t, parameter vectors θ, and functional forms $h(\cdot, \cdot)$, a variety of different models is obtained. For example, if η_t has a standard multivariate normal distribution, then the model is in the family of probit models. Several such models are considered here. The first model is a random coefficient probit model considered by Hausman and Wise (1978) and, more recently, by Horowitz (1993), among others. In this case the element of $h(\eta_t, \theta)$ that corresponds to a covariate in $Z_{t\ell}$ whose coefficient is random is of the form $\theta_i + \theta_j^{1/2}\eta_t \sim N(\theta_i, \theta_j)$. The case where p (≥ 1) random coefficient variances are zero and hence the true parameter θ_0 is on the boundary of the parameter space is considered. For notational convenience, the parameters are ordered such that the first p elements of θ are these parameters.

The second model is a binary probit panel data model with autocorrelated errors and random effects (see McFadden 1989, sec. 5). Let the first element of θ be the proportion of the total error variance due to the random effect. The case where the true proportion is zero and hence the parameter θ_0 is on the boundary of the parameter space is considered.

The third model is a probit model with measurement error on some covariates (see McFadden 1989, sec. 6). In this case some of the elements of θ correspond to the variance parameters of the measurement errors. The case where p (≥ 1) of these variance parameters equal zero and θ_0 is on a boundary is analyzed. As before, the elements of θ are ordered such that these are the first p elements of θ.

All of the foregoing cases can be treated simultaneously by analyzing the general multinomial response model introduced earlier with a parameter vector θ whose first p or more elements must be nonnegative and whose true value θ_0 has its first p elements equal to 0.

2.3 Partially Linear Regression Model

This example is a partially linear regression model with nonlinear equality and/or inequality restrictions on the parameter vector. The partially linear model is a semiparametric model.

Estimation of the finite-dimensional parameter of the model is considered using a semiparametric least squares (LS) method introduced by Robinson (1988), who considered the partially linear regression model without any restrictions. The model is defined and the same assumptions are used as in Robinson (1988). In fact, Robinson's results can be used to establish the only difficult parts of the proof of the asymptotic distribution of the semiparametric LS estimator under nonlinear equality and/or inequality restrictions.

The model is

$$Y_t = X_t'\theta + \mu(Z_t) + \varepsilon_t, \tag{11}$$

where $\{(Y_t, X_t, Z_t) : t = 1, 2, \ldots, T\}$ are the observed random variables, θ is the unknown parameter to be estimated, $\mu(\cdot)$ is an unknown function, and ε_t is an unobserved error. Following Robinson (1988), (Y_t, X_t, Z_t) are assumed to be iid across t, $E\varepsilon_t = 0$, and ε_t is independent of (X_t, Z_t).

The parameter θ is assumed to satisfy the same nonlinear restrictions as in (8) of the IV example. In addition, the true parameter θ_0 is assumed to satisfy (9). In this case the parameter θ_0 is on the boundary of the parameter space.

3. GENERAL RESULTS

3.1 Definition of the Generalized Method of Moments Estimator

Let $\mathbf{Y}_T = \{Y_t : t = 1, \ldots, T\}$ denote the sample with sample size T for $T = 1, 2, \ldots$. Considered here is a GMM objective function $L_T(\theta)$ that depends on $\mathbf{Y}_T$,

$$L_T(\theta) = \|A_T(\theta) G_T(\theta)\|^2/2, \tag{12}$$

where $G_T(\theta) : \Theta \to R^k$ is a vector of sample moments, $A_T(\theta) : \Theta \to R^{k\times k}$ is a random weight matrix, and $\|\cdot\|$ denotes the Euclidean norm. (The division by 2 is strictly for convenience. It eliminates some constants in the formulas that follow.) The random variables $G_T(\theta)$ and $A_T(\theta)$ are normalized such that each is $O_p(1)$, but not $o_p(1)$ [except $G_T(\theta_0)$, which is $O_p(T^{-1/2})$]. Note that the notation for GMM estimators is the same as that of Pakes and Pollard (1989), but with the sample size given by T rather than n.

Typically, the vector of moment conditions $G_T(\theta)$ is of the form

$$G_T(\theta) = T^{-1}\sum_{t=1}^{T} g(Y_t, \theta), \text{ where } Eg(Y_t, \theta_0) = 0, \tag{13}$$

$g(Y_t, \cdot) : \Theta \to R^k$ is a known function, and θ_0 is the true value. But $G_T(\theta)$ could also depend on some finite- or infinite-dimensional preliminary estimator. For example, $G_T(\theta)$ could be of the form

$$G_T(\theta) = T^{-1}\sum_{t=1}^{T} g(Y_t, \theta, \hat{\tau}) \quad \text{or}$$

$$G_T(\theta) = T^{-1}\sum_{t=1}^{T} g(Y_t, \theta, \hat{\tau}(Y_t)), \tag{14}$$

where $\hat{\tau}$ denotes a finite-dimensional preliminary estimator; $\hat{\tau} \to_p \tau_0$ for some $\tau_0 \in R^{d_\tau}$; $Eg(Y_t, \theta_0, \tau_0) = 0$, $\hat{\tau}(Y_t)$ denotes an infinite-dimensional estimator evaluated at Y_t, such as a nonparametric regression or density estimator; $\hat{\tau}(y) \to_p \tau_0(y) \in R^{d_\tau}$ for all y in the support of Y_t; and $Eg(Y_t, \theta_0, \tau_0(Y_t)) = 0$. For example, the partially linear regression example uses preliminary nonparametric regression estimators.

The framework considered here covers minimum distance (MD) estimators as well as GMM estimators. The MD objective function is the same as the GMM objective function, except that $G_T(\theta)$ is not a vector of moment conditions, but rather the difference between an unrestricted estimator $\hat{\xi}_T$ of a parameter ξ_0 and a vector of restrictions $h(\theta)$ on ξ_0. That is,

$$G_T(\theta) = \hat{\xi}_T - h(\theta), \text{ where } \xi_0 = h(\theta_0) \tag{15}$$

and θ_0 is the true value. Henceforth, for simplicity, the estimator and criterion function are called the GMM estimator and criterion function, but it should be understood that they also could be a MD estimator and criterion function.

By definition, the GMM estimator $\hat{\theta}$ satisfies $\hat{\theta} \in \Theta$ and

$$L_T(\hat{\theta}) = \inf_{\theta\in\Theta} L_T(\theta) + o_p(1). \tag{16}$$

That is, $\hat{\theta}$ minimizes the GMM criterion function over Θ up to some term that is $o_p(1)$ (which allows for some tolerance in the numerical minimization).

3.2 Assumptions

First, consistency of $\hat{\theta}$ is considered. The same methods for establishing consistency can be used whether $\hat{\theta}$ is on a boundary of the parameter space or is in the interior. Hence $\hat{\theta}$ is assumed to be consistent, and a standard sufficient condition is provided for this result.

Assumption GMM1. $\hat{\theta} = \theta_0 + o_p(1)$, where $\theta_0 \in \mathrm{cl}(\Theta)$.

A well-known sufficient condition for assumption GMM1 is the following.

Assumption GMM1.* (a) For some function $L(\theta) : \Theta \to R$, $\sup_{\theta\in\Theta} |L_T(\theta) - L(\theta)| \to_p 0$.

(b) For all $\varepsilon > 0$, $\inf_{\theta\in\Theta/S(\theta_0,\varepsilon)} L(\theta) > L(\theta_0)$, where $\Theta/S(\theta_0, \varepsilon)$ denotes all vectors θ in Θ but not in $S(\theta_0, \varepsilon)$.

Note that here and below that a superscript *, 2*, or 3* on an assumption denotes that the assumption is sufficient (sometimes only in the presence of other specified assumptions) for the unsuperscripted assumption.

Alternatively, assumption GMM1 can be established using theorem 3.1 and lemma 3.4 of Pakes and Pollard (1989), which is applicable even if θ_0 is on the boundary of Θ.

Next, the basic assumptions used to derive the asymptotic distribution of the GMM estimator are stated. These assumptions are quite similar to the assumptions used by Pakes and Pollard (1989), except they do not require that the true value θ_0 is in the interior of the parameter space. In particular, the assumptions allow for nonsmooth sample moments, such as sample moments that depend on indicator functions. The assumptions also allow for sample moments that depend on preliminary estimators.

Assumption GMM2. (a) For some nonrandom function $G(\theta)$, $G_T(\theta) \to_p G(\theta)\ \forall\, \theta \in \Theta \cap S(\theta_0, \varepsilon)$ for some $\varepsilon > 0$.

(b) $G(\theta) = G(\theta_0) + \Gamma(\theta - \theta_0) + o(\|\theta - \theta_0\|)$ as $\|\theta - \theta_0\| \to 0$ for $\theta \in \Theta \cap S(\theta_0, \varepsilon)$ for some $\varepsilon > 0$, where Γ is a nonrandom $k \times s$ matrix with full column rank s ($\leq k$).

(c) $G(\theta_0) = \mathbf{0}$.

(d) For all $\gamma_T \to 0$, $\sup_{\theta \in \Theta: \|\theta - \theta_0\| \leq \gamma_T} T^{1/2}\|G_T(\theta) - G(\theta) - G_T(\theta_0)\|/(1 + T^{1/2}\|\theta - \theta_0\|) = o_p(1)$ under θ_0.

(e) For some nonrandom nonsingular $k \times k$ matrix A and for all $\gamma_T \to 0$, $\sup_{\theta \in \Theta: \|\theta - \theta_0\| \leq \gamma_T} \|A_T(\theta) - A\| = o_p(1)$.

Assumption GMM3. $T^{1/2} G_T(\theta_0) \to_d \widetilde{G} \sim N(\mathbf{0}, \mathcal{V})$ for some nonrandom matrix $\mathcal{V}$.

Section 3.3 provides sufficient conditions for assumption GMM2. Here assumptions GMM2 and GMM3 are discussed and how they are verified in the most regular cases is illustrated.

Assumption GMM2(a) can be verified using a pointwise weak law of large numbers (WLLN) when $G_T(\theta)$ is of the form in (13). Assumption GMM2(a) is used because it serves to define the limit function $G(\theta)$. It is not actually used in any of the proofs. Any function $G(\theta)$ that satisfies assumption GMM2(d) could be used to define $G(\theta)$. There is little to be gained by this, however, because it is hard to imagine a case for which assumption GMM2(d) holds for a function $G(\cdot)$ that does not satisfy assumption GMM2(a).

Assumption GMM2(b) holds if $G(\theta)$ is differentiable at θ_0. In this case, $\Gamma = (\partial/\partial\theta')G(\theta_0)$. This requires that $G(\theta)$ is defined on a neighborhood of θ_0. Section 3.3 provides sufficient conditions for assumption GMM2(b) when $G(\theta)$ is not defined on a neighborhood of θ_0 using left/right partial derivatives.

Assumption GMM2(c) holds if the moment conditions are correctly specified in the GMM context or if the parameter ξ_0 satisfies the restrictions $\xi_0 = h(\theta_0)$ in the MD context.

Assumption GMM2(d) is a stochastic equicontinuity condition. It can be verified using the empirical process results of Pakes and Pollard (1989), van der Vaart and Wellner (1996), or Andrews (1994). In such cases, the condition is actually verified with the denominator "$1 + T^{1/2}\|\theta - \theta_0\|$" replaced by "1."

Alternatively, if $G_T(\theta)$ is smooth in θ, then assumption GMM2(d) can be verified easily with the denominator "$1 + T^{1/2}\|\theta - \theta_0\|$" replaced by "$T^{1/2}\|\theta - \theta_0\|$." To see this, suppose that $G_T(\theta)$ is differentiable at θ_0 with derivative matrix $(\partial/\partial\theta')G_T(\theta)$ that satisfies the following: for all $\gamma_T \to 0$,

$$\sup_{\theta \in \Theta: \|\theta - \theta_0\| \leq \gamma_T} \left\| \frac{\partial}{\partial\theta'} G_T(\theta) - \Gamma \right\| = o_p(1). \tag{17}$$

[This condition can be verified using a uniform WLLN and continuity of the probability limit of $(\partial/\partial\theta')G_T(\theta)$ at θ_0.]

Then, applying the mean-value theorem element by element yields

$$\begin{aligned} G_T(\theta) &= G_T(\theta_0) + \frac{\partial}{\partial\theta'} G_T(\theta^\dagger)(\theta - \theta_0) \\ &= G_T(\theta_0) + \Gamma(\theta - \theta_0) + o_p(\|\theta - \theta_0\|) \end{aligned} \tag{18}$$

uniformly over $\{\theta \in \Theta : \|\theta - \theta_0\| \leq \gamma_T\}$, where $\theta^\dagger$ lies between θ and θ_0 and may differ across rows. Combining this result with assumptions GMM2(b) and GMM2(c) gives

$$\|G_T(\theta) - G_T(\theta_0) - G(\theta)\| = o_p(\|\theta - \theta_0\|) \tag{19}$$

uniformly over $\{\theta \in \Theta : \|\theta - \theta_0\| \leq \gamma_T\}$. This immediately implies assumption GMM2(d) using the "$T^{1/2}\|\theta - \theta_0\|$" part of the denominator in assumption GMM2(d).

In summary, if assumptions GMM2(a)–(c) hold and $G_T(\theta)$ is differentiable in a neighborhood of θ_0 with partial derivative matrix that satisfies (17), then assumption GMM2(d) holds.

The above verification of assumption GMM2(d) using smoothness relies on $G_T(\theta)$ being defined on a neighborhood of θ_0 (to define the derivative of $G_T(\theta)$ at θ_0). Later it is shown that assumption GMM2(d) can be verified using smoothness even when $G_T(\theta)$ is not defined on a neighborhood of θ_0 by using left/right partial differentiability of $G_T(\theta)$. Furthermore, assumption GMM2(d) can be verified without smoothness of $G_T(\theta)$. This is illustrated in the multinomial response example.

Assumption GMM2(e) requires that the weight matrix $A_T(\theta)$ is well behaved. Typically, it can be verified using a uniform WLLN and continuity of the probability limit of $A_T(\theta)$ at θ_0.

Note that an asymptotically optimal choice of the weight matrix for the GMM estimator when θ_0 is not on a boundary is such that $A'A = \mathcal{V}^{-1}$.

Assumption GMM3 is verified when $G_T(\theta)$ is as in (13) by the application of a CLT. It can be verified when $G_T(\theta)$ is as in the first case of (14) by (a) taking a mean-value expansion of $T^{1/2}G_T(\theta_0)$ about $\hat{\tau}$, (b) using the asymptotic normality of $T^{1/2}(\hat{\tau} - \tau_0)$, and (c) applying a CLT to $T^{-1/2}\sum_{t=1}^{T} g(Y_t, \theta, \tau_0)$. Assumption GMM3 can be verified when $G_T(\theta)$ is as in the second case of (14) by applying results in the literature for sample averages that depend on preliminary infinite-dimensional nuisance parameters (see, e.g., Andrews 1991, Newey 1994).

The stochastic equicontinuity condition of assumption GMM2(d) is quite similar to, but slightly different from, that used by Pakes and Pollard (1989). The following result, however, shows that it is equivalent to that of Pakes and Pollard (1989) given the other assumptions. Assumption GMM2(d) is stated as is, because it is in the most convenient form for verification.

Lemma 1. Under assumptions GMM2 and GMM3 except GMM2(d), Assumption GMM2(d) is equivalent to each of the following two conditions: For all $\gamma_T \to 0$,

(a) $\sup_{\theta \in \Theta: \|\theta - \theta_0\| \leq \gamma_T} \|G_T(\theta) - G(\theta) - G_T(\theta_0)\|/(T^{-1/2} + \|G(\theta)\|) = o_p(1)$ and

(b) $\sup_{\theta \in \Theta: \|\theta - \theta_0\| \leq \gamma_T} \|G_T(\theta) - G(\theta) - G_T(\theta_0)\|/(T^{-1/2} + \|G_n(\theta)\| + \|G(\theta)\|) = o_p(1)$.

Comment. Condition (b) of Lemma 1 is the same as that of Pakes and Pollard [1989, condition (iii) of thm. 3.3].

3.3 Sufficient Conditions for Assumptions GMM2 and GMM3

This section states several sufficient conditions for assumptions GMM2 and GMM3. First, it provides sufficient conditions for assumptions GMM2 and GMM3 that are essentially the same as Hansen's (1982) conditions for asymptotic normality of the GMM estimator, except that θ_0 need not be an interior point of Θ. These conditions rely on smoothness of $G_T(\theta)$ in a neighborhood of θ_0 and use the argument of (17)–(19) to establish assumption GMM2(d).

*Assumption GMM2**. (a) $\{Y_t : t = \ldots, 0, 1, \ldots\}$ is stationary and ergodic.
(b) $G_T(\theta) = T^{-1}\sum_{t=1}^T g(Y_t, \theta)$ for some function $g(\cdot, \cdot)$.
(c) For some $\varepsilon > 0$, $g(y, \theta)$ is continuously differentiable in θ on $S(\theta_0, \varepsilon)$ for y in the support of Y_t, $E\|g(Y_t, \theta)\| < \infty$ for $\theta \in S(\theta_0, \varepsilon)$, and $E\sup_{\theta\in S(\theta_0,\varepsilon)} \|(\partial/\partial\theta')g(Y_t, \theta)\| < \infty$.
(d) $\Gamma = E(\partial/\partial\theta')g(Y_t, \theta_0)$ is full column rank.
(e) $Eg(Y_t, \theta_0) = 0$.
(f) Assumption GMM2(e) holds.

Define

$$W_t = g(Y_t, \theta_0) \quad \text{and}$$
$$v_j = E(W_0|W_{-j}, W_{-j-1}, \ldots) - E(W_0|W_{-j-1}, W_{-j-2}, \ldots) \text{ for } j \geq 0. \tag{20}$$

*Assumption GMM3**. (a) $E\|g(Y_t, \theta_0)\|^2 < \infty$.
(b) $E(W_0|W_{-j}, W_{-j-1}, \ldots)$ converges in mean square to 0 as $j \to \infty$.
(c) $\sum_{j=0}^{\infty} E\|v_j\| < \infty$.

Next, simple smoothness conditions are provided that are sufficient for Assumption GMM2 and that apply even when $G(\theta)$ and $G_T(\theta)$ are not defined on a neighborhood of θ_0. Some terminology must be introduced. Let f be a function whose domain includes $\mathcal{X} \subset R^s$. Let $a \in \mathcal{X}$. A Taylor expansion of $f(x)$ about $f(a)$ to hold for points $x \in \mathcal{X}$ is desired. It is supposed that $\mathcal{X} - a$ equals the intersection of a union of orthants and an open cube, $C(\mathbf{0}, \varepsilon)$, centered at $\mathbf{0}$ with edges of length 2ε for some $\varepsilon > 0$. (Thus $\mathcal{X} - a$ is locally equal to a union of orthants.) As defined, $\mathcal{X}$ is a cube centered at a with some "orthants" of the cube removed.

The function f is said to have *left/right* (l/r) *partial derivatives* (of order 1) *on* $\mathcal{X}$ if it has partial derivatives at each interior point of $\mathcal{X}$; if it has partial derivatives at each boundary point of $\mathcal{X}$ with respect to coordinates that can be perturbed to the left and right; and if it has left (right) partial derivatives at each boundary point of $\mathcal{X}$ with respect to coordinates that can be perturbed only to the left (right). Note that the shape of $\mathcal{X}$ is such that $\forall x \in \mathcal{X}$ and for all coordinates x_j of x, it is possible to perturb x_j to the right or left or both and stay within $\mathcal{X}$. Thus it is possible to define the left, the right, or the two-sided partial derivative of f with respect to x_j at x $\forall j \leq s$ and $\forall x \in \mathcal{X}$.

The function f is said to have l/r partial derivatives of order k on $\mathcal{X}$ for $k \geq 2$ if f has l/r partial derivatives of order $k-1$ on $\mathcal{X}$ and each of the latter has l/r partial derivatives on $\mathcal{X}$. f has continuous l/r partial derivatives of order k on $\mathcal{X}$ if f has l/r partial derivatives of order k on $\mathcal{X}$, each of which is continuous at all points in $\mathcal{X}$, where continuity is defined in terms of local perturbations only within $\mathcal{X}$.

Assumption GMM2^{2*}. (a) Assumptions GMM2(a), GMM2(c), and GMM2(e) hold.

The domain of $G(\theta)$ includes a set Θ^+ that satisfies $\Theta^+ - \theta_0$ equals the intersection of a union of orthants and an open cube $C(\mathbf{0}, \varepsilon)$ for some $\varepsilon > 0$ and $\Theta \cap S(\theta_0, \varepsilon_1) \subset \Theta^+$ for some $\varepsilon_1 > 0$, where Θ is the parameter space. Each element of the k-vector–valued function $G(\theta)$ has continuous l/r partial derivatives of order 1 on Θ^+.
(b) (c) Each element of the k-vector–valued function $G_T(\theta)$ has continuous l/r partial derivatives of order 1 on $\Theta^+ \, \forall T \geq 1$ with probability 1.
(d) For all $\gamma_T \to 0$,

$$\sup_{\theta\in\Theta:\|\theta-\theta_0\|\leq\gamma_T} \left\| \frac{\partial}{\partial\theta'}G_T(\theta) - \frac{\partial}{\partial\theta'}G_T(\theta_0) \right\| = o_p(1),$$

where $(\partial/\partial\theta')G_T(\theta)$ denotes the $k \times s$ matrix of l/r partial derivatives of $G_T(\theta)$.
(e) $(\partial/\partial\theta')G_T(\theta_0) \to_p \Gamma = (\partial/\partial\theta')G(\theta_0)$, where $(\partial/\partial\theta')\times G(\theta_0)$ denotes the $k \times s$ matrix of l/r partial derivatives of $G(\theta)$ at θ_0.

Lemma 2. (a) Assumption GMM2* implies assumption GMM2 with $G(\theta) = Eg(Y_t, \theta)$.
(b) Assumptions GMM2* and GMM3* imply that assumption GMM3 holds with $\mathcal{V} = \sum_{j=-\infty}^{\infty} Eg(Y_0, \theta_0)g(Y_{-j}, \theta_0)'$.
(c) Assumption GMM2^{2*} implies assumption GMM2.

Comment. Assumptions GMM2* and GMM3* also imply that the autocovariances $\{Eg(Y_0, \theta_0)g(Y_{-j}, \theta_0)' : j = \ldots, 0, 1, \ldots\}$ are absolutely summable.

Next, MD estimators are considered. It is easy to see that assumption GMM2 holds in the MD case with $G(\theta) = \xi_0 - h(\theta)$ under the following conditions.

Assumption MD2. (a) $\hat{\xi}_T \to_p \xi_0$.
(b) $h(\theta) = h(\theta_0) + \Gamma(\theta - \theta_0) + o(\|\theta - \theta_0\|)$ as $\|\theta - \theta_0\| \to 0$ for $\theta \in \Theta \cap S(\theta_0, \varepsilon)$ for some $\varepsilon > 0$, where Γ is a nonrandom $k \times s$ matrix with full column rank.
(c) $\xi_0 = h(\theta_0)$.
(d) The weight matrix $A_T(\theta)$ satisfies assumption GMM2(e).

With the MD estimator, assumption GMM2(d) holds automatically because $G_T(\theta) - G(\theta) - G_T(\theta_0) = 0$. Assumption GMM3 is established for the MD estimator by showing that $T^{1/2}(\hat{\xi}_T - \xi_0) \to_d \widetilde{G} \sim N(\mathbf{0}, \mathcal{V})$ for some nonrandom matrix $\mathcal{V}$.

3.4 Quadratic Approximation of the Generalized Method of Moments Criterion Function

The criterion function $L_T(\theta)$ has a quadratic approximation given by

$$L_T(\theta) = \|A_T(\theta_0)G_T(\theta_0)\|^2/2 + G_T(\theta_0)'M\Gamma(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)'(\Gamma'M\Gamma)(\theta - \theta_0) + R_T(\theta), \quad \text{where}$$
$$M = A'A \tag{21}$$

and $R_T(\theta)$ is a remainder term. Under assumptions GMM2 and GMM3, it can be shown that the remainder $R_T(\theta)$ is small.

Lemma 3. Assumptions GMM2 and GMM3 imply that for all $\gamma_T \to 0$,

$$\sup_{\theta\in\Theta:\|\theta-\theta_0\|\le\gamma_T} T|R_T(\theta)|/(1+\|T^{1/2}(\theta-\theta_0)\|)^2 = o_p(1). \quad (22)$$

Comment. The property of the remainder in (22) is shown to be sufficiently strong that the difference between the normalized GMM estimator, $T^{1/2}(\hat{\theta}-\theta_0)$, and the normalized estimator that minimizes the quadratic approximation to $L_T(\theta)$, say $T^{1/2}(\tilde{\theta}-\theta_0)$, is $o_p(1)$, and hence the two estimators have the same asymptotic distribution. Of course, it is much simpler to determine the asymptotic distribution of $T^{1/2}(\tilde{\theta}-\theta_0)$ than that of $T^{1/2}(\hat{\theta}-\theta_0)$, because a quadratic criterion function is very well behaved.

Next, the quadratic approximation to the GMM criterion function $L_T(\theta)$ is simplified. Define

$$\mathcal{J} = \Gamma' M \Gamma,$$
$$Z_T = \mathcal{J}^{-1}\Gamma' M T^{1/2} G_T(\theta_0),$$

and

$$q_T(\lambda) = (\lambda - Z_T)'\mathcal{J}(\lambda - Z_T) \text{ for } \lambda \in R^s. \quad (23)$$

Then (21) can be written as

$$\begin{aligned} L_T(\theta) &= L_T(\theta_0) + (\Gamma' M G_T(\theta_0))'(\theta-\theta_0) \\ &\quad + \frac{1}{2}(\theta-\theta_0)'\mathcal{J}(\theta-\theta_0) + R_T(\theta) \\ &= L_T(\theta_0) - \frac{1}{2T}Z_T'\mathcal{J}Z_T \\ &\quad + \frac{1}{2T}q_T\big(T^{1/2}(\theta-\theta_0)\big) + R_T(\theta). \quad (24) \end{aligned}$$

Note that the estimator that minimizes the quadratic approximation to $L_T(\theta)$ (i.e., the right side of (24) excluding $R_T(\theta)$), call this estimator $\hat{\theta}_q$, equals $\theta_0 + T^{-1/2}Z_T$. Hence, $T^{1/2}(\hat{\theta}_q - \theta_0) = Z_T$ and Z_T determines the asymptotic distribution of the unrestricted estimator.

3.5 The Parameter Space

This section provides conditions on the parameter space under which the asymptotic distribution of the GMM estimator $\hat{\theta}$ is derived. The condition used is from Chernoff (1954), who considered likelihood ratio tests when a parameter is on a boundary.

The asymptotic distribution of $\hat{\theta}$ depends on a local approximation to the shifted parameter space $\Theta - \theta_0$. The local approximation is given by a cone. By definition, a set $\Lambda \subset R^s$ is a *cone* if $\lambda \in \Lambda$ implies $a\lambda \in \Lambda\ \forall a \in R$ with $a > 0$. Examples of cones include R^s, linear subspaces, orthants, unions of orthants, and sets defined by linear equalities and/or inequalities of the form $\Gamma_a\lambda = \mathbf{0}$ and $\Gamma_b\lambda \le \mathbf{0}$, where Γ_j is a $k_j \times s$ matrix for $j = a, b$.

Define the distance between a point $y \in R^s$ and a set $\Lambda \subset R^s$ by

$$\text{dist}(y, \Lambda) = \inf_{\lambda\in\Lambda}\|y-\lambda\|. \quad (25)$$

A set $\Phi \subset R^s$ is *locally approximated* (at the origin) by a cone $\Lambda \subset R^s$ if

$$\text{dist}(\phi, \Lambda) = o(\|\phi\|) \text{ as } \|\phi\| \to 0 \text{ for } \phi \in \Phi$$

and

$$\text{dist}(\lambda, \Phi) = o(\|\lambda\|) \quad \text{as } \|\lambda\| \to 0 \text{ for } \lambda \in \Lambda. \quad (26)$$

Assumption GMM4. $\Theta - \theta_0$ is locally approximated by a cone Λ.

Assumption GMM4 allows for linear, kinked, and curved boundaries.

Now two easily verifiable sufficient conditions for Assumption GMM4 are given. The conditions are specified in terms of the parameter space Θ shifted to be centered at the origin rather than at θ_0, that is, in terms of $\Theta - \theta_0$. A set $\Gamma \subset R^s$ is said to be *locally equal* to a set $\Lambda \subset R^s$ if $\Gamma \cap S(\mathbf{0}, \varepsilon) = \Lambda \cap S(\mathbf{0}, \varepsilon)$ for some $\varepsilon > 0$.

Assumption GMM4.* $\Theta - \theta_0$ is locally equal to a cone $\Lambda \subset R^s$.

Assumption GMM4* allows for parameter spaces $\Theta - \theta_0$, which are defined by multivariate linear equality and/or inequality constraints. For example, one could have

$$\Theta = \{\theta \in R^s : \Gamma_a\theta = r_1, \Gamma_b\theta \le r_2, \ \|\theta\| \le c < \infty\}, \quad (27)$$

$\Gamma_a\theta_0 = r_1$, and $\Gamma_b\theta_0 \le r_2$ with equality for zero or more elements of r_2, where Γ_j is an $\ell_j \times s$ matrix, r_j is an ℓ_j-vector, and $0 \le \ell_j \le s$ for $j = a, b$. In this example,

$$\Lambda = \{\lambda \in R^s : \Gamma_a\lambda = \mathbf{0}, \ \Gamma_{b1}\lambda \le \mathbf{0}\}, \quad (28)$$

where Γ_{b1} denotes the submatrix of Γ_b that consists of the rows of Γ_b for which $\Gamma_b\theta_0 \le r_2$ holds as an equality.

Assumption GMM4* does not allow for any curvature in the boundary of Θ near θ_0. Such curvature arises in some examples, such as cases where Θ is a sphere, ellipse, cylinder, or a set defined by smooth nonlinear equality and/or inequality constraints and θ_0 is on its boundary. Assumption GMM4 can be verified in these cases using the following conditions.

The following sufficient condition for assumption GMM4 considers the case where θ_0 is on the boundary of Θ and some smooth nonlinear equality and/or inequality constraints are binding at θ_0.

Assumption GMM4^{2*}. For some $\varepsilon > 0$, $\Theta \cap S(\theta_0, \varepsilon) = \{\theta \in R^s : m_a(\theta) = \mathbf{0}, m_b(\theta) \le \mathbf{0}, \|\theta - \theta_0\| \le \varepsilon\}$, where $m_j(\theta) \in R^{c_j}$ for $0 \le c_j < \infty$ for $j = a, b$; $m_j(\theta_0) = \mathbf{0}$ for $j = a, b$; and $m(\cdot) = (m_a(\cdot)', m_b(\cdot)')'$ is continuously differentiable on some neighborhood of θ_0 with $(\partial/\partial\theta')m(\theta_0)$ of full row rank.

Note that Assumption GMM4^{2*} specifies the parameter space only locally to θ_0 and, by definition, $m_a(\theta) = \mathbf{0}$ and $m_b(\theta) \le \mathbf{0}$ are constraints that are binding at θ_0. If the true parameter vector θ_0 changes, then the inequality constraints that are binding at θ_0, $m_b(\cdot)$, typically change as well.

Lemma 4. Each of assumptions GMM4* and GMM4^{2*} is sufficient for assumption GMM4. Under assumption GMM4^{2*}, assumption GMM4 holds with $\Lambda = \{\lambda \in R^s : (\partial/\partial\theta')m_a(\theta_0)\lambda = \mathbf{0}, (\partial/\partial\theta')m_b(\theta_0)\lambda \le \mathbf{0}\}$.

The approximating cone Λ depends on the inequality constraints that are binding at θ_0, that is, $m_a(\theta) = \mathbf{0}$ and $m_b(\theta) \le \mathbf{0}$ in the case of assumption GMM4^{2*}. For a different true parameter vector, Λ typically is different.

Obtaining the asymptotic distribution of $\hat{\theta}$ requires the following assumption.

Assumption GMM5. Λ is convex.

Assumption GMM5 holds for all of the examples of Section 2.

3.6 Asymptotic Distribution of the Generalized Method of Moments Estimator

This section specifies the asymptotic distribution of $\hat{\theta}$. First, note that by assumptions GMM2(e) and GMM3,

$$Z_T \to_d Z = \mathcal{J}^{-1} G, \text{ where } G = \Gamma' M \widetilde{G} \quad \text{ and}$$
$$q_T(\lambda) \to_d q(\lambda) = (\lambda - Z)' \mathcal{J} (\lambda - Z) \quad (29)$$

for each $\lambda \in R^s$.

Define $\hat{\lambda}$ by $\hat{\lambda} \in \mathrm{cl}(\Lambda)$ and

$$q(\hat{\lambda}) = \inf_{\lambda \in \Lambda} q(\lambda). \quad (30)$$

Under assumption GMM5, $\hat{\lambda}$ is uniquely defined.

The asymptotic distribution of $T^{1/2}(\hat{\theta} - \theta_0)$ is that of $\hat{\lambda}$. This is the main result of this article.

Theorem 1. Suppose that assumptions GMM1–GMM5 hold. Then $T^{1/2}(\hat{\theta} - \theta_0) \to_d \hat{\lambda}$.

Comment. The proof of Theorem 1 is made by verifying the conditions of theorem 3 of Andrews (1999), which applies to general extremum estimators.

3.7 Asymptotic Distributions of Subvectors of the Generalized Method of Moments Estimator

In this section the asymptotic distribution of $T^{1/2}(\hat{\theta} - \theta_0)$ is simplified by partitioning θ into two subvectors and providing separate expressions for each of the two corresponding subvectors of $\hat{\lambda}$.

θ is partitioned as follows:

$$\theta = \begin{pmatrix} \beta \\ \delta \end{pmatrix} \quad \text{and} \quad \theta_0 = \begin{pmatrix} \beta_0 \\ \delta_0 \end{pmatrix}, \quad (31)$$

where $\beta, \beta_0 \in R^p$, $\delta, \delta_0 \in R^q$, $0 \le p, q \le s$, and $p + q = s$. Later it is assumed that δ_0 is a parameter not on a boundary. This feature characterizes the subvectors β and δ.

The vectors and matrices $\hat{\theta}$, G, Z, $\mathcal{J}$, $\hat{\lambda}_T$, and $\hat{\lambda}$ are partitioned conformably with θ. Let

$$\hat{\theta} = \begin{pmatrix} \hat{\beta} \\ \hat{\delta} \end{pmatrix}, \quad G = \begin{pmatrix} G_\beta \\ G_\delta \end{pmatrix}, \quad Z = \begin{pmatrix} Z_\beta \\ Z_\delta \end{pmatrix},$$
$$\mathcal{J} = \begin{bmatrix} \mathcal{J}_\beta & \mathcal{J}_{\beta\delta} \\ \mathcal{J}_{\delta\beta} & \mathcal{J}_\delta \end{bmatrix}, \quad \hat{\lambda}_T = \begin{pmatrix} \hat{\lambda}_{\beta T} \\ \hat{\lambda}_{\delta T} \end{pmatrix}, \quad \text{and} \quad \hat{\lambda} = \begin{pmatrix} \hat{\lambda}_\beta \\ \hat{\lambda}_\delta \end{pmatrix}. \quad (32)$$

The defining feature of the parameter δ is given in part (b) of the following assumption.

Assumption GMM6. (a) The cone Λ of assumption GMM4 is a product set $\Lambda_\beta \times \Lambda_\delta$, where $\Lambda_\beta \subset R^p$ and $\Lambda_\delta \subset R^q$. (b) $\Lambda_\delta = R^q$.

Assumption GMM6 requires that δ_0 is not on a boundary. Any element of θ that does not satisfy this condition is lumped in with β.

Some calculations show that

$$Z_\beta = HZ = \mathcal{J}_\beta^{-1} G_\beta + \mathcal{J}_\beta^{-1} \mathcal{J}_{\beta\delta} (\mathcal{J}_\delta - \mathcal{J}_{\delta\beta} \mathcal{J}_\beta^{-1} \mathcal{J}_{\beta\delta})^{-1}$$
$$\times (\mathcal{J}_{\delta\beta} \mathcal{J}_\beta^{-1} G_\beta - G_\delta), \text{ where}$$
$$H = [I_p \,\vdots\, \mathbf{0}] \in R^{p \times (p+q)}. \quad (33)$$

Define

$$q_\beta(\lambda_\beta) = (\lambda_\beta - Z_\beta)' \left(H \mathcal{J}^{-1} H'\right)^{-1} (\lambda_\beta - Z_\beta). \quad (34)$$

Theorem 2. Suppose that assumptions GMM1–GMM6 hold. Then,

$$T^{1/2}(\hat{\beta} - \beta_0) \to_d \hat{\lambda}_\beta, \text{ where}$$
$$\hat{\lambda}_\beta \in \mathrm{cl}(\Lambda_\beta) \text{ solves } q_\beta(\hat{\lambda}_\beta) = \inf_{\lambda_\beta \in \Lambda_\beta} q_\beta(\lambda_\beta),$$
$$T^{1/2}(\hat{\delta} - \delta_0) \to_d \mathcal{J}_\delta^{-1} G_\delta - \mathcal{J}_\delta^{-1} \mathcal{J}_{\delta\beta} \hat{\lambda}_\beta,$$

and the convergence of the two terms holds jointly.

Comments. 1. Theorem 2 shows that the asymptotic distribution of $\hat{\delta}$ depends on whether β_0 is on a boundary if and only if $\mathcal{J}_{\delta\beta} \ne \mathbf{0}$.

2. Theorem 2 is a special case of corollary 1(b) of Andrews (1999).

3.8 A Closed-Form Expression for $\hat{\lambda}_\beta$

Now an assumption on Λ_β is considered under which there is a closed-form expression for $\hat{\lambda}_\beta$ and hence for $\hat{\lambda}_\delta$ as well.

Assumption GMM7. $\Lambda_\beta = \{\lambda_\beta \in R^p : \Gamma_a \lambda_\beta = \mathbf{0}, \Gamma_b \lambda_\beta \le \mathbf{0}\}$, where $\Gamma = [\Gamma_a' \,\vdots\, \Gamma_b']'$ is a full-row rank matrix.

Note that assumption GMM7 allows for the case where Γ_a or Γ_b does not appear. Assumption GMM7 holds in all of the examples considered in this article. For Λ_β as in assumption GMM7, $\hat{\lambda}_\beta$ is the solution to a quadratic programming (QP) problem with mixed linear equality and inequality constraints. Andrews's (1999) theorem 5 gives a closed-form solution to this problem. As an example, suppose that $\Lambda_\beta = R^+ \times R^{p-1}$. Then

$$\hat{\lambda}_\beta = \begin{cases} K Z_{\beta K} & \text{if } Z_{\beta K 1} \ge 0 \\ K(0, Z_{\beta K2} - \rho_{12} Z_{\beta K1}, \ldots, Z_{\beta Kp} - \rho_{p1} Z_{\beta K1})' & \\ \quad \text{otherwise,} & \end{cases}$$

where

$$K = \mathrm{diag}^{1/2}(H \mathcal{J}^{-1} H'), \; Z_{\beta K} = (Z_{\beta K1}', \ldots, Z_{\beta Kp}')' = K^{-1} Z_\beta,$$

and

$$\rho_{ij} = [K^{-1} H \mathcal{J}^{-1} H' K^{-1}]_{ij} \qquad \text{for } i, j = 1, \ldots, p. \quad (35)$$

When $\Lambda_\beta = R^- \times R^{p-1}$, the inequality in (35) is reversed.

As a second example, suppose that $\Lambda_\beta = (R^+)^2 \times R^{p-2}$. Then

$$\hat{\lambda}_\beta = KP_{L_K(\hat{j})} Z_{\beta K},$$

where

$$\begin{aligned} P_{L_K(\hat{j})} Z_{\beta K} = {} & 1(Z_{\beta K1} > 0, Z_{\beta K2} > 0) Z_{\beta K} \\ & + 1(Z_{\beta K1} - \rho_{21} Z_{\beta K2} > 0, Z_{\beta K2} \le 0) \\ & \times (Z_{\beta K1} - \rho_{21} Z_{\beta K2}, 0, Z_{\beta K3} - \rho_{23} Z_{\beta K2}, \\ & \quad \dots, Z_{\beta Kp} - \rho_{2p} Z_{\beta K2})' \\ & + 1(Z_{\beta K1} \le 0, Z_{\beta K2} - \rho_{12} Z_{\beta K1} > 0) \\ & \times (0, Z_{\beta K2} - \rho_{12} Z_{\beta K1}, \dots, Z_{\beta Kp} - \rho_{1p} Z_{\beta K1})', \quad (36) \end{aligned}$$

where K and ρ_{ij} are as in (35). For the case where $\Lambda_\beta = R^- \times R^+ \times R^{p-2}$, (36) holds but with the first of the two inequalities reversed in each of the indicator functions in the definition of $P_{L_K(\hat{j})} Z_{\beta K}$. Adjustments of (36) for the cases where $\Lambda_\beta = R^+ \times R^- \times R^{p-2}$ and $\Lambda_\beta = (R^-)^2 \times R^{p-2}$ are analogous.

For the case where Λ_β is of the form $\Lambda_\beta = \{\lambda_\beta \in R^p : \lambda_{\beta 1} \ge 0, \Gamma_a \lambda_\beta = \mathbf{0}\}$, $\hat{\lambda}_\beta$ is as defined in (35), but with $Z_{\beta K}$ replaced by $P_{\Gamma_a K} Z_{\beta K}$, where $P_{\Gamma_a K} = I_p - K^{-1} H \mathcal{J}^{-1} H' \Gamma_a' (\Gamma_a H \mathcal{J}^{-1} H' \Gamma_a')^{-1} \Gamma_a K$. For the case where Λ_β is of the form $\Lambda_\beta = \{\lambda_\beta \in R^p : \lambda_{\beta 1} \ge 0, \lambda_{\beta 2} \ge 0, \Gamma_a \lambda_\beta = \mathbf{0}\}$, $\hat{\lambda}_\beta$ is as defined in (36), but with $Z_{\beta K}$ replaced by $P_{\Gamma_a K} Z_{\beta K}$.

One can simulate the distribution of $\hat{\lambda}_\beta$ when Λ_β is as in assumption GMM7 by simulating Z_β or $Z_{\beta K}$ and computing $\hat{\lambda}_\beta$ using a standard quadratic programming algorithm (see, e.g., Gill, Murray, and Wright 1981). The programs GAUSS and Matlab have built-in procedures for doing so, called QPROG and QP. The GAUSS procedure QPROG is very quick. Alternatively, one can use the formulas of theorem 5 of Andrews (1999) or the foregoing equations.

4. INSTRUMENTAL VARIABLES REGRESSION EXAMPLE

4.1 Instrumental Variables Objective Function

The first example considers an IV estimator. The IV estimator minimizes the following quadratic form subject to the restrictions

$$L_T(\theta) = T^{-1} \sum_{t=1}^T (Y_t - X_t'\theta) Z_t' \left(T^{-1} \sum_{t=1}^T Z_t Z_t' \right)^{-1} \times T^{-1} \sum_{t=1}^T (Y_t - X_t'\theta) Z_t / 2. \quad (37)$$

The parameter space Θ is a compact set given by

$$\Theta = \{\theta \in R^s : m_a(\theta) = \mathbf{0}, m_b(\theta) \le \mathbf{0}, h(\theta) \le \mathbf{0}\}. \quad (38)$$

The errors, regressors, and IVs $\{(\varepsilon_t, X_t, Z_t) : t \le T\}$ are iid with

$$\begin{aligned} & E\varepsilon_t Z_t = \mathbf{0}, \qquad E\|\varepsilon_t Z_t\|^2 < \infty, \qquad E\|X_t Z_t'\|^2 < \infty, \\ & EZ_t Z_t' > 0, \qquad \text{and } EZ_t X_t' \text{ has full row rank.} \quad (39) \end{aligned}$$

In this example,

$$G_T(\theta) = T^{-1} \sum_{t=1}^T (Y_t - X_t'\theta) Z_t, \quad A_T(\theta) = \left(T^{-1} \sum_{t=1}^T Z_t Z_t' \right)^{-1/2},$$
$$G(\theta) = E(Y_t - X_t'\theta) Z_t, \quad \Gamma = -EZ_t X_t', \quad M = (EZ_t Z_t')^{-1},$$

and

$$\mathcal{J} = EX_t Z_t' (EZ_t Z_t')^{-1} EZ_t X_t' > 0. \quad (40)$$

Assumption GMM1*(a) holds because

$$\begin{aligned} & \sup_{\theta \in \Theta} \|G_T(\theta) - G(\theta)\| \\ & \quad = \sup_{\theta \in \Theta} \left\| T^{-1} \sum_{t=1}^T \varepsilon_t Z_t - T^{-1} \sum_{t=1}^T (Z_t X_t' - EZ_t X_t')(\theta - \theta_0) \right\| \\ & \quad \le \left\| T^{-1} \sum_{t=1}^T \varepsilon_t Z_t \right\| + \sup_{\theta \in \Theta} \|\theta - \theta_0\| \\ & \qquad \cdot \left\| T^{-1} \sum_{t=1}^T Z_t X_t' - EZ_t X_t' \right\|, \quad (41) \end{aligned}$$

the right side is $o_p(1)$ by the WLLN and the assumption that Θ is compact, and $A_T(\theta)' A_T(\theta) \to_p (EZ_t Z_t')^{-1} > 0$ by the WLLN and Slutsky's theorem.

Assumption GMM1*(b) holds because

$$\begin{aligned} L(\theta) & = \|AG(\theta)\|^2/2 \\ & = (\theta - \theta_0)' EX_t Z_t' (EZ_t Z_t')^{-1} EZ_t X_t' (\theta - \theta_0)/2 \quad (42) \end{aligned}$$

and hence for $\varepsilon > 0$,

$$\inf_{\theta \in \Theta / S(\theta_0, \varepsilon)} L(\theta) \ge \lambda_{\min}(\mathcal{J}) \varepsilon > 0 = L(\theta_0). \quad (43)$$

Assumption GMM2(a) holds by the WLLN. Assumption GMM2(b) holds because $G_T(\theta)$ is differentiable in θ with partial derivative matrix $\Gamma = -EZ_t X_t'$. Assumption GMM2(c) holds because $G(\theta_0) = E\varepsilon_t Z_t = 0$. Assumption GMM2(d) holds because

$$G_T(\theta) - G(\theta) - G_T(\theta_0) = -T^{-1} \sum_{t=1}^T (Z_t X_t' - EZ_t X_t')(\theta - \theta_0)$$

and

$$\begin{aligned} & \sup_{\|\theta - \theta_0\| \le \gamma_T} T^{1/2} \left\| T^{-1} \sum_{t=1}^T (Z_t X_t' - EZ_t X_t')(\theta - \theta_0) \right\| \\ & \qquad \Big/ \left(1 + T^{1/2} \|\theta - \theta_0\|\right) \\ & \quad \le \left\| T^{-1} \sum_{t=1}^T (Z_t X_t' - EZ_t X_t') \right\| = o_p(1) \quad (44) \end{aligned}$$

using the WLLN.

Assumption GMM2(e) holds by the WLLN and Slutsky's theorem given that EZ_tZ_t' is positive definite. Assumption GMM3 holds with

$$\widetilde{G} \sim N(\mathbf{0}, \mathcal{V}), \qquad \text{where} \qquad \mathcal{V} = E\varepsilon_t^2 Z_t Z_t'. \tag{45}$$

This follows from the CLT for iid mean 0 finite variance random variables.

4.2 Parameter Space

Assumption GMM4[2*] holds in this example provided that $m_a(\theta)$ and $m_b(\theta)$ are continuously differentiable on some neighborhood of θ_0 and $(\partial/\partial\theta')m(\theta_0)$ is full row rank, where $m(\theta) = (m_a(\theta)', m_b(\theta)')'$. In consequence, by Lemma 4, assumption GMM4 holds with

$$\Lambda = \left\{\lambda \in R^s : \frac{\partial}{\partial\theta'} m_a(\theta_0)\lambda = \mathbf{0},\ \frac{\partial}{\partial\theta'} m_b(\theta_0)\lambda \le \mathbf{0}\right\}. \tag{46}$$

For Λ as such, assumption GMM5 holds.

For example, suppose that

$$m_j(\theta) = v_j'\theta - d_j \tag{47}$$

for some given $v_j \in R^s$ and $d_j \in R$ for $j = a, b$. Then

$$\Lambda = \{\lambda \in R^s : v_a'\lambda = 0,\ v_b'\lambda \le 0\}. \tag{48}$$

Alternatively, suppose

$$m_j(\theta) = \theta' V_j \theta - d_j \tag{49}$$

for $V_j \in R^{s\times s}$ and d_j as before for $j = a, b$. Then the boundary of Θ at θ_0 is elliptical and

$$\Lambda = \{\lambda \in R^s : \theta_0{}' V_a \lambda = 0,\ \theta_0{}' V_b \lambda \le 0\}. \tag{50}$$

4.3 Asymptotic Distribution of the Instrumental Variables Estimator

In this example the quadratic approximation of (24) holds with

$$Z_T = \mathcal{J}^{-1}\Gamma' M T^{-1/2} \sum_{t=1}^{T} \varepsilon_t Z_t \qquad \text{and}$$
$$\mathcal{J} = EX_tZ_t'(EZ_tZ_t')^{-1}EZ_tX_t'. \tag{51}$$

By Theorem 1, $T^{1/2}(\hat{\theta} - \theta_0) \to_d \hat{\lambda}$. By definition, $\hat{\lambda} \in \mathrm{cl}(\Lambda)$ and

$$q(\hat{\lambda}) = \inf_{\lambda\in\Lambda} q(\lambda),$$

where

$$\begin{aligned} q(\lambda) &= (\lambda - Z)'\mathcal{J}(\lambda - Z);\\ G &= \Gamma' M \widetilde{G} \sim N(0, \mathcal{I});\\ \mathcal{I} &= \Gamma' M \mathcal{V} M \Gamma;\\ Z &= \mathcal{J}^{-1}G \sim N(0, \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}); \end{aligned} \tag{52}$$

Γ, M, and $\mathcal{J}$ are defined in (40); and Λ is defined in (46). If the errors $\{\varepsilon_t : t \ge 1\}$ are homoscedastic given Z_t [i.e., $E(\varepsilon_t|Z_t) = \sigma^2$ a.s.], then $\mathcal{I} = \sigma^2\mathcal{J}$ and $Z \sim N(0, \sigma^2\mathcal{J}^{-1})$.

4.4 Asymptotic Distribution of Subvectors of the Instrumental Variables Estimator

Typically, the restrictions $m_a(\theta) = \mathbf{0}$ and $m_b(\theta) \le \mathbf{0}$ of the IV example involve only some of the elements of θ. In this case, the vector $(\partial/\partial\theta')m(\theta_0)$, where $m(\theta) = (m_a(\theta)', m_b(\theta)')'$, that determines Λ contains some nonzero columns, say p of them, and some columns of zeros, say $s - p$ of them. Without loss of generality, assume that the first p columns of $(\partial/\partial\theta')m(\theta_0)$ are nonzero vectors and that the last $q = s - p$ columns are zero vectors for $1 \le p \le s$.

The vectors $\hat{\theta}$, θ_0, and θ are partitioned such that

$$\hat{\theta} = (\hat{\beta}', \hat{\delta}')', \theta_0 = (\beta_0', \delta_0')', \quad \text{and} \quad \theta = (\beta', \delta')', \tag{53}$$

where $\hat{\beta}, \beta_0, \beta \in R^p$, and $\hat{\delta}, \delta_0, \delta \in R^q$.

Now, with the foregoing partitioning, Assumption GMM6 holds. The set Λ is a product set $\Lambda_\beta \times \Lambda_\delta$ with

$$\Lambda_\beta = \left\{\lambda_\beta \in R^p : \frac{\partial}{\partial\beta'} m_a(\theta_0)\lambda_\beta = \mathbf{0}, \frac{\partial}{\partial\beta'} m_b(\theta_0)\lambda_\beta \le \mathbf{0}\right\} \text{ and}$$
$$\Lambda_\delta = R^q, \qquad \text{where } \frac{\partial}{\partial\beta'} m_j(\theta_0) \in R^{c_j \times p} \text{ for } j = a, b. \tag{54}$$

For example, if $m_j(\theta) = v_j'\theta - d_j$ with $v_j = (v_{j\beta}', \mathbf{0}')'$ and $v_{j\beta} \in R^p$ for $j = a, b$, then

$$\Lambda_\beta = \{\lambda_\beta \in R^p : v_{a\beta}'\lambda_\beta = 0, v_{b\beta}'\lambda_\beta \le 0\}. \tag{55}$$

Alternatively, if $m_j(\theta) = \theta' V_j \theta - d_j$ with $V_j = \mathrm{diag}\{V_{j\beta}, \mathbf{0}\}$ and $V_{j\beta} \in R^{p\times p}$ for $j = a, b$, then

$$\Lambda_\beta = \{\lambda_\beta \in R^p : \beta_0{}' V_{a\beta}\lambda_\beta = 0, \beta_0{}' V_{b\beta}\lambda_\beta \le 0\}. \tag{56}$$

By Theorem 2,

$$T^{1/2}(\hat{\beta} - \beta_0) \to_d \hat{\lambda}_\beta, \tag{57}$$

where $\hat{\lambda}_\beta$ solves $q_\beta(\hat{\lambda}_\beta) = \inf_{\lambda_\beta\in\Lambda_\beta} q_\beta(\lambda_\beta)$ with Λ_β as in (54) and $q_\beta(\lambda_\beta)$ defined in (34). In the simplest case where $p = 1$, which occurs when $m(\theta)$ places an upper or lower bound on a single parameter at $\theta = \theta_0$, the closed-form expression for $\hat{\lambda}_\beta$ given in (35) is applicable. If $p > 1$, then a closed-form expression for $\hat{\lambda}_\beta$ is given in (36) or theorem 5 of Andrews (1999).

By Theorem 2,

$$T^{1/2}(\hat{\delta} - \delta_0) \to_d \hat{\lambda}_\delta = \mathcal{J}_\delta^{-1}G_\delta - \mathcal{J}_\delta^{-1}\mathcal{J}_{\delta\beta}\hat{\lambda}_\beta,$$

where

$$G = \begin{pmatrix} G_\beta \\ G_\delta \end{pmatrix}, \quad G_\delta \sim N(\mathbf{0}, \mathcal{I}_\delta),$$
$$\mathcal{J} = \begin{bmatrix} \mathcal{J}_\beta & \mathcal{J}_{\beta\delta} \\ \mathcal{J}_{\delta\beta} & \mathcal{J}_\delta \end{bmatrix}, \mathcal{I} = \begin{bmatrix} \mathcal{I}_\beta & \mathcal{I}_{\beta\delta} \\ \mathcal{I}_{\delta\beta} & \mathcal{I}_\delta \end{bmatrix}, \tag{58}$$

$G_\delta \in R^q$, and $\mathcal{J}_\delta \in R^{q\times q}$.

5. MULTINOMIAL RESPONSE MODEL EXAMPLE

5.1 Method of Simulated Moments Estimator

This example considers a method of simulated moments (MSM) estimator. This is a GMM estimator. The moment conditions are defined as follows. Let $\pi(Z_t, \theta)$ denote the conditional expectation of $D(Z_t h(\eta_t, \theta))$ given Z_t. Let $W(Z_t, \theta)$ denote an $s \times m$ matrix of instruments (see McFadden 1989 regarding the choice of instruments). Then the moment conditions are

$$T^{-1} \sum_{t=1}^{T} W(Z_t, \theta)(d_t - \pi(Z_t, \theta)). \tag{59}$$

These moment conditions have expectation zero when $\theta = \theta_0$, as desired. Following McFadden (1989) and Pakes and Pollard (1989), the number of moment conditions is taken to equal the dimension s of θ. In this case the choice of weight matrix is immaterial, so $A_T(\theta) = I_s$.

The conditional probability vector $\pi(Z_t, \theta)$ is computationally intractable in many cases because it is a vector of high-dimensional integrals. Consequently, it is replaced in the moment conditions by a simulated probability

$$\hat{\pi}_S(Z_t, \theta) = S^{-1} \sum_{j=1}^{S} D(Z_t h(\eta_{tj}, \theta)), \tag{60}$$

where $\eta_{t1}, \ldots, \eta_{tS}$ are simulated random variables each with the same distribution as η_t and $\xi_t = (Z_t, \eta_t, \eta_{t1}, \ldots, \eta_{tS})$ is iid across $t = 1, \ldots, T$. (With crude frequency simulators, $\eta_{t1}, \ldots, \eta_{tS}$ are iid. With variance-reduced simulators, they are not necessarily independent.) The same simulated random variables are used for all θ.

The simulated moment conditions on which the GMM estimator is based are

$$G_T(\theta) = T^{-1} \sum_{t=1}^{T} W(Z_t, \theta)(d_t - \hat{\pi}_S(Z_t, \theta)). \tag{61}$$

The true parameter vector is

$$\theta_0 = (\theta'_{10}, \theta'_{20}, \theta'_{30})' = (\mathbf{0}', \theta'_{20}, \theta'_{30})', \tag{62}$$

where $\theta_{10} \in R^p$, $\theta_{20} > 0$ (element by element), (θ_1, θ_2) are the parameters that must be nonnegative, and θ_3 contains the remaining parameters. The parameter space Θ is

$$\Theta = \{\theta \in R^s : \theta = (\theta'_1, \theta'_2, \theta'_3)', \theta_1 \geq 0, \theta_2 \geq 0, \theta_3 = (\theta_{31}, \ldots, \theta_{3J})', \\ c_{\ell j} \leq \theta_{3j} \leq c_{uj} \ \forall j = 1, \ldots, J\} \tag{63}$$

for some constants $-\infty \leq c_{\ell j} < c_{uj} \leq \infty$ for $j = 1, \ldots, J$. The true subvector θ_{30} of θ_0 is assumed to not lie on a boundary. The parameter space could incorporate additional restrictions without affecting the results given later, provided that none are binding at θ_0. Note that Θ is not necessarily a bounded subset of R^s.

5.2 Verification of Assumptions GMM1–GMM3

The function $G(\theta)$ and the matrix Γ that appear in assumption GMM2 are

$$G(\theta) = EG_T(\theta) = EW(Z_t, \theta)(\pi(Z_t, \theta_0) - \pi(Z_t, \theta))$$

and

$$\Gamma = \frac{\partial}{\partial \theta'} G(\theta_0) = -EW(Z_t, \theta_0) \frac{\partial}{\partial \theta'} \pi(Z_t, \theta_0), \tag{64}$$

where $(\partial/\partial\theta')G(\theta_0)$ and $(\partial/\partial\theta')\pi(Z_t, \theta_0)$ denote the matrices of right partial derivatives of $G(\theta)$ and $\pi(Z_t, \theta_0)$ at θ_0 (see Sec. 3.3).

The quadratic approximation of the GMM criterion function depends on

$$Z_T = \Gamma^{-1} T^{1/2} G_T(\theta_0) \quad \text{and} \quad \mathcal{J} = \Gamma'\Gamma. \tag{65}$$

Assumptions GMM1–GMM3 are verified for this example taking the approach of Pakes and Pollard (1989, Sec. 4.2), using the previously stated assumptions plus the following:

(a) $\inf_{\theta \in \Theta / S(\theta_0, \varepsilon)} \|G(\theta)\| > 0 \ \forall \varepsilon > 0$.
(b) $G(\theta)$ has continuous right partial derivatives with respect to θ_1 and continuous partial derivatives with respect to θ_2 and θ_3 at θ_0.
(c) Γ is nonsingular.
(d) $E \sup_{\theta \in \Theta} \|W(Z_t, \theta)\| < \infty$. (66)
(e) $E \sup_{\theta \in \Theta \cap S(\theta_0, \varepsilon)} \|W(Z_t, \theta)\|^2 < \infty$ for some $\varepsilon > 0$.
(f) $\{B(\theta) : \theta \in \Theta\}$ is a VC class of sets, where $B(\theta) = \{(z, \eta) \in R^b \times R^r : z'h(\eta, \theta) \geq 0\}$.
(g) $\mathcal{F}_W = \{W(\cdot, \theta) : \theta \in \Theta\}$ is a Euclidean class of functions with envelope F that satisfies $EF^2(Z_t) < \infty$.

VC and Euclidean classes were defined by Pakes and Pollard (1989, Sec. 2).

A sufficient condition for (a) is that Θ is compact, $G(\theta)$ is continuous on Θ, and $G(\theta)$ has a unique zero at θ_0. A sufficient condition for (f) is that $h(\eta, \theta)$ is of the form

$$h(\eta, \theta) = \beta_1(\theta) + \beta_2(\theta)\eta + \eta'\beta_3(\theta)\eta \tag{67}$$

for some functions $\beta_j(\theta)$, $j = 1, 2, 3$. This holds for the probit models considered by McFadden (1989) and those discussed earlier. Sufficient conditions for (g) are that Θ is bounded, condition (e) holds, and $W(\cdot, \theta)$ satisfies the Lipschitz condition

$$\|W(Z, \theta^*) - W(Z, \theta)\| \\ \leq \phi(Z)\|\theta^* - \theta\| \ \forall Z \in R^{k \times m}, \ \forall \theta^*, \theta \in \Theta$$

and

$$E\phi^2(Z_t) < \infty. \tag{68}$$

Sufficiency of these conditions was shown by Pakes and Pollard (1989, Sec. 4.2).

Now assumptions GMM1–GMM3 are verified. Assumption GMM1 is verified by verifying assumption GMM1*. Assumption GMM1*(a) holds by an empirical process uniform WLLN by the argument of Pakes and Pollard (1989, Sec. 4.2) using (d), (f), and (g). Assumption GMM1*(b) holds by (a).

Assumption GMM2(a) holds by a pointwise WLLN for iid random variables with finite mean using (d). GMM2(b) holds by the one-term Taylor expansion of theorem 6 of Andrews (1999) and (c). GMM2(c) holds because $Ed_t = E\pi(Z_t, \theta_0) = E\hat{\pi}_S(Z_t, \theta_0)$ using the assumption of identical distributions of $\eta_t, \eta_{t1}, \ldots, \eta_{tS}$. GMM2(d) holds via empirical process results by the argument of Pakes and Pollard (1989, Sec. 4.2) using (e)–(g). GMM2(e) holds because $A_T(\theta) = I_s$.

Assumption GMM3 holds by the CLT for iid square-integrable random variables using (v) with

$$\begin{aligned} &\widetilde{G} \sim N(\mathbf{0}, \mathcal{V}) \qquad \text{and} \\ &\mathcal{V} = EW(Z_t, \theta_0)(d_t - \hat{\pi}_S(Z_t, \theta_0)) \\ &\qquad \times (d_t - \hat{\pi}_S(Z_t, \theta_0))' W(Z_t, \theta_0)'. \end{aligned} \tag{69}$$

If $\eta_t, \eta_{t1}, \ldots, \eta_{tS}$ are independent conditional on Z_t a.s., then $\mathcal{V}$ simplifies to

$$\begin{aligned} \mathcal{V} = \left(1 + \frac{1}{S}\right) &EW(Z_t, \theta_0)(\text{diag}(\pi(Z_t, \theta_0)) \\ &- \pi(Z_t, \theta_0)\pi(Z_t, \theta_0)')W(Z_t, \theta_0)'. \end{aligned} \tag{70}$$

Hence in this example,

$$\begin{aligned} &G = \Gamma'\widetilde{G} \sim N(\mathbf{0}, \Gamma'\mathcal{V}\Gamma) \qquad \text{and} \\ &Z = \mathcal{J}^{-1}G = \Gamma^{-1}\widetilde{G} \sim N(\mathbf{0}, \Gamma^{-1}\mathcal{V}(\Gamma^{-1})'). \end{aligned} \tag{71}$$

5.3 Parameter Space

Assumptions GMM4* and GMM5 hold in this example with $\Lambda = (R^+)^p \times R^{s-p}$.

5.4 Asymptotic Distribution of the Method of Simulated Moments Estimator

By Theorem 1, $T^{1/2}(\hat{\theta} - \theta_0) \to_d \hat{\lambda}$, where $\hat{\lambda}$ satisfies (30) with $(Z, \mathcal{J})$ defined in (65) and (71) and $\Lambda = (R^+)^p \times R^{s-p}$.

5.5 Asymptotic Distribution of Subvectors of the Method of Simulated Moments Estimator

In this case, write

$$\theta = (\beta', \delta')', \qquad \beta = \theta_1, \qquad \text{and} \qquad \delta = (\theta_2', \theta_3')'. \tag{72}$$

The set Λ is a product set $\Lambda_\beta \times \Lambda_\delta$ with

$$\Lambda_\beta = (R^+)^p \qquad \text{and} \qquad \Lambda_\delta = R^{s-p}. \tag{73}$$

Thus assumption GMM6 holds.

Recall that the parameter $\theta_1 (=\beta)$ corresponds to the random coefficient variances that are zero in the random coefficient probit model, the proportion of the error variance due to the random effect in the binary probit panel data model, or the measurement error variances that are zero in the measurement error probit model. By Theorem 2, the MSM estimator of this parameter has asymptotic distribution given by

$$T^{1/2}(\hat{\theta}_1 - \theta_{10}) \to_d \hat{\lambda}_\beta, \tag{74}$$

where $\hat{\lambda}_\beta = Z_\beta 1(Z_\beta \geq 0)$ and $\hat{\lambda}_\beta$ has a half-normal distribution when $p = 1$, $\hat{\lambda}_\beta$ is as in (36) when $p = 2$, and $\hat{\lambda}_\beta$ is as in theorem 5 of Andrews (1999) for $p > 2$.

Also, by Theorem 2, the asymptotic distribution of the remaining parameters is given by

$$T^{1/2}\left((\hat{\theta}_2', \hat{\theta}_3')' - (\theta_{20}', \theta_{30}')'\right) \to_d \hat{\lambda}_\delta,$$

where

$$\hat{\lambda}_\delta = \mathcal{J}_\delta^{-1}G_\delta - \mathcal{J}_\delta^{-1}\mathcal{J}_{\delta\beta}\hat{\lambda}_\beta, \qquad G = \binom{G_\beta}{G_\delta} \sim N(\mathbf{0}, \Gamma'\mathcal{V}\Gamma), \tag{75}$$

and $\mathcal{J}$ is partitioned as before.

6. PARTIALLY LINEAR REGRESSION EXAMPLE

In this example, the estimator objective function is semi-parametric,

$$\begin{aligned} L_T(\theta) = \Bigg\| T^{-1}\sum_{t=1}^{T}[Y_t - \widehat{E}(Y_t \mid Z_t) \\ - (X_t - \widehat{E}(X_t \mid Z_t))'\theta][X_t - \widehat{E}(X_t \mid Z_t)]\Bigg\|^2 \Big/ 2, \end{aligned} \tag{76}$$

where $\widehat{E}(Y_t \mid Z_t)$ and $\widehat{E}(X_t \mid Z_t)$ are nonparametric bias-reducing kernel estimators of $\eta_1(Z_t) = E(Y_t \mid Z_t)$ and $\eta_2(Z_t) = E(X_t \mid Z_t)$, as defined by Robinson (1988). The parameter space is the same as in (38) for the IV example.

In addition to the assumptions stated in Section 2.3, assume that

$$\begin{aligned} &\Phi = E(X_t - E(X_t \mid Z_t))(X_t - E(X_t \mid Z_t))' > 0, \\ &E\varepsilon_t^2 = \sigma^2 < \infty, \quad E|X_t|^4 < \infty; \end{aligned} \tag{77}$$

Z_t has a density $f(\cdot)$ with respect to Lebesgue measure; the functions $\mu(\cdot)$, $\eta_1(\cdot)$, and $\eta_2(\cdot)$ satisfy the smoothness and boundedness conditions of Robinson (1988, thm. 1); and the bandwidth and trimming parameters and the kernel used in the kernel estimators $\widehat{E}(Y_t \mid X_t, Z_t)$ and $\widehat{E}(X_t \mid Z_t)$ satisfy the conditions of Robinson (1988, thm. 1). In this example, $G_T(\theta)$ equals the average inside the absolute value signs in (76) and $A_T(\theta) = I_s$.

Assumption GMM1* holds by the same argument as in the IV example using the propositions in the appendix of Robinson (1988) to establish that

$$T^{-1}\textstyle\sum_{t=1}^{T}[\widehat{E}(Y_t \mid Z_t) - E(Y_t \mid Z_t)] \to_p 0$$

and

$$T^{-1}\textstyle\sum_{t=1}^{T}[\widehat{E}(X_t \mid Z_t) - E(X_t \mid Z_t)] \to_p 0. \tag{78}$$

The function $G(\theta)$ in this example is

$$\begin{aligned} G(\theta) &= -E[Y_t - E(Y_t \mid Z_t) - (X_t - E(X_t \mid Z_t))'\theta] \\ &\qquad \times [X_t - E(X_t \mid Z_t)] \\ &= -\Phi(\theta - \theta_0), \end{aligned} \tag{79}$$

where the second equality uses $E(Y_t \mid Z_t) = E(X_t \mid Z_t)'\theta_0 + \mu(Z_t)$, $E\varepsilon_t = 0$, and ε_t is independent of X_t and Z_t.

Assumption GMM2(a) holds by (78) and the WLLN. Assumption GMM2(b) holds with $\Gamma = -\Phi$ because $G(\theta)$ is differentiable. Assumption GMM2(c) holds by (79). Assumption GMM2(d) holds by the same argument as in (44) for the IV example with X_t and Z_t each replaced by $X_t - \widehat{E}(X_t \mid Z_t)$.

Assumption GMM3 holds in this example with

$$\widetilde{G} \sim N(\mathbf{0}, \mathcal{V}) \qquad \text{and} \qquad \mathcal{V} = \sigma^2\Phi, \tag{80}$$

by the propositions in the appendix of Robinson (1988). Hence,

$$\Gamma = -\Phi, \qquad M = I_s, \qquad \mathcal{J} = \Gamma^2 = \Phi^2 > 0,$$
$$G = -\Phi\widetilde{G} \sim N(\mathbf{0}, \sigma^2\Phi^3),$$

and

$$Z = \mathcal{J}^{-1}\Gamma M \widetilde{G} = \Gamma^{-2}G \sim N(\mathbf{0}, \sigma^2\Phi^{-1}). \tag{81}$$

Assumptions GMM4^{2*} and GMM5 hold in this example under the same conditions on $m_1(\theta)$ and $m_2(\theta)$ and with the same Λ matrix as in the IV example. Given the foregoing results, by Theorem 1, $T^{1/2}(\hat{\theta} - \theta_0) \rightarrow_d \hat{\lambda}$, where $\hat{\lambda}$ satisfies (30) with $(Z, \mathcal{J})$ defined in (81) and Λ defined in (46).

The parameter θ can be partitioned in this example in the same way as in the IV example. Then the asymptotic distributions of $T^{1/2}(\hat{\beta} - \beta_0)$ and $T^{1/2}(\hat{\delta} - \delta_0)$ are the same as in the IV example but with $(Z, \mathcal{J})$ defined as in (81).

APPENDIX: PROOFS

Proof of Lemma 1

Condition (a) and Assumption GMM2(d) are easily shown to be equivalent given the assumption that $G(\theta) = \Gamma(\theta - \theta_0) + o(\|\theta - \theta_0\|)$. Condition (a) obviously implies Condition (b). To obtain the converse, assume that Condition (b) holds and write, uniformly over $\{\theta \in \Theta : \|\theta - \theta_0\| \le \gamma_T\}$,

$$\begin{aligned}\|G_T(\theta) - G(\theta) - G_T(\theta_0)\| &= o_p(T^{-1/2} + \|G_T(\theta)\| + \|G(\theta)\|) \\ &\le o_p(T^{-1/2} + \|G_T(\theta) - G(\theta) - G_T(\theta_0)\| \\ &\quad + \|G_T(\theta_0)\| + 2\|G(\theta)\|),\end{aligned} \tag{A.1}$$

where the first equality uses condition (b) and the second uses Minkowski's inequality. Rearranging this equation and using the assumption that $\|G_T(\theta_0)\| = O_p(T^{-1/2})$ yields

$$\|G_T(\theta) - G(\theta) - G_T(\theta_0)\| = o_p(T^{-1/2} + \|G(\theta)\|) \tag{A.2}$$

uniformly over $\{\theta \in \Theta : \|\theta - \theta_0\| \le \gamma_T\}$. Hence Condition (a) of the lemma holds.

Proof of Lemma 2

Part (a) of the lemma is proved as follows. Assumption GMM2(a) holds by GMM2*(a)–(c) and the ergodic theorem. Assumption GMM2(b) holds with $G(\theta) = Eg(Y_t, \theta)$ and $\Gamma = E(\partial/\partial\theta')g(Y_t, \theta_0)$ by GMM2*(c) and (d) using the mean value and dominated convergence theorems. Assumption GMM2(c) holds by GMM2*(e). Assumption GMM2(d) holds by the argument of (17)–(19). Equation (17) holds by GMM2*(a)–(c) using a uniform WLLN such as theorem 5 of Andrews (1992) applied with assumption TSE-1D plus continuity of $E(\partial/\partial\theta')g(Y_t, \theta)$ at θ_0 which holds by the dominated convergence theorem.

Part (b) of the lemma follows by a CLT of Gordin, as in Hansen (1982).

Part (c) is established as follows. Assumption GMM2^{2*}(b) implies Assumption GMM2(b) by theorem 6 of Andrews (1999). To establish Assumption GMM2(d), write

$$\begin{aligned}G_T(\theta) - G_T(\theta_0) - G(\theta) &= \frac{\partial}{\partial\theta'}G_T(\theta^\dagger)(\theta - \theta_0) - G(\theta) \\ &= \left(\frac{\partial}{\partial\theta'}G_T(\theta_0) + o_p(1)\right)(\theta - \theta_0) \\ &\quad - \Gamma(\theta - \theta_0) + o(\|\theta - \theta_0\|) \\ &= o_p(\|\theta - \theta_0\|)\end{aligned} \tag{A.3}$$

uniformly over $\theta \in \Theta : \|\theta - \theta_0\| \le \gamma_T$, where the first equality holds for some $\theta^\dagger$ between θ and θ_0 by theorem 6 of Andrews (1999) [where $\theta^\dagger$ may differ across rows of $(\partial/\partial\theta')G_T(\theta^\dagger)$], the second equality holds by Assumptions GMM2^{2*}(d) and GMM2(b), and the last equality holds by Assumption GMM2^{2*} (e). Multiplying (A.3) by $T^{1/2}/\|T^{1/2}(\theta - \theta_0)\|$ and taking the supremum over $\{\theta \in \Theta : \|\theta - \theta_0\| \le \gamma_T\}$ establishes Assumption GMM2(d).

Proof of Lemma 3

First, the case where $A_T(\theta) = I_k$ is considered. Define the remainder term $R_T(\theta)$ and a close approximation to it, $R_T^*(\theta)$, which is obtained by replacing $\Gamma(\theta - \theta_0)$ by $G(\theta)$:

$$\begin{aligned}R_T(\theta) &= G_T(\theta)'G_T(\theta)/2 - G_T(\theta_0)'G_T(\theta_0)/2 \\ &\quad - G_T(\theta_0)'\Gamma(\theta - \theta_0) - (\theta - \theta_0)'\Gamma'\Gamma(\theta - \theta_0)/2\end{aligned}$$

and

$$\begin{aligned}R_T^*(\theta) &= G_T(\theta)'G_T(\theta)/2 - G_T(\theta_0)'G_T(\theta_0)/2 \\ &\quad - G_T(\theta_0)'G(\theta) - G(\theta)'G(\theta)/2.\end{aligned} \tag{A.4}$$

Let a, b, and c be k-vectors for which $a = b + c$. By the Cauchy–Schwarz inequality,

$$|a'a - b'b| = |c'c + 2b'c| \le c'c + 2\|b\| \cdot \|c\|. \tag{A.5}$$

Now,

$$\begin{aligned}&\sup_{\theta\in\Theta:\|\theta-\theta_0\|\le\gamma_T} T|R_T^*(\theta) - R_T(\theta)|/(1 + T^{1/2}\|\theta - \theta_0\|)^2 \\ &= \frac{1}{2}\sup_{\theta\in\Theta:\|\theta-\theta_0\|\le\gamma_T} T|2G_T(\theta_0)'(G(\theta) - \Gamma(\theta - \theta_0)) \\ &\qquad + G(\theta)'G(\theta) - (\theta - \theta_0)'\Gamma'\Gamma(\theta - \theta_0)| \\ &\quad /(1 + T^{1/2}\|\theta - \theta_0\|)^2 = o_p(1),\end{aligned}$$

where the first part of the second equality holds because $G(\theta) - \Gamma(\theta - \theta_0) = o(\|\theta - \theta_0\|)$ and $T^{1/2}G_T(\theta_0) = O_p(1)$, and the second part of the second equality holds by applying (A.5) with $a = G(\theta)$, $b = \Gamma(\theta - \theta_0)$, and $c = o(\|\theta - \theta_0\|)$. Thus it suffices to show that (22) holds with $R_T(\theta)$ replaced by $R_T^*(\theta)$.

Some algebra shows that

$$\begin{aligned} R_T^*(\theta) &= (G_T(\theta) - G(\theta) - G_T(\theta_0))' \\ &\quad \times (G_T(\theta) - G(\theta) - G_T(\theta_0))/2 \\ &\quad + (G(\theta) + G_T(\theta_0))(G_T(\theta) - G(\theta) - G_T(\theta_0)). \\ &= \|G_T(\theta) - G(\theta) - G_T(\theta_0)\|^2/2 \\ &\quad + (O(\|\theta - \theta_0\|) + G_T(\theta_0))' \\ &\quad \times (G_T(\theta) - G(\theta) - G_T(\theta_0)), \end{aligned} \tag{A.6}$$

where the second equality uses assumptions GMM2(b) and (c).

Define

$$\eta_T = \sup_{\theta \in \Theta: \|\theta - \theta_0\| \le \gamma_T} \frac{T^{1/2}\|G_T(\theta) - G(\theta) - G_T(\theta_0)\|}{1 + T^{1/2}\|\theta - \theta_0\|}. \tag{A.7}$$

By assumption GMM2(d), $\eta_T = o_p(1)$.

By (A.6) and (A.7),

$$\begin{aligned} &\sup_{\theta \in \Theta: \|\theta - \theta_0\| \le \gamma_T} \frac{2T|R_T^*(\theta)|}{(1 + T^{1/2}\|\theta - \theta_0\|)^2} \\ &\le \eta_T^2 + 2 \sup_{\theta \in \Theta: \|\theta - \theta_0\| \le \gamma_T} \frac{T^{1/2}O_p(\|\theta - \theta_0\|) + \|T^{1/2}G_T(\theta_0)\|}{(1 + T^{1/2}\|\theta - \theta_0\|)} \eta_T \\ &= \eta_T^2 + O_p(1)\eta_T \\ &= o_p(1), \end{aligned} \tag{A.8}$$

which establishes the lemma for the case where $A_T(\theta) = I_k$.

Next, part (a) is established for the case where $A_T(\theta)$ is as in assumption GMM2(e). The idea is to use the same proof as before, but with $G_T(\theta)$, $G(\theta)$, and Γ replaced by $A_T(\theta)G_T(\theta)$, $AG(\theta)$, and $A\Gamma$. This method works provided that assumptions GMM2(b), GMM2(c), and GMM2(d), which are used in the foregoing proof, hold with the same changes. Assumptions GMM2(b) and GMM2(c) obviously do. By lemma 3.5 of Pakes and Pollard (1989), assumption GMM2(e) and condition (b) of Lemma 1 imply that condition (b) of Lemma 1 holds with $G_T(\theta)$ and $G(\theta)$ replaced by $A_T(\theta)G_T(\theta)$ and $AG(\theta)$. In addition, Lemma 1 holds with these changes made to its conditions (a) and (b) and to assumption GMM2(d) by the proof given for the lemma. The last two results imply that assumption GMM2(d) holds with the aforementioned changes, as desired.

Proof of Lemma 4

The proof that assumption GMM4* implies assumption GMM4 is immediate given the definition of local approximation by a cone.

We now show that assumption GMM4^{2*} implies assumption GMM4. By assumption, $m_j(\theta_0) = \mathbf{0}$ for $j = a, b$. Let $\Gamma_j = (\partial/\partial\theta')m_j(\theta_0) \in R^{c_j \times s}$ for $j = a, b$. Let

$$\Gamma = \begin{bmatrix} \Gamma_a \\ \Gamma_b \\ \Gamma_c \end{bmatrix} \quad \text{and} \quad m^+(\theta) = \begin{pmatrix} m_a(\theta) \\ m_b(\theta) \\ \Gamma_c(\theta - \theta_0) \end{pmatrix}, \tag{A.9}$$

where $\Gamma_c \in R^{(s - c_a - c_b) \times s}$ is chosen such that Γ is nonsingular. Then $m^+(\theta_0) = \mathbf{0}$ and $(\partial/\partial\theta')m^+(\theta_0) = \Gamma$.

Let $\Phi = \Theta - \theta_0$. Given $\phi \in \Phi$ with ϕ close to 0, define

$$\lambda^* = \Gamma^{-1}m^+(\theta_0 + \phi). \tag{A.10}$$

Then $\Gamma\lambda^* = m^+(\theta_0 + \phi)$, $\Gamma_a\lambda^* = m_a(\theta_0 + \phi) = m_a(\theta) = \mathbf{0}$, and $\Gamma_b\lambda^* = m_b(\theta_0 + \phi) = m_b(\theta) \le \mathbf{0}$ for $\theta = \theta_0 + \phi \in \Theta$. Hence $\lambda^* \in \Lambda$. Element-by-element mean-value expansions give

$$\begin{aligned} \lambda^* &= \Gamma^{-1}m^+(\theta_0 + \phi) \\ &= \Gamma^{-1}m^+(\theta_0) + \Gamma^{-1}\frac{\partial}{\partial\theta'}m^+(\theta_0)\phi + o(\|\phi\|) \\ &= \mathbf{0} + \Gamma^{-1}\Gamma\phi + o(\|\phi\|) = \phi + o(\|\phi\|). \end{aligned} \tag{A.11}$$

It is concluded that $\text{dist}(\phi, \Lambda) \le \|\phi - \lambda^*\| = o(\|\phi\|)$, as required by assumption GMM4.

Next, the function $\tilde{m}(\cdot) = m^+(\theta_0 + \cdot) : R^s \to R^s$ is continuously differentiable on a neighborhood of $\mathbf{0}$ with nonsingular Jacobian matrix at $\mathbf{0}$ and $\tilde{m}(\mathbf{0}) = \mathbf{0}$. Hence, by the inverse function theorem, there exists a function $\tilde{m}^{-1}(\cdot) : R^s \to R^s$ that satisfies $\tilde{m}^{-1}(\phi)$ is continuously differentiable and $\tilde{m}(\tilde{m}^{-1}(\phi)) = \phi$ for all ϕ in a neighborhood of $\mathbf{0}$, $\tilde{m}^{-1}(\mathbf{0}) = \mathbf{0}$, and $(\partial/\partial\phi')\tilde{m}^{-1}(\mathbf{0}) = [(\partial/\partial\phi')\tilde{m}(0)]^{-1}$ $(= \Gamma^{-1})$.

Given $\lambda \in \Lambda$ with λ close to $\mathbf{0}$, define

$$\phi^* = \tilde{m}^{-1}(\Gamma\lambda). \tag{A.12}$$

Then $m^+(\theta_0 + \phi^*) = \tilde{m}(\phi^*) = \tilde{m}(\tilde{m}^{-1}(\Gamma\lambda)) = \Gamma\lambda$, $m_a(\theta_0 + \phi^*) = \Gamma_a\lambda = \mathbf{0}$, and $m_b(\theta_0 + \phi^*) = \Gamma_b\lambda \le \mathbf{0}$. Hence $\phi^* \in \Phi$. Element-by-element mean-value expansions give

$$\begin{aligned} \phi^* &= \tilde{m}^{-1}(\Gamma\lambda) = \tilde{m}^{-1}(\mathbf{0}) + \frac{\partial}{\partial\phi'}\tilde{m}^{-1}(\mathbf{0})\Gamma\lambda + o(\|\lambda\|) \\ &= \mathbf{0} + \left[\frac{\partial}{\partial\phi'}\tilde{m}(\mathbf{0})\right]^{-1}\Gamma\lambda + o(\|\lambda\|) = \lambda + o(\|\lambda\|). \end{aligned} \tag{A.13}$$

Hence $\text{dist}(\lambda, \Theta - \theta_0) \le \|\lambda - \phi^*\| = o(\|\lambda\|)$, and Assumption GMM4 holds.

Proof of Theorem 1

The theorem holds by theorem 3 of Andrews (1999) with $\ell_T(\theta)$, B_T, and $R_T(\theta)$ of Andrews (1999) equal to $-TL_T(\theta)$, $T^{1/2}I_s$, and $-TR_T(\theta)$ of this article provided that assumptions 2–6 of Andrews (1999) hold. The latter are implied by Andrews's (1999) assumptions 1, 2*, 3, 5, and 6. These are verified as follows. Andrews's assumption 1 holds by assumption GMM1. Assumption 2* holds by lemma 3 given assumptions GMM2 and GMM3. Assumption 3 holds immediately from assumption GMM3 and same as in the nonsingularity of $\mathcal{J}$, which holds by assumption GMM2. Assumptions 5 and 6 are implied by assumptions GMM4 and GMM5.

Proof of Theorem 2

Theorem 2 is a special case of corollary 1(b) of Andrews (1999).

ACKNOWLEDGMENTS

The author thanks Arthur Lewbel, Whitney Newey, David Pollard, and Chris Sims for helpful comments. The author also gratefully acknowledges the research support of National Science Foundation grants SBR-9730277 and SES-0001706.

[Received May 2002. Revised June 2002.]

REFERENCES

Aitchison, J., and Silvey, S. D. (1958), "Maximum Likelihood Estimation of Parameters Subject to Constraint," *Annals of Mathematical Statistics*, 29, 813–828.

Andrews, D. W. K. (1991), "Asymptotics for Kernel-Based Nonorthogonal Semiparametric Estimators," unpublished manuscript, Cowles Foundation, Yale University.

——— (1992), "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

——— (1994), "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, Vol. IV, eds. R. F. Engle and D. McFadden, New York: North-Holland.

——— (1999), "Estimation When a Parameter Is on a Boundary," *Econometrica*, 67, 1341–1383.

——— (2001), "Testing When a Parameter Is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734.

Andrews, D. W. K., and McDermott, C. J. (1995), "Nonlinear Econometric Models With Deterministically Trending Variables," *Review of Economic Studies*, 62, 343–360.

Berry, S. (1994), "Estimating Discrete Choice Models of Product Differentiation," *Rand Journal of Economics*, 25, 242–262.

Berry, S., Levinsohn, J., and Pakes, A. (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Chant, D. (1974), "On Asymptotic Tests of Composite Hypotheses in Nonstandard Conditions," *Biometrika*, 61, 291–298.

Chernoff, H. (1954), "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, 54, 573–578.

Gallant, A. R., and Golub, G. H. (1984), "Imposing Curvature Restrictions on Flexible Functional Forms," *Journal of Econometrics*, 26, 295–321.

Geyer, C. J. (1994), "On the Asymptotics of Constrained M-Estimation," *The Annals of Statistics*, 22, 1993–2010.

Gill, P. E., Murray, W., and Wright, M. H. (1981), *Practical Optimization*, New York: Academic Press.

Gourieroux, C., and Monfort, A. (1989), *Statistique et Modeles Econometriques*, Vol. 2, Paris: Economica. English translation (1995), *Statistics and Econometric Models*, Vol. 2, trans. by Q. Vuong, Cambridge, UK: Cambridge University Press.

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica*, 50, 1029–1054.

Hausman, J. A., and Wise, D. (1978), "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46, 403–426.

Horowitz, J. L. (1993), "Semiparametric Estimation of a Work-Trip Mode Choice Model," *Journal of Econometrics*, 58, 49–70.

Jeganathan, P. (1982), "On the Asymptotic Theory of Estimation When the Limit of the Loglikelihood Is Mixed Normal," *Sankhya*, Ser. A, 44, part 2, 172–212.

Le Cam, L. (1960), "Locally Asymptotically Normal Families of Distributions," *University of California Publications in Statistics*, 3, 37–98.

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

Moran, P. A. P. (1971), "Maximum-Likelihood Estimation in Non-Standard Conditions," *Proceedings of the Cambridge Philosophical Society*, 70, 441–450.

Newey, W. K. (1994), "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.

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

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

Robinson, P. M. (1988), "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

Self, S. G., and Liang, K.-Y. (1987), "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions," *Journal of the American Statistical Association*, 82, 605–610.

van der Vaart, A. W., and Wellner, J. (1996), *Weak Convergence and Empirical Processes*, New York: Springer.

Wang, J. (1996), "Asymptotics of Least-Squares Estimators for Constrained Nonlinear Regression," *The Annals of Statistics,* 24, 1316–1326.