ASYMPTOTICS FOR SEMIPARAMETRIC ECONOMETRIC MODELS: II. STOCHASTIC EQUicontinuity AND NONPARAMETRIC KERNEL ESTIMATION

by Donald W. K. Andrews

March 1989

Revised: July 1990
ASYMPTOTICS FOR
SEMIPARAMETRIC ECONOMETRIC MODELS:
II. STOCHASTIC EQUICONTINUITY
AND NONPARAMETRIC KERNEL ESTIMATION

Donald W. K. Andrews¹
Cowles Foundation for Research in Economics
Yale University

March 1989
Revised: July 1990
ABSTRACT

This paper presents several stochastic equicontinuity results that are useful for establishing the asymptotic properties of estimators and tests in parametric, semi-parametric, and nonparametric econometric models. In particular, they can be applied straightforwardly in the estimation and testing results of Andrews (1989b). The paper takes various stochastic equicontinuity results from the probability literature, which rely on entropy conditions of one sort or another, and provides primitive conditions under which the entropy conditions hold. This yields stochastic equicontinuity results that are readily applicable in a variety of contexts.

This paper also presents a number of consistency results for nonparametric kernel estimators of density and regression functions and their derivatives. These results are particularly useful in semiparametric estimation and testing problems that rely on preliminary nonparametric estimators, as in Andrews (1989b). The results allow for near epoch dependent non–identically distributed random variables, data–dependent bandwidth sequences, preliminary estimation of parameters (e.g., regression based on residuals), and nonparametric regression on index functions. Some of the results make use of the stochastic equicontinuity results of the paper.

JEL Classification Number: 211

Keywords: Asymptotic results, bracketing method, empirical process, kernel estimator, nonparametric density estimator, nonparametric regression estimator, semiparametric estimator, semiparametric test, series expansion, Sobolev norm, stochastic equicontinuity.
1. INTRODUCTION

This paper is a sequel to Andrews (1989b) — hereafter referred to as ASEM:I. In ASEM:I it is shown that the asymptotic properties of a large class of parametric and semiparametric estimators and tests can be established if one is able to establish (i) the stochastic equicontinuity of certain stochastic processes and (ii) various consistency and rate of convergence properties of preliminary nonparametric estimators. Here we provide primitive conditions under which (i) the stochastic equicontinuity property holds and (ii) the consistency and rate of convergence properties hold for one type of nonparametric estimator, viz., kernel estimators.

In addition to their use with the results of ASEM:I, the results given here should be of independent interest and applicability. For example, the stochastic equicontinuity results are used (or can be used) in the semiparametric and nonparametric misspecification testing results of Yatchew (1988), Wooldridge (1989), and Whang and Andrews (1990), in the semiparametric estimation results of Newey (1989) and in some of the uniform consistency results given below for nonparametric kernel estimators. Similarly, the kernel estimation results offer several innovations and generalizations that should be useful in the pure nonparametric estimation context.

Numerous results are available in the probability literature concerning sufficient conditions for stochastic equicontinuity (references are given below). Most of these results rely on some sort of entropy condition. For application to specific estimation and testing problems in parametric and semiparametric models, such entropy conditions are not sufficiently primitive. This paper provides an array of primitive conditions under which such entropy conditions hold, and hence, under which stochastic equicontinuity obtains. These primitive conditions are designed specifically for their ease of application in the estimation and testing problems of ASEM:I and the other applications referred to above. The primitive conditions considered here include: differentiability conditions, Lipschitz
conditions, $L^p$ continuity conditions, Vapnik-Chervonenkis conditions, and combinations thereof. The various stochastic equicontinuity results given here cover independent, $m-$dependent, strong mixing, and near epoch dependent random variables. Examples are given to show how the results can be used in estimation and testing problems.

The paper gives four sets of kernel estimation results. The first set establishes uniform consistency and $L^Q$ consistency, with rates of convergence, for kernel estimators of density and regression functions and their derivatives. The results apply to near epoch dependent (NED) non-identically distributed random variables (rv's) and allow for data-dependent bandwidth parameters. These results are the most general available with respect to temporal dependence and non-identical distributions.

The second set of kernel results extends the first set to the case where the dependent and/or regressor variables are not observed, but estimates of them are available that are based on finite dimensional preliminary estimators. This situation arises quite frequently when nonparametric estimators are used in semiparametric criterion functions, as in ASEM:I. For example, the results cover nonparametric regression when the dependent variable is a residual or a squared residual from some preliminary estimation.

The third set of kernel results considers nonparametric regressions of a dependent variable that is indexed by a parameter $\alpha_1$ on a regressor variable that is indexed by a parameter $\alpha_2$, where $\alpha_1$ and $\alpha_2$ may be finite or infinite dimensional. The uniform convergence results referred to above are extended to cover uniform convergence over the parameters $\alpha_1$ and $\alpha_2$ as well as over values of the regressor variable. Such results are useful in semiparametric examples when nonparametric regressions on index functions are employed, such as with the MAD-DUC and three-step sample selection estimator examples of ASEM:I.

Last, for the case of regression on an indexed set of regressor variables, it is shown how the third set of results can be used to obtain consistency results for kernel estimators
of derivatives of the regression functions with respect to the index parameter \( \alpha_2 \). Such results are useful in the MAD-DUC and sample selection examples referred to above.

The remainder of this paper is organized as follows: Section 2 defines stochastic equicontinuity and gives a heuristic discussion of this property. Sections 3–6 provide the stochastic equicontinuity results of the paper. Sections 7–10 provide the kernel estimation results.

All limits in the paper are taken as \( T \to \infty \) unless stated otherwise. Throughout we let "with probability \( \to 1 \)" abbreviate "with probability that goes to one as \( T \to \infty \)."

2. STOCHASTIC EQUICONTINUITY

We begin by introducing some notation. Let \( \{W_{Tt} : t < T, T \geq 1\} \) be a triangular array of \( \mathcal{N} \)-valuedrv's defined on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\), where \( \mathcal{N} \subset \mathbb{R}^k \). For notational simplicity we often abbreviate \( W_{Tt} \) by \( W_t \). Let \( \mathcal{T} \) be a pseudo-metric space with pseudo-metric \( \rho_{\mathcal{T}}(\cdot, \cdot) \). Let

\[
\mathcal{M} = \{m(\cdot, \tau) : \tau \in \mathcal{T}\}
\]

be a class of real functions defined on \( \mathcal{N} \) and indexed by \( \tau \in \mathcal{T} \). Define an empirical process \( \nu_T(\cdot) \) by

\[
\nu_T(\tau) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (m(W_{Tt}, \tau) - \mathbb{E}(m(W_{Tt}, \tau))) \quad \text{for} \quad \tau \in \mathcal{T},
\]

where \( \sum_{t=1}^{T} \) abbreviates \( \sum_{t=1}^{T} \).

Next we define stochastic equicontinuity of \( \{\nu_T(\cdot) : T \geq 1\} \). The definition is a natural extension to the stochastic context of the notion of equicontinuity of functions. Loosely speaking, \( \{\nu_T(\cdot) : T \geq 1\} \) is stochastically equicontinuous if \( \nu_T(\tau_1) \) is close to \( \nu_T(\tau_2) \) with large probability for \( T \) large whenever \( \tau_1 \) is close to \( \tau_2 \). More precisely, we have:
DEFINITION: \( \{ \nu_T(\cdot) : T \geq 1 \} \) is stochastically equicontinuous if \( \forall \epsilon > 0 \) and \( \eta > 0 \),
\[ \exists \delta > 0 \text{ such that} \]
\[ \lim_{T \to \infty} \mathbb{P}^* \left[ \sup_{\rho_T(\tau_1, \tau_2) < \delta} |\nu_T(\tau_1) - \nu_T(\tau_2)| > \eta \right] < \epsilon, \]
where \( \mathbb{P}^* \) denotes \( \mathbb{P} \)-outer measure.²

(The use of \( \mathbb{P}^* \) is a measure theoretic subtlety. If the quantity in parentheses is measurable, then \( \mathbb{P}^* \) can be replaced by \( \mathbb{P} \).

Two equivalent definitions of stochastic equicontinuity are the following:

(i) \( \{ \nu_T(\cdot) : T \geq 1 \} \) is stochastically equicontinuous if for every sequence of constants \( \{ \delta_T \} \)
that converges to zero, we have
\[ \sup_{\rho_T(\tau_1, \tau_2) \leq \delta_T} |\nu_T(\tau_1) - \nu_T(\tau_2)| \leq 0. \]

(ii) \( \{ \nu_T(\cdot) : T \geq 1 \} \)
is stochastically equicontinuous if for all sequences of random elements \( \{ \hat{\tau}_{1T} \} \) and \( \{ \hat{\tau}_{2T} \} \)
that satisfy \( \rho_T(\hat{\tau}_{1T}, \hat{\tau}_{2T}) \leq 0 \), we have
\[ \nu_T(\hat{\tau}_{1T}) - \nu_T(\hat{\tau}_{2T}) \leq 0. \]

Note that the second characterization of stochastic equicontinuity reflects its use in ASEM:1, where we take
\( \hat{\tau}_{1T} = \hat{\tau} \), \( \hat{\tau}_{2T} = \hat{\tau}_0 \), and \( \rho_T(\hat{\tau}, \tau_0) \leq 0 \) to conclude that
\( \nu_T(\hat{\tau}) - \nu_T(\tau_0) \leq 0. \) Allowing
\( \{ \hat{\tau}_{1T} \} \) and \( \{ \hat{\tau}_{2T} \} \) to be random in the second characterization is crucial. If only fixed
sequences were considered, then the property would be substantially weaker — it would not
deliver the result just stated that is used in ASEM:1 — and its proof would be substantially
simpler — the property would follow directly from Chebyshev’s inequality.

As noted above, the stochastic equicontinuity property can be usefully exploited in
establishing the asymptotic properties of parametric, semiparametric, and nonparametric
estimators and test statistics. Its primary use in the probability literature, however, is in
the proof of weak convergence results, including abstract functional central limit theorems
(CTLs). If \( (\mathcal{T}, \rho_T) \) is a totally bounded pseudo–metric space, \( \{ \nu_T(\cdot) : T \geq 1 \} \) is stochas-
tically equicontinuous, and the finite dimensional distributions of \( \{ \nu_T(\cdot) : T \geq 1 \} \)
converge, then \( \{ \nu_T(\cdot) : T \geq 1 \} \) converges weakly to a stochastic process on \( \mathcal{T} \) that has
uniformly \( \rho_T \)-continuous sample paths almost surely (and is a Borel measurable element of
the space \( B(\mathcal{T}) \) of all bounded, real-valued functions on \( \mathcal{T} \), where \( B(\mathcal{T}) \) is endowed with the sup norm. Conversely, if \( (\mathcal{T}, \rho_\mathcal{T}) \) is a totally bounded pseudo-metric space and \( \{\nu_{\mathcal{T}}(\cdot) : \mathcal{T} \geq 1\} \) converges weakly to such a process on \( \mathcal{T} \), then \( \{\nu_{\mathcal{T}}(\cdot) : \mathcal{T} \geq 1\} \) is stochastically equicontinuous and its finite dimensional distributions converge, see Pollard (1990, Theorem 10.2). When the finite dimensional distributions of \( \{\nu_{\mathcal{T}}(\cdot) : \mathcal{T} \geq 1\} \) are asymptotically normal, the limit process is Gaussian and the stochastic equicontinuity condition serves the dual role of helping to establish the existence of the Gaussian process and of proving a functional CLT. For examples of the use of stochastic equicontinuity in establishing functional CLTs, see Ossiander (1987), Andrews (1989a), Pollard (1990, Theorem 10.7), and Andrews and Pollard (1990).

To demonstrate the plausibility of the stochastic equicontinuity property, suppose \( \mathcal{K} \) contains only linear functions, i.e., \( \mathcal{K} = \{g : g(w) = w'\tau \text{ for some } \tau \in \mathbb{R}^k\} \), and \( \rho_\mathcal{T}(\cdot, \cdot) \) is the Euclidean metric. In this case, the left-hand side expression in (2.3) equals

\[
\limsup_{T \to \infty} P^* \left[ \sup_{\rho_\mathcal{T}(\tau_1, \tau_2) < \delta} \left| \frac{1}{\sqrt{T}} \Sigma_1^T (W_t - EW_t)'(\tau_1 - \tau_2) \right| > \eta \right]
\]

\[
\leq \limsup_{T \to \infty} P \left[ \left| \frac{1}{\sqrt{T}} \Sigma_1^T (W_t - EW_t) \right| > \eta/\delta \right] < \epsilon,
\]

where the second inequality holds for \( \delta \) sufficiently small provided \( \frac{1}{\sqrt{T}} \Sigma_1^T (W_t - EW_t) \) = \( O_p(1) \). Thus, \( \{\nu_{\mathcal{T}}(\cdot) : \mathcal{T} \geq 1\} \) is stochastically equicontinuous in this case if the rv's \( \{W_t - EW_t\} \) satisfy an ordinary CLT.

For classes of nonlinear functions, the stochastic equicontinuity property is substantially more difficult to verify than for linear functions. Indeed, it is not difficult to demonstrate that it does not hold for all classes of functions \( \mathcal{K} \). Some restrictions on \( \mathcal{K} \) are necessary — \( \mathcal{K} \) cannot be too complex/large.

To see this, suppose \( \{W_t\} \) are iid with distribution \( P_1 \) that is absolutely continuous with respect to Lebesgue measure and \( \mathcal{K} \) is the class of indicator functions of all Borel sets in \( \mathcal{W} \). Let \( \tau \) denote a Borel set in \( \mathcal{W} \) and let \( \mathcal{T} \) denote the collection of all such sets.
Then, \( m(w, \tau) = 1(w \in \tau) \). Take \( \rho_T(\tau_1, \tau_2) = \left[ \int (m(w, \tau_1) - m(w, \tau_2))^2 dP_1(w) \right]^{1/2} \). For any two finite sets \( \tau_1, \tau_2 \) in \( T \), \( \nu_T(\tau_j) = \frac{1}{\sqrt{T}} \sum_{1}^{T} 1(W_{T_t} \in \tau_j) \) and \( \rho_T(\tau_1, \tau_2) = 0 \), since \( P_1(W_t \in \tau_j) = 0 \) for \( j = 1, 2 \). Given any \( T \geq 1 \) and any realization \( \omega \in \Omega \), there exist finite sets \( \tau_{1T \omega} \) and \( \tau_{2T \omega} \) in \( T \) such that \( W_t(\omega) \in \tau_{1T \omega} \) and \( W_t(\omega) \notin \tau_{2T \omega} \) \( \forall t \in T \), where \( W_t(\omega) \) denotes the value of \( W_t \) when \( \omega \) is realized. This yields \( \nu_T(\tau_{1T \omega}) = \sqrt{T} \), \( \nu_T(\omega_{2T \omega}) = 0 \), and \( \sup_{\rho_T(\tau_1, \tau_2) < \delta} |\nu_T(\tau_1) - \nu_T(\tau_2)| \geq \sqrt{T} \). In consequence, \( \{\nu_T(\cdot) : T \geq 1\} \) is not stochastically equicontinuous. The class of functions \( \mathcal{K} \) is too large.

Sections 3–6 below are devoted to obtaining a useful array of different restrictions on \( \mathcal{K} \) that are sufficient for stochastic equicontinuity of \( \{\nu_T(\cdot) : T \geq 1\} \). Section 3 considers a symmetrization result based on a theorem of Pollard (1990). This stochastic equicontinuity result allows for \( m \)-dependent non-identically distributed triangular arrays \( \{W_{T_t}\} \). This rules out some time-series examples. On the other hand, it allows for a wide variety of smooth and non-smooth unbounded functions \( \{m(\cdot, \tau)\} \) and unbounded random variables \( \{W_{T_t}\} \). Primitive conditions are presented under which Pollard's entropy condition holds.

Section 4 considers a bracketing result based on the work of Ossiander (1987) and its extension by Pollard (1989). This stochastic equicontinuity result also applies to \( m \)-dependent non-identically distributed rv's \( \{W_{T_t}\} \) and to a wide variety of smooth and non-smooth unbounded functions \( \{m(\cdot, \tau)\} \). Primitive conditions are developed under which Ossiander's entropy condition holds.

Section 5 briefly describes a series expansion result of Andrews (1989a). This stochastic equicontinuity result is the most general result considered here with respect to temporal dependence of \( \{W_{T_t}\} \), but is less general than those of Sections 3 and 4 with respect to the classes of functions \( \{m(\cdot, \tau)\} \) to which it applies. It allows for NED rv's,
but requires the functions \( \{m(w, \tau)\} \) to be bounded and smooth in \( w \) and requires some of the elements of the underlying rv's \( W_{Tt} \) to be bounded.

Section 6 presents a new stochastic equicontinuity result that combines the series result of Section 5 with a bracketing result of Andrews and Pollard (1990). This result allows for strong mixing rv's \( \{W_{Tt}\} \) and functions \( \{m(\cdot, \tau)\} \) that have one component that may be non-smooth but is finite dimensional, plus another component that is smooth but may be infinite dimensional.

For use in establishing the stochastic equicontinuity condition Assumption 2(e) of ASEM:I, one takes \( m(W_{Tt}, \tau) \) of this paper to equal one element of the \( \mathbb{R}^v \)-valued function \( m_{Tt}(W_{Tt}, \theta, \tau) \) of ASEM:I. Note that stochastic equicontinuity of each of the \( v \) elements of \( \{\nu_{Tt}(\cdot)\} \) of ASEM:I is equivalent to stochastic equicontinuity of the vector process \( \{\nu_{Tt}(\cdot)\} \) itself. For use in establishing the stochastic equicontinuity condition Assumption 2*(e) of ASEM:I, one takes \( m(W_{Tt}, \tau) \) of this paper to equal one element of the \( \mathbb{R}^v \)-valued function \( m_{Tt}(W_{Tt}, \theta, \tau) \) of ASEM:I, one sets \( \tau \) of this paper equal to \( (\theta, \tau) \) of ASEM:I, and one lets \( \tau \) of this paper equal \( \Theta \times \tau \) of ASEM:I.

When one uses the stochastic equicontinuity results given below to verify conditions given in ASEM:I, the stochastic equicontinuity results effectively define an appropriate set \( \mathcal{T} \) for use in ASEM:I. That is, \( \mathcal{T} \) must be such that \( \mathcal{H} = \{m(\cdot, \tau) : \tau \in \mathcal{T}\} \) is sufficiently restricted that one of the stochastic equicontinuity results holds. In addition, Assumptions 1(b), 2(b), and 2*(b) of ASEM:I require that \( P(\tau \in \mathcal{T}) \rightarrow 1 \). Hence, to obtain the asymptotic normality of a MINPIN estimator, one needs to have sufficient control over the preliminary nuisance parameter estimator \( \hat{\tau} \) such that \( m(\cdot, \hat{\tau}) \) is in one of the classes of functions described below with probability \( \rightarrow 1 \). For example, this may require showing that with probability \( \rightarrow 1 \) a nonparametric regression estimator is partially differentiable up to some order with its partial derivatives bounded above. For kernel regression estimators, the results of Sections 7–10 can be used for this purpose.
3. STOCHASTIC EQUICONtinuity VIA SYMMETRIZATION

In this section we consider a result of Pollard (1990) altered to encompass m–dependent rather than independent rv's and reduced in generality somewhat to achieve a simplification of the conditions. This result depends on a condition, referred to as Pollard's entropy condition, that is based on how well the functions in $\mathcal{M}$ can be approximated by a finite number of functions, where the distance between functions is measured by the largest $L^2(Q)$ distance over all distributions $Q$ that have finite support. The main purpose of this section is to establish primitive conditions under which the entropy condition holds. Following this, a number of examples are provided to illustrate the ease of verification of the entropy condition.

The pseudo–metric $\rho_{\mathcal{T}}$ on $\mathcal{T}$ is defined in this section by

$$\rho_{\mathcal{T}}(\tau_1, \tau_2) = \sup_{N \geq 1} \left( \frac{1}{N} \sum_{i=1}^{N} (m(W_{Nt_i}, \tau_1) - m(W_{Nt_i}, \tau_2))^2 \right)^{1/2}.$$  \hspace{1cm} (3.1)

Let $Q$ denote a probability measure on $\mathcal{M}$. For a real function $f$ on $\mathcal{M}$, let $Qf^2 = \int_{\mathcal{M}} f^2(w) dQ(w)$. Let $\mathcal{F}$ be a class of functions in $L^2(Q)$. The $L^2(Q)$ cover numbers of $\mathcal{F}$ are defined as follows:

**DEFINITION:** For any $\epsilon > 0$, the cover number $N_2(\epsilon, Q, \mathcal{F})$ is the smallest value of $n$ for which there exist functions $f_1, \ldots, f_n$ in $\mathcal{F}$ such that $\min_{j \leq n} Q(f - f_j)^2 \leq \epsilon$. $\forall f \in \mathcal{F}$. $N_2(\epsilon, Q, \mathcal{F}) = \infty$ if no such $n$ exists.

The log of $N_2(\epsilon, Q, \mathcal{F})$ is referred to as the $L^2(Q)$ $\epsilon$–entropy of $\mathcal{F}$. Let $Q$ denote the class of all probability measures $Q$ on $\mathcal{M}$ that concentrate on a finite set. The following entropy/cover number condition was introduced in Pollard (1982):

**DEFINITION:** A class $\mathcal{F}$ of real functions defined on $\mathcal{M}$ satisfies Pollard's entropy condition if
\[
\int_0^1 \sup_{Q \in \mathcal{Q}} \left[ \log N_2 \left[ \epsilon (QF^2)^{1/2}, Q, \mathcal{F} \right] \right]^{1/2} \, d\epsilon < \infty,
\]
where \( F \) is some envelope function for \( \mathcal{F} \), i.e. \( F \) is a real function on \( \mathcal{Y} \) for which \( |f(\cdot)| \leq F(\cdot) \) \( \forall f \in \mathcal{F} \).

As \( \epsilon \downarrow 0 \), the cover number \( N_2 \left[ \epsilon (QF^2)^{1/2}, Q, \mathcal{F} \right] \) increases. Pollard's entropy condition requires that it cannot increase too quickly as \( \epsilon \downarrow 0 \). This restricts the complexity/size of \( \mathcal{F} \) and does so in a way that is sufficient for stochastic equicontinuity.

**ASSUMPTION A:** \( \mathcal{M} \) satisfies Pollard's entropy condition with some envelope \( \tilde{M} \).

**ASSUMPTION B:** \( \lim_{T \to \infty} \frac{1}{T} \Sigma_1^T \sum_{t=1}^T EM^{2+\delta}(W_{Tt}) < \infty \) for some \( \delta > 0 \), where \( \tilde{M} \) is as in Assumption A.

**ASSUMPTION C:** \( \{W_{Tt}\} \) is an \( m \)-dependent triangular array of rv's.

**THEOREM II.1** (Pollard): Under Assumptions A–C, \( \{\nu_T(\cdot) : T \geq 1\} \) is stochastically equicontinuous with \( \rho_T \) given by (3.1).

**COMMENT:** Theorem II.1 is proved using a symmetrization argument. In particular, one obtains a maximal inequality for \( \nu_T(\tau) \) by showing that \( \sup_{\tau \in \mathcal{F}} |\nu_T(\tau)| \) is less variable than \( \sup_{\tau \in \mathcal{F}} \left| \frac{1}{T} \sum_{t=1}^T \sigma_t m(W_t, \tau) \right| \), where \( \sigma_t : t \leq T \) are iid rv's that are independent of \( \{W_t : t \leq T\} \) and have Rudemacher distribution \((-1, 1)\), each with probability \( 1/2 \). Conditional on \( \{W_t\} \) one performs a chaining argument that relies on Hoeffding's inequality for tail probabilities of sums of bounded, mean zero, independent rv's. The bound in this case is small when the average sum of squares of the bounds on the individual rv's is small. In the present case, the latter is just \( \frac{1}{T} \sum_{t=1}^T m^2(W_t, \tau) \). The maximal inequality ultimately is applied to the empirical measure constructed from differences of the form \( m(W_t^1, \tau_1) - m(W_t^2, \tau_2) \) rather than to just \( m(W_t, \tau) \). In consequence, the measure of distance between \( m(\cdot, \tau_1) \) and \( m(\cdot, \tau_2) \) that makes the bound effective is an \( L^2(\mathbb{P}) \).
pseudo-metric, where \( P_T \) denotes the empirical distribution of \( \{W_t : t \leq T\} \). This pseudo-metric is random and depends on \( T \), but is conveniently dominated by the largest \( L^2(Q) \) pseudo-metric over all distributions \( Q \) with finite support. This explains the appearance of the latter in the definition of Pollard's entropy condition. To see why Pollard's entropy condition takes the precise form given above, one has to inspect the details of the chaining argument. The interested reader can do so, see Pollard (1990, Sec. 3).

Combinatorial arguments have been used to establish that certain classes of functions, often referred to as Vapnik–Červonenkis (VC) classes of one sort or another, satisfy Pollard's entropy condition, see Pollard (1984, Ch. 2; 1990, Sec. 4) and Dudley (1987). Here we consider the most important of these VC classes for applications (type I classes below) and we show that several other classes of functions satisfy Pollard's entropy condition. These include Lipschitz functions indexed by finite dimensional parameters (type II classes) and infinite dimensional classes of smooth functions (type III classes). The latter are particularly important for applications to semiparametric and nonparametric problems because they cover realizations of nonparametric estimators (under suitable assumptions).

Having established that Pollard's entropy condition holds for several useful classes of functions, we proceed below to show that functions from these classes can be "mixed and matched," e.g., by addition, multiplication, and division, to obtain new classes that satisfy Pollard's entropy condition. In consequence, one can routinely build up fairly complicated classes of functions that satisfy Pollard's entropy condition. In particular, one can build up classes of functions that are suitable for use in ASEM:I and in other parametric, semiparametric, and nonparametric applications.

**DEFINITION:** A class \( \mathcal{F} \) of real functions on \( \mathcal{W} \) is called a type I class if it is of the form
\( (a) \; \mathcal{F} = \{ f : f(w) = w \cdot \xi \; \forall w \in \mathcal{W} \text{ for some } \xi \in \Psi \subset \mathbb{R}^k \} \) or \( (b) \; \mathcal{F} = \{ f : f(w) = h(w', \xi) \; \forall w \in \mathcal{W} \text{ for some } \xi \in \Psi \subset \mathbb{R}^k, h \in V_K \} \), where \( V_K \) is some set of functions from \( \mathbb{R} \) to \( \mathbb{R} \) each with total variation less than or equal to \( K < \infty \).
Common choices for \( h \) in (b) include the indicator function, the sign function, and Huber \( \psi \)-functions, among others.

The second class of functions we consider contains functions that are indexed by a finite dimensional parameter and are Lipschitz with respect to that parameter:

**DEFINITION:** A class \( \mathcal{F} \) of real functions on \( \mathcal{W} \) is called a type II class if each function \( f \) in \( \mathcal{F} \) satisfies: \( f(\cdot) = f(\cdot, \tau) \) for some \( \tau \in \mathcal{T} \), where \( \mathcal{T} \) is some bounded subset of Euclidean space and \( f(\cdot, \tau) \) is Lipschitz in \( \tau \), i.e.

\[
|f(\cdot, \tau_1) - f(\cdot, \tau_2)| \leq B(\cdot) \|\tau_1 - \tau_2\| \quad \forall \tau_1, \tau_2 \in \mathcal{T}
\]

for some function \( B(\cdot) : \mathcal{W} \to \mathbb{R} \).

The third class of functions we consider contains functions that depend on \( w = (w_a', w_b')' \) only through a subvector \( w_a \) that has dimension \( k_a \leq k \). The functions are smooth on a restricted subset of \( \mathcal{W} \) and are equal to a constant elsewhere. Define

\[
\mathcal{W}_a = \{ w_a \in \mathbb{R}^{k_a} : \exists w_b \text{ s.t. } (w_a', w_b')' \in \mathcal{W} \}. 
\]

For \( w, h \in \mathbb{R}^k \), we write \( w = (w_a', w_b')' \) and \( h = (h_a', h_b')' \).

**DEFINITION:** A class \( \mathcal{F} \) of real functions on \( \mathcal{W} \) is called a type III class if

(i) each \( f \) in \( \mathcal{F} \) depends on \( w \) only through a subvector \( w_a \) of dimension \( k_a \leq k \),

(ii) for some real number \( q > k_a/2 \), some constant \( C < \infty \), and some set \( \mathcal{W}_a^* \), which is a subset of \( \mathcal{W}_a \) and is a finite union of connected compact subsets of \( \mathbb{R}^{k_a} \), each \( f \in \mathcal{F} \) satisfies the smoothness condition: \( \forall w \in \mathcal{W} \) and \( w + h \in \mathcal{W} \) such that \( w_a \) and \( w_a + h_a \) are in the same connected set in \( \mathcal{W}_a^* \),

\[
f(w+h) = \sum_{v=0}^{[q]} \frac{1}{v!} B_v(h_a, w_a) + R(h_a, w_a) \quad \text{and}
\]

\[
R(h_a, w_a) \leq C\|h_a\|^q,
\]

where \( B_v(h_a, w_a) \) is homogeneous of degree \( v \) in \( h_a \) and \( (q, C, \mathcal{W}_a^*) \) do not depend on
(iii) for some constant $K$ and all $f \in \mathcal{F}$, $f(w) = K \forall w \in \mathcal{W}$ such that $w_a \in \mathcal{W}_a - \mathcal{W}^*$.

Typically the expansion of $f(w+h)$ in (3.4) is a Taylor expansion of order $[q]$ and the function $B_v(h_a, w_a)$ is the $v$--th differential of $f$ at $w$, i.e., $B_v(h_a, w_a) = \Sigma_v \frac{v!}{v_1! \cdots v_k!} \frac{\partial^v f(w)}{\partial w_1 \cdots \partial w_k} h_1^{v_1} \cdots h_k^{v_k}$, where $\Sigma_v$ denotes the sum over all ordered $k_a$--tuples $(v_1, \ldots, v_k)$ of nonnegative integers such that $v_1 + \cdots + v_k = v$, $w_a = (w_1, \ldots, w_k)$, and $h_a = (h_1, \ldots, h_k)$.

Sufficient conditions for condition (ii) above are: (a) for some real number $q > k_a/2$, $f \in \mathcal{F}$ has partial derivatives of order $[q]$ on $\mathcal{W}^* = \{ w \in \mathcal{W} : w_a \in \mathcal{W}_a^* \}$, (b) the $[q]$--th order partial derivatives of $f$ satisfy a Lipschitz condition with exponent $q-[q]$ and some Lipschitz constant $C^*$ that does not depend on $f \forall f \in \mathcal{F}$, and (c) $\mathcal{W}^*$ is the finite union of convex compact sets.

The envelope of a type III class $\mathcal{F}$ can be taken to be a constant function, since the functions in $\mathcal{F}$ are uniformly bounded in absolute value over $w \in \mathcal{W}$ and $f \in \mathcal{F}$.

In applications, type III classes of functions typically are classes of realizations of nonparametric function estimates. Since these realizations usually depend on only a subvector $W_{at}$ of $W_t = (W_{at}, W_{bt})'$, it is advantageous to define type III classes to contain functions that may depend on only part of $W_t$. By "mixing and matching" functions of type III with functions of types I and II (see below), classes of functions are obtained that depend on all of $w$.

In applications where the subvector $W_{at}$ of $W_t$ is a bounded rv, one may have $\mathcal{W}^*_a = \mathcal{W}_a$. In applications where $W_{at}$ is an unbounded rv, $\mathcal{W}^*_a$ must be a proper subset of $\mathcal{W}_a$ for $\mathcal{F}$ to be a type III class. A common case where the latter arises in the examples of ASEM:I is when $W_{at}$ is an unbounded rv, all the observations are used to estimate a non-
parametric function $\tau_0(w_a)$ for $w_a \in \mathcal{W}_a$, and a MINPIN estimator is defined that only uses observations $W_t$ such that $W_{at}$ is in a bounded set $\mathcal{W}_a^*$. In this case, one sets the nonparametric estimator of $\tau_0(w_a)$ equal to zero outside $\mathcal{W}_a^*$ and the realizations of this trimmed estimator form a type III class if they satisfy the smoothness condition (ii) for $w_a \in \mathcal{W}_a^*$.

**THEOREM II.2:** If $\mathcal{F}$ is a class of functions of type I, II, or III, then Pollard's entropy condition (3.2) (i.e. Assumption A) holds with envelope $F(\cdot)$ given by $1 \vee \sup_{f \in \mathcal{F}} |f(\cdot)|$, $1 \vee \sup_{f \in \mathcal{F}} |f(\cdot)| \vee B(\cdot)$, or $1 \vee \sup_{f \in \mathcal{F}} |f(\cdot)|$, respectively, where $\vee$ is the maximum operator.

**COMMENTS:** 1. For type I classes, the result of Theorem II.2 follows from results in the literature such as Pollard (1984, Ch. II) and Dudley (1987) (see the Appendix for details). For type II classes, Theorem II.2 is established directly. It is similar to Lemma 2.13 of Pakes and Pollard (1989). For type III classes, Theorem II.2 is established using uniform metric entropy results of Kolmogorov and Tihomirov (1961).

2. Type I classes of functions can be extended to include various VC classes of functions. By definition, such classes include (i) classes of indicator functions of VC sets, (ii) VC major classes of uniformly bounded functions, (iii) VC hull classes, (iv) VC subgraph classes, and (v) VC subgraph hull classes, where each of these classes is as defined in Dudley (1987) (but without the restriction that $f \geq 0$ $\forall f \in \mathcal{F}$). For brevity and simplicity, we do not discuss all of these classes here. Rather, we define type I classes to include the two examples of VC classes that are the most important for the applications of ASEM:I.

We now show how one can "mix and match" functions of types I, II, and III to obtain a wide variety of classes that satisfy Pollard's entropy condition (Assumption A). Let $\mathcal{G}$ and $\mathcal{G}^*$ be classes of $r \times s$ matrix-valued functions defined on $\mathcal{W}$ with scalar envelopes $G$ and $G^*$ respectively (i.e., $G : \mathcal{W} \to \mathbb{R}$ and $|g_{ij}(\cdot)| \leq G(\cdot)$ $\forall i = 1, \ldots, r$, $j = 1, \ldots, s$, $\forall g \in \mathcal{G}$). Let $g$ and $g^*$ denote generic elements of $\mathcal{G}$ and $\mathcal{G}^*$. Let $\mathcal{A}$
be defined as \( G \) is, but with \( s \times u \)-valued functions. Let \( h \) denote a generic element of \( \mathcal{X} \).

We say that a class of matrix-valued functions \( G, G^* \), or \( \mathcal{X} \) satisfies Pollard's entropy condition or is of type I, II, or III if it does so, or if it is, element by element for each of the rs or su elements of its functions.

Let \( G \ast G^* = \{ g + g^* : g \in G, g^* \in G \} \), \( G \mathcal{X} = \{ gh \} \), \( G \vee G^* = \{ g \vee g^* \} \), \( G \wedge G^* = \{ g \wedge g^* \} \), and \( |G| = \{|g|\} \), where \( \vee, \wedge \), and \( |\cdot| \) denote the element by element maximum, minimum, and absolute value operators respectively. If \( r = s \) and \( g(w) \) is non-singular for all \( w \in \mathcal{W} \) and \( g \in G \), let \( G^{-1} = \{ g^{-1} \} \). Let \( \lambda_{\text{min}}(\cdot) \) denote the smallest eigenvalue of the matrix \( \cdot \).

**Theorem II.3:** If \( G, G^* \), and \( \mathcal{X} \) satisfy Pollard's entropy condition with envelopes \( G \), \( G^* \), and \( \mathcal{X} \), respectively, then so do each of the following classes (with envelopes given in parentheses): \( G \cup G^* (G \vee G^*) \), \( G \otimes G^* (G + G^*) \), \( G \mathcal{X} ((G \vee 1)(H \vee 1)) \), \( G \vee G^* (G \vee G^*) \), \( G \wedge G^* (G \vee G^*) \), and \( |G| (G) \). If in addition \( r = s \) and \( G^{-1} \) has a finite envelope \( \overline{G} \), then \( G^{-1} \) also satisfies Pollard's entropy condition (with envelope \( (G \vee 1)^2 \overline{G}^2 \)).

**Comments:** 1. The stability properties of Pollard's entropy condition given in Theorem II.3 are quite similar to stability properties of packing numbers considered in Pollard (1990).

2. If \( r = s \) and \( \inf_{g \in G} \inf_{w \in \mathcal{W}} \lambda_{\text{min}}(g(w)) > 0 \), then \( G^{-1} \) has an envelope that is uniformly bounded by a finite constant.

**Examples:** We now consider several examples that illustrate the use of Theorems II.1–II.3 to verify Assumptions 2(e) and 2*(e) of ASEM: I.

1. Weighted least squares/partially linear regression (WLS/PLR) example of ASEM: I.

In this example, we need to verify Assumption 2(e). We have
\[ m(w, \theta_0, \tau) = \xi(w)[y - \tau_1(z) - (x - \tau_2(z))' \theta_0][x - \tau_2(z)]/\tau_3(z) \]  
and
\[ \mathcal{M} = \{ m(\cdot, \theta_0, \tau) : \tau \in \mathcal{T} \}, \]

where \( w = (y, x', z')' \), \( \xi(w) = 1(z \in \mathcal{Z}^*) \) for some set \( \mathcal{Z}^* \subset \mathcal{Z} \subset \mathbb{R}^k \) is the domain of \( \tau_j(z) \) for \( j = 1, 2, 3 \) and includes the support of \( Z_t \forall t \geq 1 \), and \( \mathcal{T} \) is as defined below. In (3.5), the possibility of trimming is made explicit. If \( \mathcal{Z}^* = \mathcal{Z} \), then no trimming occurs and the indicator function \( \xi(w) \) is redundant. If \( \mathcal{Z}^* \) is a proper subset of \( \mathcal{Z} \), then trimming occurs and the WLS estimator \( \hat{\theta} \) is based on only non-trimmed observations.

With some abuse of notation, let \( \tau_j(w) \) denote a function on \( \mathcal{W} \) that depends on \( w \) only through the \( k_a \)-subvector \( z \) and equals \( \tau_j(z) \) above for \( j = 1, 2, 3 \).

By applying Theorems II.1–II.3, we find that the following conditions are sufficient for Assumption 2(e) (i.e., stochastic equicontinuity of \( \{ \nu_T(\cdot) : T \geq 1 \} \) with the index set \( \mathcal{M} = \{ m(\cdot, \theta_0, \tau) : \tau \in \mathcal{T} \} \) in this example with \( \rho_T \) given by (3.1) with \( m(\cdot, \tau) \) replaced by \( m(\cdot, \theta_0, \tau) \):

(i) \( \{(Y_t, X_t, Z_t) : t \geq 1\} \) is an \( m \)-dependent sequence of rv's.

(ii) \( \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} (E\|X_t - X_t \theta_0\|^2 + E\|X_t\|^2 + E\|(Y_t - X_t \theta_0)X_t\|^2) < \infty \) for some \( \delta > 0 \).

(iii) \( \mathcal{T} = \{ \tau : \tau = (\tau_1, \tau_2, \tau_3), \tau_j \in \mathcal{T}_j \text{ for } j = 1, 2, 3 \} \), \( \mathcal{T}_j \) is a type III class of \( R^{p_j} \)-valued functions on \( \mathcal{W} \subset \mathbb{R}^k \) that depend on \( w = (y, x', z')' \) only through the \( k_a \)-vector \( z \) for \( j = 1, 2, 3 \), where \( p_1 = 1 \), \( p_2 = p \), and \( p_3 = 1 \), and \( \mathcal{T}_3 \subset \{ \tau_3 : \inf_{w \in \mathcal{W}} |\tau_3(w)| \geq \epsilon \} \) for \( \epsilon > 0 \).

Usually one takes the set \( \mathcal{W}^*_a \) of the definition of the type III class \( \mathcal{T}_j \) to be the same for \( j = 1, 2, 3 \) and to be the same as \( \mathcal{Z}^* \). No trimming occurs if \( \mathcal{W}^*_a = \mathcal{W}_a \) and \( \mathcal{Z}^* = \mathcal{Z} \). Since \( \mathcal{W}^*_a \) is bounded by condition (iii), conditions (i)–(iii) can be satisfied without trimming only if the rv's \( \{ Z_t : t \geq 1 \} \) are bounded.
Sufficiency of conditions (i)—(iii) for Assumption 2(e) is established as follows: Let $h_1(w) = y - x' \theta_0$ and $h_2(w) = x$. By Theorem II.2, $\{\xi\}, \{h_1\}, \{h_2\}$, and $T_j$ satisfy Pollard's entropy condition with envelopes $1, |h_1|, |h_2|$, and $C_j$, respectively, for some constant $C_j \in (1, w)$, for $j = 1, 2, 3$. By the $\mathcal{G}^{-1}$ result of Theorem II.3, so does $\{1/\tau_3 : \tau_3 \in T_3\}$ with envelope $C_3^2/\epsilon^2$. By the $\mathcal{G} \mathcal{X}$ and $\mathcal{G} \circ \mathcal{G}^*$ results of Theorem II.3 applied several times, $\mathcal{N}$ satisfies Pollard's entropy condition with envelope $(|h_1| \vee 1)C_4 + (|h_2| \vee 1)C_5 + (|h_1| \vee 1)(|h_2| \vee 1)C_6$ for some finite constants $C_4, C_5, C_6$. Hence, Theorem II.1 yields the stochastic equicontinuity of $\{\nu_T(\cdot)\}$, since (ii) suffices for Assumption B.

Next, we consider the conditions $P(\tilde{\tau} \in T) \rightarrow 1$ and $\tilde{\tau} \mathcal{B} \tau_0$ of Assumption 2(b) of ASEM-I. Suppose $\tilde{\tau}_j(z)$ is a nonparametric estimator of $\tau_{j0}(z)$ that is trimmed outside a set $\mathcal{W}_a^*$ to equal zero for $j = 1, 2$ and one for $j = 3$, where $\mathcal{W}_a^*$ is a finite union of convex compact subsets of $\mathbb{R}^k_a$. Suppose $\tilde{\tau}_j(z)$ and its partial derivatives of order $\leq [q] + 1$ are uniformly consistent over $z \in \mathcal{W}_a^*$ for $\tau_{j0}(z)$ and its corresponding partial derivatives, for $j = 1, 2, 3$, for some $q > k_a/2$. Suppose the partial derivatives of order $[q] + 1$ of $\tau_{j0}(z)$ are uniformly bounded over $z \in \mathcal{W}_a^*$ and $\inf_{z \in \mathcal{W}_a^*} \lambda_{\min}(\tau_{30}(z)) > 0$.

Suppose the trimming set $Z^*$ equals $\mathcal{W}_a^*$. Then, the realizations of $\tilde{\tau}_j(z)$, viewed as a function of $w$, lie in a type III class of functions with probability $\rightarrow 1$ for $j = 1, 2, 3$ and $\tilde{\tau} \mathcal{B} \tau_0$ (where $\tau_{j0}(z)$ is defined for $z \notin \mathcal{W}_a^*$ to equal zero for $j = 1, 2$ and one for $j = 3$). Hence, the above conditions plus (i) and (ii) of (3.6) imply that Assumption 2(e) and the first two parts of Assumption 2(b) hold. If $\tilde{\tau}_j(z)$ is a kernel regression estimator for $j = 1, 2, 3$, then sufficient conditions for the above uniform consistency properties are given in Sections 7–10 below.

If $\{(Y_t, X_t, Z_t) : t \geq 1\}$ is not an $m$–dependent sequence of rv's, then one must use the results of Section 5 or 6 below, rather than Theorems II.1–II.3, to establish Assumption 2(e). If the regressor variables $\{Z_t : t \geq 1\}$ are unbounded rv's and one does not trim the
nonparametric estimators \( \hat{\tau}_j(z) \) for \( j = 1, 2, 3 \) to equal constants outside some fixed bounded set, then the results of Section 4 below, rather than Theorems II.1–II.3, must be used to establish Assumption 2(e).

(2) Weighted censored least absolute deviations (WC–LAD) example of ASEM:1. In this example, we need to verify Assumption 2*(e). We have

\[
m(w, \theta, \tau) = \tau(x)1(x'\theta < c)\text{sgn}(y - x'\theta)x
\]

and

\[
\mathcal{M} = \{m(\cdot, \theta, \tau) : (\theta, \tau) \in \Theta \times \mathcal{T}\},
\]

where \( w = (y, x', c)' \), \( \Theta \) is some subset of \( \mathbb{R}^p \), and \( \mathcal{T} \) is as defined below. With some abuse of notation, let \( \tau(w) \) be a real–valued function on \( \mathcal{W} \) that depends on \( w \) only through \( x \) and equals \( \tau(x) \).

By Theorems II.1–II.3, the following assumptions are sufficient for Assumption 2*(e) with \( \rho_{\Theta \times \mathcal{T}} \) given by (3.1) with \( \Theta \times \mathcal{T} \) and \( m(\cdot, \theta, \tau) \) in place of \( \mathcal{T} \) and \( m(\cdot, \tau) \), respectively:

\[
(i) \ \{(Y_t, X_t, C_t) : t \geq 1\} \text{ is an m–dependent sequence of rv's.}
\]

(ii) \( \mathcal{T} \) is a type III class of real–valued functions defined on \( \mathcal{W} \subset \mathbb{R}^k \) that depend on \( w = (y, x', c)' \) only through the p–vector \( x \).

Sufficiency is established as follows: Let \( \mathcal{G}_1 = \{g : g(w) = 1(x'\theta < c) \text{ for some } \theta \in \Theta\} \), \( \mathcal{G}_2 = \{g : g(w) = \text{sgn}(y - x'\theta) \text{ for some } \theta \in \Theta\} \), and \( \mathcal{G}_3 = \{g : g(w) = x1(x \in \mathcal{W}_a^*)\} \), where \( \mathcal{W}_a^* \) is the bounded set that appears in the definition of the type III class \( \mathcal{T} \). \( \mathcal{G}_j \) is a type I class for \( j = 1, 2 \) and a type II class for \( j = 3 \). By Theorem II.2, \( \mathcal{T} \) and \( \mathcal{G}_j \) satisfy Pollard's entropy condition with envelopes \( \tilde{C} \) and \( C_j \), where \( \tilde{C} \) and \( C_j \) are constants in \( [1, m] \), for \( j = 1, 2, 3 \). By the \( \mathcal{GA} \) result of Theorem II.3, \( \mathcal{M} \) satisfies Pollard's entropy condition with envelope \( C^* \) for some constant \( C^* \in [1, m] \). Theorem II.1 now yields the stochastic equicontinuity result.
Comments similar to those made in Example (1) apply here regarding the boundedness of $X_t$ versus the trimming of $\hat{\gamma}(x)$ outside a bounded set and regarding sufficient conditions for nonparametric estimators $\hat{\gamma}(x)$ such that their realizations lie in a type III class with probability $\rightarrow 1$.

(3) $M-$estimators for standard, censored, and truncated linear regression models. In the models considered here, $\{(Y_t, X_t^*) : t \leq T\}$ are observed rv's and $\{(Y_t^*, X_t^*) : t \leq T\}$ are latent rv's. The models are defined by

$$Y_t^* = X_t^* \theta_0 + U_t^* , \quad t = 1, \ldots, T ,$$

linear regression (LR): $(Y_t, X_t) = (Y_t^*, X_t^*)$,

$$censored \ regression \ (CR): \quad (Y_t, X_t) = (Y_t^* 1(Y_t^* \geq 0), X_t^*)$$

truncated regression (TR): $(Y_t, X_t) = (Y_t^* 1(Y_t^* \geq 0), X_t^* 1(Y_t^* \geq 0))$.

Depending upon the context, the errors $\{U_t^*\}$ may satisfy any one of a number of assumptions such as constant conditional mean or quantile for all $t$ or symmetry about zero for all $t$. We need not be specific for present purposes.

We consider MINPIN estimators of $\theta_0$ with $W_t = (Y_t, X_t^*)'$, $d(m, \gamma) = m \cdot m/2$, and

$$m(w, \theta) = \psi_1(y - x' \theta) \psi_2(w, \theta)x \quad \text{or}$$

$$m(w, \theta, \tau) = \psi_1((y - x' \theta)/\tau) \psi_2(w, x)x ,$$

where $w = (y, x')'$, $\psi_1(\cdot)$ is a function of bounded variation, and $\tau$ ($\epsilon \subset \mathbb{R}$) corresponds to a preliminary scale estimator.

Examples of such MINPIN estimators include the following:

(a) LR model: Let $\psi_1(z) = \text{sgn}(z)$ and $\psi_2 = 1$ to obtain the LAD estimator. Let $\psi_1(z) = q - 1(y - x' \theta < 0)$ and $\psi_2 = 1$ to obtain Koenker and Bassett's (1978) regression quantile estimator for quantile $q \in (0,1)$.

Let $\psi_1(z) = (z \wedge c) \vee (-c)$ and
\[ \psi_2 = 1 \] to obtain Huber’s (1973) M-estimator with truncation at \( \pm c \). Let \( \psi_1(z) = |q - 1(y - x'\theta < 0)| \) and \( \psi_2(w,\theta) = y - x'\theta \) to obtain Newey and Powell’s (1987) asymmetric LS estimator.

(b) CR model: Let \( \psi_1(z) = q - 1(y - x'\theta < 0) \) and \( \psi_2(w,\theta) = 1(x'\theta > 0) \) to obtain Powell’s (1984, 1986a) censored regression quantile estimator for quantile \( q \in (0,1) \). Let \( \psi_1 = 1 \) and \( \psi_2(w,\theta) = 1(x'\theta > 0)(y - x'\theta) \wedge x'\theta \) to obtain Powell’s (1986b) symmetrically trimmed LS estimator.

(c) TR model: Let \( \psi_1 = 1 \) and \( \psi_2(w,\theta) = 1(y < 2x'\theta)(y - x'\theta) \) to obtain Powell’s (1986b) symmetrically trimmed LS estimator.

Only the Huber M-estimator of the LR model requires a preliminary scale estimator \( \tau \).

The following conditions are sufficient for Assumption 2*(c) with pseudo-metric \( \rho_\Theta \) given by (3.1) with \( \Theta \) and \( m(\cdot,\theta) \) in place of \( T \) and \( m(\cdot,\tau) \), or with \( \Theta \times T \) and \( m(\cdot,\theta,\tau) \) in place of \( T \) and \( m(\cdot,\tau) \), as is appropriate:

(i) \( \{(Y_t, X_t) : t \geq 1\} \) is an \( m \)-dependent sequence of rv’s.

(ii) \( \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E\|X_t\|^{2+\delta} < \infty \) for some \( \delta > 0 \).

(iii) \( \{\psi_2(\cdot,\theta) : \theta \in \Theta\} \) satisfies Pollard’s entropy condition with envelope

\[ \sup_{\theta \in \Theta} |\psi_2(\cdot,\theta)| \quad \text{and} \quad \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E\left[\|X_t\|^{2+\delta} + 1\right] \sup_{\theta \in \Theta} |\psi_2(W_t, \theta)|^{2+\delta} < \infty \]

for some \( \delta > 0 \).

(iv) \( \psi_1(\cdot) \) is a function of bounded variation.

In the examples above, condition (iv) is always satisfied and condition (iii) is automatically satisfied given (ii) whenever \( \psi_2 = 1 \) or \( \psi_2(w,\theta) = 1(x'\theta > 0) \). When \( \psi_2(w,\theta) = y - x'\theta \), \( \psi_2(w,\theta) = 1(x'\theta > 0)(y - x'\theta) \wedge x'\theta \), or \( \psi_2(w,\theta) = 1(y < 2x'\theta)(y - x'\theta) \), condition (iii) is satisfied provided \( \Theta \) is bounded and \( \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E\|U_t\|^{2+\delta} + E\|X_t\|^{4+\delta} + E\|U_tX_t\|^{2+\delta} < \infty \) for some \( \delta > 0 \). The latter follows using Theorem II.3, since
\{1(x' \theta > 0) \colon \theta \in \Theta \}, \ \{y - x' \theta \colon \theta \in \Theta \}, \ \{x' \theta \colon \theta \in \Theta \}, \ \text{and} \ \{1(y < 2x' \theta) \colon \theta \in \Theta \} \text{ are type I classes with envelopes } 1, \ |u| + \|x\| \sup_{\theta \in \Theta} \|\theta - \theta_0\|, \ \|x\| \sup_{\theta \in \Theta} \|\theta\|, \ \text{and } 1, \ \text{respectively, where } u = y - x' \theta_0.

Sufficiency of conditions (i)–(iv) for Assumption 2*(e) is established as follows: The sets \( \{g : g(w) = \psi_1(y - x' \theta) \text{ for some } \theta \in \Theta \}, \ \{g : g(w) = \psi_1((y - x' \theta)/\tau) \text{ for some } \theta \in \Theta \text{ and } \tau \in \mathcal{T} \subset \mathbb{R} \}, \ \{h : h(w) = x\} \text{ are type I classes with envelopes } C_1, \ C_1, \ \text{and} \ \|x\|, \ \text{respectively, for some constant } C_1 < \infty, \ \text{and hence satisfy Pollard's entropy condition by Theorem II.2. This result, condition (iii), and the } \mathcal{G}\mathcal{T} \ \text{result of Theorem II.3 show that } \mathcal{H} \ \text{satisfies Pollard's entropy condition with envelope } \left(\|x\| \vee 1 \right) \left(\sup_{\theta \in \Theta} \psi_2(w, \theta) \right) \vee 1\right).\]

Stochastic equicontinuity now follows from Theorem II.1, since Assumption B is implied by conditions (ii) and (iii).

(4) Method of simulated moments (MSM) estimator for multinomial probit. The model and estimator considered here are as in McFadden (1989) and Pakes and Pollard (1989). We consider a discrete response model with \( r \) possible responses. Let \( D_t \) be an observed response vector that takes values in \( \{e_i : i = 1, \ldots, r\} \), where \( e_i = (0, \ldots, 0, 1, 0, \ldots, 0)' \) is the \( i \)-th elementary \( r \)-vector. Let \( Z_{ti} \) denote an observed \( b \)-vector of covariates \( - \) one for each possible response \( i = 1, \ldots, r \). Let \( Z_t = [Z_{t1} : Z_{t2} \cdots Z_{tr}]' \). The model is defined such that

\[
D_t = e_i \ \text{if} \ \left(Z_{ti} - Z_t\right)' \left(\beta(\theta_0) + A(\theta_0)U_t\right) \geq 0 \ \forall \ell = 1, \ldots, r,
\]

where \( U_t \sim \mathcal{N}(0, I_c) \) is an unobserved normal rv, \( \beta(\cdot) \) and \( A(\cdot) \) are known \( R^{b \times 1} \) and \( R^{b \times c} \) valued functions of an unknown parameter \( \theta_0 \in \Theta \subset \mathbb{R}^p \).

McFadden's MSM estimator of \( \theta_0 \) is constructed using \( s \) independent simulated \( \mathcal{N}(0, I_c) \) rv's \( (Y_{t1}, \ldots, Y_{ts})' \) and a matrix of instruments \( g(Z_t, \theta) \), where \( g(\cdot, \cdot) \) is a known \( R^{v \times b} \)-valued function. The MSM estimator is a MINPIN estimator with \( W_t = (D_t, Z_t, Y_{t1}, \ldots, Y_{ts}), \ d(m, \gamma) = m \gamma/2, \text{ and} \)
\[(3.13) \quad m(w, \theta) = g(z, \theta) \left[ d - \frac{1}{s} \sum_{j=1}^{s} \mathbb{H}[z(\beta(\theta) + A(\theta)y_j)] \right], \]

where \( w = (d, z, y_1, \ldots, y_s) \). Here, \( \mathbb{H}[\cdot] \) is a \( \{0,1\}^{\mathbb{R}} \)-valued function whose \( i \)-th element is of the form

\[(3.14) \quad \prod_{\ell=1}^{L}(z_{i_{\ell}} - z_{j_{\ell}}) \cdot (\beta(\theta) + A(\theta)y_{j_{\ell}}) \geq 0. \]

In this example, the following conditions are sufficient for Assumption 2*(e) with \( p_{\Theta} \) given by (3.1) with \( \Theta \) and \( m(\cdot, \theta) \) in place of \( T \) and \( m(\cdot, \tau) \):

(i) \( \{(T_t, Z_t, Y_{t1}, \ldots, Y_{ts}) : t \geq 1\} \) is an \( m \)-dependent sequence of rv's.

(ii) \( \{g(\cdot, \theta) : \theta \in \Theta\} \) is a type II class of functions with Lipschitz function

\[(3.15) \quad B(\cdot) \text{ that satisfies } \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \left[ \mathbb{E} B^{2+\delta}(Z_t) + \mathbb{E} \sup_{\theta \in \Theta} \|g(Z_t, \theta)\|^{2+\delta} \right] < \infty \text{ for some } \delta > 0. \]

Note that condition (ii) holds if \( g(w, \theta) \) is differentiable in \( \theta \) \( \forall w \in \mathcal{W}, \forall \theta \in \Theta, \Theta \) is open, and

\[ \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \left[ \mathbb{E} \sup_{\theta \in \Theta} \left\| \frac{\partial}{\partial \theta} g(Z_t, \theta) \right\|^{2+\delta} + \mathbb{E} \sup_{\theta \in \Theta} \|g(Z_t, \theta)\|^{2+\delta} \right] < \infty \text{ for some } \delta > 0. \]

Sufficiency of conditions (i) and (ii) for assumption 2*(e) is established as follows:

Classes of functions of the form \( \{1((z_i - z_j) \cdot (\beta(\theta) + A(\theta)y_j) \geq 0) : \theta \in \Theta \subset \mathbb{R}^p\} \) are type 1 classes with envelopes equal to 1 (by including products \( z_i y_j \) and \( z_j y_j \) as additional elements of \( w \)) and hence satisfy Pollard's entropy condition by Theorem II.2. \( \{g(\cdot, \theta) : \theta \in \Theta\} \) also satisfies Pollard's entropy condition with envelope

\[ 1 \vee \sup_{\theta \in \Theta} \|g(\cdot, \theta)\| \vee B(\cdot) \] by condition (ii) and Theorem II.2. The \( \mathcal{G} \) result of Theorem II.3 now implies that \( \mathcal{M} \) satisfies Pollard's entropy condition with envelope

\[ 1 \vee \sup_{\theta \in \Theta} \|g(\cdot, \theta)\| \vee B(\cdot). \] Stochastic equicontinuity now follows by Theorem II.1.

Other examples for which Theorems II.1–II.3 can be used to verify Assumption 2(e) or 2*(e) of ASEM:I include the GMM/CMR, MAD–DUC, three–step sample selection, and
adaptive linear regression examples, provided the rv's \( \{W_t : t \geq 1\} \) are \( m \)-dependent. For the adaptive linear regression example, \( m \)-dependence requires the errors to be 0-th order Markov. For this example when the errors are \( r \)-th order Markov for \( r \geq 1 \) and for the regression with unobserved risk variables example, the results of Sections 5 and 6 below must be used, because they allow for more general forms of temporal dependence.

4. STOCHASTIC EQUICONTINUITY VIA BRACKETING

In this section, we present a bracketing result of Ossiander (1987) for iid rv's altered to encompass \( m \)-dependent rather than independent rv's and extended as in Pollard (1989) to allow for non-identically distributed rv's. This result depends on a condition, referred to as Ossiander's entropy condition, that is based on how well the functions in \( \mathcal{H} \) can be approximated by a finite number of functions that "bracket" each of the functions in \( \mathcal{H} \). The bracketing error is measured by the largest \( L^2(P_{Tt}) \) distance over all distributions \( P_{Tt} \) of \( W_{Tt} \) for \( t \leq T \), \( T \geq 1 \). The main purpose of this section is to give primitive conditions under which Ossiander's entropy condition holds.

The results given here are particularly useful in three contexts. The first is in verifying Assumption 2(e) of ASEM:I when the nuisance parameter \( \tau \) is infinite dimensional and is a bounded smooth function with an unbounded domain. For example, realizations of smooth nonparametric estimators are sometimes of this form. Theorem II.2 above does not apply in this case. The second occurs in verifying Assumption 2(e) of ASEM:I when \( \tau \) is infinite dimensional, is a bounded smooth function on one set out of a countable collection of sets, and is constant outside this set. For example, realizations of trimmed nonparametric estimators with variable trimming sets are sometimes of this form. The third context occurs in verifying Assumption 2*(e) of ASEM:I when \( \tau \) is finite dimensional (or does not exist) and \( m(W_{Tt}, \theta, \tau) \) is a non-smooth function of some nonlinear function of \( \theta \), \( \tau \), and \( W_{Tt} \). For example, the \( m(W_{Tt}, \theta, \tau) \) function for the LAD
estimator of a nonlinear regression model is of this form. In this case, it is difficult to verify Pollard’s entropy condition, so Theorems II.1–II.3 are difficult to apply.

The pseudo–metric \( \rho_f \) on \( \mathcal{F} \) that is used in this section is defined by

\[
\rho_f(\tau_1, \tau_2) = \sup_{t \leq N, N \geq 1} \left[ \mathbb{E}(m(W_{Nt}, \tau_1) - m(W_{Nt}, \tau_2))^2 \right]^{1/2}.
\]

We adopt the following notational convention: For any real function \( f \) on \( \mathcal{Y} \),

\[
\left[ \mathbb{E}|f(W_{Tt})|^p \right]^{1/p} = \sup_{w \in \mathcal{Y}} |f(w)| \text{ if } p = \infty.
\]

An entropy condition analogous to Pollard’s is defined using the following bracketing cover numbers:

**DEFINITION:** For any \( \epsilon > 0 \) and \( p \in [2, \infty] \), the \( L^p \) bracketing cover number \( N^B_p(\epsilon, p, \mathcal{F}) \) is the smallest value of \( n \) for which there exist real functions \( a_1, \ldots, a_n \) and \( b_1, \ldots, b_n \) on \( \mathcal{Y} \) such that for each \( f \in \mathcal{F} \) \( |f - a_j| \leq b_j \) for some \( j \leq n \) and \( \max_{j \leq n} \sup_{t \leq T, T \geq 1} \left( \mathbb{E}b_j(W_{Tt}) \right)^{1/p} \leq \epsilon \), where \( \{W_{Tt} : t \leq T, T \geq 1\} \) has distribution determined by \( \mathcal{P} \).

The log of \( N^B_p(\epsilon, p, \mathcal{F}) \) is referred to as the \( L^p \) bracketing \( \epsilon \)-entropy of \( \mathcal{F} \). The following entropy condition was introduced by Ossiander (1987) (for the case \( p = 2 \)):

**DEFINITION:** A class \( \mathcal{F} \) of real functions on \( \mathcal{Y} \) satisfies Ossiander’s \( L^p \) entropy condition for some \( p \in [2, \infty] \) if

\[
\int_0^1 \left[ \log N^B_p(\epsilon, p, \mathcal{F}) \right]^{1/2} d\epsilon < \infty.
\]

As with Pollard’s entropy condition, Ossiander’s entropy condition restricts the complexity/size of \( \mathcal{F} \) by restricting the rate of increase of the \( L^p \) bracketing cover numbers as \( \epsilon \downarrow 0 \).

Often our interest in Ossiander’s \( L^p \) entropy condition is limited to the case where \( p = 2 \), as in Ossiander (1987) and Pollard (1986). To show that Ossiander’s \( L^p \) entropy
condition holds for $p = 2$ for a class of products of functions $\mathcal{G}$, however, we need to consider the case $p > 2$. The examples of ASEM:I show that the latter situation arises quite frequently in applications of interest.

ASSUMPTION D: $\mathcal{M}$ satisfies Ossiander's $L^p$ entropy condition with $p = 2$ and has envelope $\bar{M}$.

THEOREM II.4: Under Assumptions B–D (with $\bar{M}$ in Assumption B given by Assumption D rather than Assumption A), $\{\nu_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous with $\rho_T$ given by (4.1).

COMMENTS: 1. The proof of this Theorem follows easily from Theorem 2 of Pollard (1989) (as shown in the Appendix). Pollard's result in turn is based on methods introduced by Ossiander (1987).

2. As in Section 3, one establishes stochastic equicontinuity here via maximal inequalities. With the bracketing approach, however, one applies a chaining argument directly to the empirical measure rather than to a symmetrized version of it. The chaining argument relies on the Bernstein inequality for the tail probabilities of a sum of mean zero, independent rv's. The upper bound in Bernstein's inequality is small when the $L^2(P_{T_t})$ norms of the underlying rv's are small, where $P_{T_t}$ denotes the distribution of the $t$–th underlying rv. The bound ultimately is applied with the underlying rv's given by the centered difference between an arbitrary function in $\mathcal{M}$ and one of the functions from a finite set of approximating functions, each evaluated at $W_{T_t}$. In consequence, these functions need to be close in an $L^2(P_{T_t})$ sense for all $t \leq T$ for the bound to be effective, where $P_{T_t}$ denotes the distribution of $W_{T_t}$. This explains the appearance of the supremum $L^2(P_{T_t})$ norm as the measure of approximation error in Ossiander's $L^2$ entropy condition.

Next, we provide primitive conditions under which Ossiander's entropy condition is satisfied. The method is analogous to that used for Pollard's entropy condition. First, we
show that several useful classes of functions satisfy the condition. Then, we show how functions from these classes can be mixed and matched to obtain more general classes that satisfy the condition.

**DEFINITION:** A class \( \mathcal{F} \) of real functions on \( \mathcal{H} \) is called a **type IV class under \( P \) with index \( p \in [2,\infty] \)** if each function \( f \) in \( \mathcal{F} \) satisfies: \( f(\cdot) = f(\cdot, \tau) \) for some \( \tau \in \mathcal{T} \), where \( \mathcal{T} \) is some bounded subset of Euclidean space, and

\[
\sup_{t \leq T, T \geq 1} \frac{1}{\tau_1} \left[ \mathbf{E} \sup_{\|\tau_1 - \tau\| < \delta} |f(W_{Tt}, \tau_1) - f(W_{Tt}, \tau)|^p \right]^{1/p} \leq C\delta \psi
\]

\( \forall \tau \in \mathcal{T} \) and \( \forall \delta > 0 \) in a neighborhood of 0, for some finite positive constants \( C \) and \( \psi \), where \( \{W_{Tt} : t \leq T, T \geq 1\} \) has distribution determined by \( P \).

Condition (4.3) is an \( L^p \) continuity condition that weakens the Lipschitz condition (3.3) of type II classes (provided \( \sup_{t \leq T, T \geq 1} (EBP(W_{Tt}))^{1/p} < \infty \)). The \( L^p \) continuity condition allows for discontinuous functions such as sign and indicator functions. For example, for the LAD estimator of a nonlinear regression model one takes \( \theta = \tau \) and 
\[
f(W_{Tt}, \tau) = \text{sgn}(Y_t - g(X_t, \theta)) \frac{\partial}{\partial \theta_j} g(X_t, \theta)
\]
for different elements \( \theta_j \) of \( \theta \). Under appropriate conditions on \( (Y_t, X_t) \) and on the regression function \( g(\cdot, \cdot) \), the resultant class of functions can be shown to be of type IV under \( P \) with index \( p \).

Note that the conditions placed on a type IV class of functions are weaker in several respects than those placed on the functions in Huber's (1967, Lemma 3, p. 227) stochastic equicontinuity result. (Huber's conditions N–2, N–3(i), and N–3(ii) are not used here, nor is his independence assumption on \( \{W_{Tt}\} \).) Huber's result has been used extensively in the literature on M–estimators.

Next we consider an analogue of type III classes that allows for uniformly bounded functions that are smooth on an unbounded domain. (Recall that the functions of type III
are smooth only on a bounded domain and equal a constant elsewhere.) Define \( \mathcal{H}_a \) as above and let \( w = (w^a_a, w^b_b)^\prime \), \( h = (h^a_a, h^b_b)^\prime \), and \( W_{TT} = (W^a_{aTT}, W^b_{bTT})^\prime \).

**DEFINITION:** A class \( \mathcal{F} \) of real functions on \( \mathcal{H} \) is called a type \( V \) class under \( P \) with index \( p \in [2, \infty] \), if

(i) each \( f \) in \( \mathcal{F} \) depends on \( w \) only through a subvector \( w_a \) of dimension \( k_a \leq k \),

(ii) \( \mathcal{H}_a \) is such that \( \mathcal{H}_a \cap \{ w_a \in \mathbb{R}^{k_a} : \|w_a\| \leq r \} \) is a finite union of connected compact sets \( \forall r > 0 \),

(iii) for some real number \( q > k_a/2 \) and some finite constants \( C_0, \ldots, C_{[q]} \), \( C_q \), each \( f \in \mathcal{F} \) satisfies the smoothness condition: \( \forall w \in \mathcal{H} \) and \( w+h \in \mathcal{H} \) such that \( w_a \) and \( w_a + h_a \) are in the same connected set in \( \mathcal{H}_a \cap \{ w_a : \|w_a\| \leq r \} \) for some \( r > 0 \),

\[
f(w+h) = \sum_{v=0}^{[q]} \frac{[q]_1}{v!} B_v(h_a, w_a) + R(h_a, w_a),
\]

(4.4)

\[
R(h_a, w_a) \leq C_q \|h_a\|^q, \quad \text{and} \quad |B_v(h_a, w_a)| \leq C_v \|h_a\|^v \quad \text{for} \quad v = 0, \ldots, [q],
\]

where \( B_v(h_a, w_a) \) is homogeneous of degree \( v \) in \( h_a \) and \( (q, C_0, \ldots, C_q) \) do not depend on \( f \), \( w \), or \( h \),

(iv) \( \sup_{t \leq T, T \geq 1} E \|W_{aTT}\|^\zeta < \infty \) for some \( \zeta > p q k_a / (2q - k_a) \) under \( P \).

In condition (iv) above, the condition \( \zeta > \alpha \), which arises when \( p = \infty \), is taken to hold if \( \zeta = \alpha \).

Condition (ii) above holds, for example, if \( \mathcal{H}_a = \mathbb{R}^{k_a} \).

As with type III classes, the expansion of \( f(w+h) \) in (4.4) typically is a Taylor expansion and \( B_v(h_a, w_a) \) is usually the \( v \)-th differential of \( f \) at \( w \). In this case, the third condition of (4.4) holds if the partial derivatives of \( f \) of order \( \leq [q] \) are uniformly bounded.

Sufficient conditions for condition (iii) above are: (a) for some real number \( q > k_a/2 \), each \( f \in \mathcal{F} \) has partial derivatives of order \( [q] \) on \( \mathcal{H} \) that are bounded.
uniformly over \( w \in \mathcal{W} \) and \( f \in \mathcal{F} \), (b) the \([q]-\text{th}\) order partial derivatives of \( f \) satisfy a Lipschitz condition with exponent \( q - [q] \) and some Lipschitz constant \( C_q \) that does not depend on \( f \), and (c) \( \mathcal{W}_a \) is the finite union of convex sets.

The envelope of a type \( V \) class \( \mathcal{F} \) can be taken to be a constant function, since the functions in \( \mathcal{F} \) are uniformly bounded over \( w \in \mathcal{W} \) and \( f \in \mathcal{F} \).

In applications, the functions in type \( V \) classes usually are the realizations of non-parametric function estimates. For example, nonparametric kernel density estimates for bounded and unbounded rv's satisfy the uniform smoothness conditions of type \( V \) classes under suitable assumptions, see Sections 7–10 below. In addition, kernel regression estimates for bounded and unbounded regressor variables satisfy the uniform smoothness conditions if they are trimmed to equal a constant outside a suitable bounded set and then smoothed (e.g., by convolution with another kernel), see Sections 7–10 below. The bounded set in this case may depend on \( T \).

In some cases one may wish to consider nonparametric estimates that are trimmed (i.e., set equal to a constant outside some set), but not subsequently smoothed. Realizations of such estimates do not comprise a type \( V \) class because the trimming procedure creates a discontinuity. The following class of functions is designed for this scenario. The trimming sets are restricted to come from a countably infinite number of sets \( \{\mathcal{W}_{aj} : j \geq 1\} \). (This can be restrictive in practice.)

**DEFINITION:** A class \( \mathcal{F} \) of real functions on \( \mathcal{W} \) is called a type \( VI \) class under \( P \) with index \( p \in [2,\infty] \), if

(i) each \( f \) in \( \mathcal{F} \) depends on \( w \) only through a subvector \( w_a \) of \( w \) of dimension \( k_a \leq k \),

(ii) for some real number \( q > k_a/2 \), some sequence \( \{\mathcal{W}_{aj} : j \geq 1\} \) of connected compact subsets of \( \mathbb{R}^a \) that lie in \( \mathcal{W}_a \), some sequence \( \{K_j : j \geq 1\} \) of constants that satisfy \( \sup_j |K_j| < \infty \), and some finite constants \( C_0, \ldots, C_{[q]} \), \( C_q \), each \( f \in \mathcal{F} \) satisfies the smoothness condition: For some integer \( J \),
(a) \( f(w) = K_f \forall w \in \mathcal{H} \) for which \( w \notin \mathcal{H}_{\mathcal{A}} \) and

(b) \( \forall w \in \mathcal{H} \) and \( w + h \in \mathcal{H} \) for which \( w \in \mathcal{H}_{\mathcal{A}} \) and \( w + h \in \mathcal{H}_{\mathcal{A}} \),

\[
f(w+h) = \sum_{v=0}^{q} \frac{q!}{v!} B_v(h, w) + R(h, w),
\]

(4.5) \( R(h, w) \leq C_q \|h\|^q \), and \( |B_v(h, w)| \leq C_v \|h\|^v \) for \( v = 0, \ldots, [q] \),

where \( B_v(h, w) \) is homogeneous of degree \( v \) in \( h \) and \( (q, \{\mathcal{H}_j : j \geq 1\}, C_0, \ldots, C_q) \) do not depend on \( f, w, \) or \( h \).

(iii) \( \sup_{t \leq T, t \geq 1} E\|W_{\text{At}}\|^\zeta < \infty \) for some \( \zeta > pqk_2/(2q-k_2) \) under \( P \),

(iv) \( n(\varepsilon) \leq K_1 \exp(K_2 \varepsilon) \) for some \( \xi < 2\xi/p \) and some finite constants \( K_1, K_2 \), where \( n(\varepsilon) \) is the number of sets \( \mathcal{H}_{\mathcal{A}} \) in the sequence \( \{\mathcal{H}_j : j \geq 1\} \) that do not include \( \{w \in \mathcal{H}_a : \|w\| \leq \varepsilon\} \).

Conditions (i)–(iii) in the definition of a type VI class are quite similar to conditions used above to define type III and type V classes. The difference is that with a type VI class, the set on which the functions are smooth is not a single set, but many vary from one function to the next among a countably infinite number of sets.

Condition (iv) restricts the number of \( \mathcal{H}_{\mathcal{A}} \) sets that may be of a given radius or less. Sufficient conditions for condition (iv) are the following: Suppose \( \mathcal{H}_j \supset \{w \in \mathcal{H}_a : \|w\| \leq \eta(j)\} \) for all \( j \) sufficiently large, where \( \eta(\cdot) \) is a nondecreasing real function on the positive integers that diverges to infinity as \( j \to \infty \). For example, \( \{\mathcal{H}_j : j \geq 1\} \) could contain spheres, ellipses, and/or rectangles whose "radii" are large for large \( j \). If

\[
\eta(j) \geq D_1 \left[ \log j \right]^{1/\xi} D_2 \ \
\]

(4.6) for some positive finite constants \( D_1, D_2 \), then condition (iv) holds. Thus, the "radii" of the sets \( \{\mathcal{H}_j : j \geq 1\} \) are only required to increase logarithmically for condition (iv). This condition is not too restrictive, given that the number of trimming sets \( \{\mathcal{H}_j\} \) is
countable. More restrictive is the latter condition that the number of trimming sets \( \{ \mathcal{N}_j \} \) is countable.

As with type III and type V classes, the envelope of a type VI class of functions can be taken to be a constant function.

The trimmed kernel regression estimators discussed in Sections 7-10 below provide examples of nonparametric function estimates for which type VI classes are applicable. For suitable trimming sets \( \{ \mathcal{N}_j : j \geq 1 \} \) and suitable smoothness conditions on the true regression function, one can specify a type VI class that contains all of the realizations of such kernel estimators in a set whose probability \( \rightarrow 1 \).

The following result establishes Ossiander's \( L^p \) entropy condition for classes of type II-VI:

**Theorem II.5:** Let \( p \in [2,\infty] \). If \( \mathcal{F} \) is a class of functions of type II with

\[
\sup_{t \leq T_1, T_2 \geq t} \left[ \mathbb{E}_P(W_{T_1}) \right]^{1/p} < \infty, \text{ of type III, of type IV under } P \text{ with index } p, \text{ or of type V under } P \text{ with index } p, \text{ then Ossiander's } L^p \text{ entropy condition (4.2) holds (with envelope } F(\cdot) \text{ given by } \sup_{f \in \mathcal{F}} |f(\cdot)| \text{).}
\]

**Comments:**

1. To obtain Assumption D for any of the classes of functions considered above, one only needs to consider \( p = 2 \) in Theorem II.5. To obtain Assumption D for a class of the form \( \mathcal{G}_\mathcal{V} \), where \( \mathcal{G} \) and \( \mathcal{V} \) are classes of types II, III, IV, V, or VI, however, one needs to apply Theorem II.5 to \( \mathcal{G} \) and \( \mathcal{V} \) for values of \( p \) greater than 2, see Theorem II.6 below.

2. Theorem II.5 covers classes containing a finite number of functions, because such functions are of type IV under any distribution \( P \) and for any index \( p \in [2,\infty] \). In particular, this is true for classes containing a single function. This observation is useful when establishing Ossiander's \( L^p \) entropy condition for classes of functions that can be obtained by mixing and matching functions from several classes, see below.
We now show how one can "mix and match" functions of types II–VI. Let \( \mathcal{G} \), \( \mathcal{G}^* \), \( \mathcal{X} \), \( \mathcal{G} \odot \mathcal{G}^* \), etc., be as defined in Section 3. We say that a class of matrix–valued functions \( \mathcal{G} \), \( \mathcal{G}^* \), or \( \mathcal{X} \) satisfies Ossiander's \( L_p \) entropy condition or is of type II, III, IV, V, or VI if it does so, or if it is, element by element for each of the rs or su elements of its functions. We adopt the convention that \( \lambda_{\mu/}(\lambda+\mu) = \mu \in (0,\omega] \) if \( \lambda = \omega \) and vice versa.

**THEOREM II.6:** (a) If \( \mathcal{G} \) and \( \mathcal{G}^* \) satisfy Ossiander's \( L_p \) entropy condition for some \( p \in [2,\omega] \), with envelopes \( \mathcal{G} \) and \( \mathcal{G}^* \), respectively, then so do each of the following classes (with envelopes given in parentheses): \( \mathcal{G} \cup \mathcal{G}^* \) \( G \vee G^* \), \( \mathcal{G} \otimes \mathcal{G}^* \) \( G + G^* \), \( \mathcal{G} \vee \mathcal{G}^* \) \( G \vee G^* \), \( \mathcal{G} \wedge \mathcal{G}^* \) \( G \wedge G^* \), and \( \mathcal{G} \) \( G \). If in addition \( r = s \) and \( \inf_{g \in \mathcal{G}} \inf_{w \in \mathcal{W}} \lambda_{\min}(g(w)) = \lambda_* \) for some \( \lambda_* > 0 \), then \( \mathcal{G}^{-1} \) also satisfies Ossiander's \( L_p \) entropy condition (with envelope \( r/\lambda_* \)).

(b) The class \( \mathcal{G} \mathcal{X} \) satisfies Ossiander's \( L_p \) entropy condition with \( p \) equal to \( \alpha \in [2,\omega] \) and envelope \( s \mathcal{G} \mathcal{H} \), if (i) \( \mathcal{G} \) and \( \mathcal{X} \) satisfy Ossiander's \( L_p \) entropy condition with \( p \) equal to \( \lambda \in (\alpha,\omega] \) and \( p \) equal to \( \mu \in (\alpha,\omega] \), respectively, (ii) \( \frac{\lambda \mu}{\lambda + \mu} \geq \alpha \), and (iii) the envelopes \( \mathcal{G} \) and \( \mathcal{H} \) of \( \mathcal{G} \) and \( \mathcal{X} \) satisfy \( \sup_{t \leq T, T \geq 1} \left[ E_{\mathcal{G}}(W_{Tt}) \right]^{1/\lambda} < \omega \) and \( \sup_{t \leq T, T \geq 1} \left[ E_{\mathcal{H}}(W_{Tt}) \right]^{1/\mu} < \omega \).

**EXAMPLE:** We now illustrate the use of Theorems II.4–II.6 in verifying Assumption 2(e) of ASEM:1.

(1) **Generalized method of moments/conditional moments restriction (GMM/CMR)**

example of ASEM:1. In this example,

\[ m(w, \theta_0, \tau) = \tau(x) \psi(z, \theta_0) = \Delta(x) \Omega^{-1}(x) \psi(z, \theta_0) \]

and

\[ \mathcal{U} = \{ m(\cdot, \theta_0, \tau) : \tau \in \mathcal{T} \}, \]

where \( w = (z', x') \) and \( \mathcal{T} \) is as defined below. With some abuse of notation, let \( \Delta(w) \) and \( \Omega(w) \) denote functions on \( \mathcal{W} \) whose values depend on \( w \) only through the \( k_a \)–vector
x and equal \( \Delta(x) \) and \( \Omega(x) \) respectively. Similarly, let \( \psi(w, \theta_0) \) denote the function on \( \mathcal{W} \) that depends on \( w \) only through \( z \) and equals \( \psi(z, \theta_0) \).

The following conditions are sufficient for Assumption 2(e) of ASEM:I with \( \rho_T \) given by (4.1) with \( m(\cdot, \tau) \) replaced by \( m(\cdot, \theta_0, \tau) \):

(i) \( \{ (Z_t, X_t) : t \geq 1 \} \) is an \( m \)-dependent sequence of rv's.

(ii) \( \sup_{t \geq 1} E[|\psi(Z_t, \theta_0)|^6] < \infty \).

(iii) \( J = \{ \tau : \tau = \Delta \cdot \Omega^{-1} \text{ for some } \Delta \in \mathcal{D} \text{ and } \Omega \in \mathcal{A} \} \), where \( \mathcal{D} \) and \( \mathcal{A} \) are type V or type VI classes of functions on \( \mathcal{W} \subset \mathbb{R}^k \) with index \( p = 6 \) whose functions depend on \( w \) only through the \( k_a \)-vector \( x \), and

\[ \mathcal{A} \subset \{ \Omega : \inf_{w \in \mathcal{W}} \lambda \min_{w \in \mathcal{W}} (\Omega(w)) \geq \epsilon \} \text{ for some } \epsilon > 0. \]

By definition, the functions \( \Delta \cdot \Omega \in \mathcal{D} \) and \( \Omega \in \mathcal{A} \) are \( \mathbb{R}^{p \times \alpha} \) and \( \mathbb{R}^{\alpha \times \alpha} \) valued, where \( p \) in this context denotes the dimension of \( \theta \) not the index of some class. Note that condition (iii) of (4.8) includes a moment condition on \( X_t : \sup_{t \geq 1} E[||X_t||^\zeta] < \infty \) for some \( \zeta > 6q_{k_a}/(2q - k_a) \).

Sufficiency of conditions (i)-(iii) for Assumption 2(e) is established as follows: By Theorem II.5, \( \{ \psi(\cdot, \theta_0) \} \), \( \mathcal{D} \), and \( \mathcal{A} \) satisfy Ossiander's \( L^p \) entropy condition with \( p = 6 \) and with envelopes \( \{|\psi(\cdot, \theta_0)|, C_1, \text{ and } C_2\} \), respectively, for some finite constants \( C_1, C_2 \). By the \( G^{-1} \) result of Theorem II.6, so does \( \mathcal{A}^{-1} \) with some constant envelope \( C_3 < \infty \). By the \( G\mathcal{A} \) result of Theorem II.6 applied with \( \alpha = 3 \) and \( \lambda = \mu = 6 \), \( \mathcal{D}\mathcal{A}^{-1} \) satisfies Ossiander's \( L^p \) entropy condition with \( p = 3 \) and some constant envelope \( C_4 < \infty \). By this result, condition (ii), and the \( G\mathcal{A} \) result of Theorem II.6 applied with \( \alpha = 2, \lambda = 3, \mu = 6, \mathcal{G} = \mathcal{D}\mathcal{A}^{-1} \), and \( \mathcal{X} = \{ \psi(\cdot, \theta_0) \} \), \( \mathcal{H} \) satisfies Ossiander's \( L^p \) entropy condition with \( p = 2 \) and envelope \( C_5 \) \( \{ \psi(\cdot, \theta_0) \} \) for some constant \( C_5 < \infty \). Theorem II.4 now yields stochastic equicontinuity, since condition (ii) is sufficient for Assumption B.
Condition (iii) above covers the case where the domain of the nonparametric functions is unbounded and the nonparametric estimators $\hat{\Delta}$ and $\hat{\Omega}$ are not trimmed to equal zero outside a single fixed bounded set, as is required when the symmetrization results of Section 3 are applied. As discussed above and in Section 7 below, nonparametric kernel regression estimators that are trimmed and smoothed or trimmed on variable sets provide examples where condition (iii) holds under suitable assumptions for realizations of the estimators that lie in a set whose probability $\to 1$. Sections 7–10 below provide uniform consistency on expanding sets and $L^Q$ consistency results for such estimators, as are required to establish that $\mathbb{P}(\hat{\tau} \in \mathcal{T}) \to 1$ and $\hat{\tau} \in \Omega_{0}$ (the first and second parts of Assumptions 2(b) of ASEM:I) when Assumption 2(e) is established using conditions (i)–(iii) above.

Note that it is much more complicated to establish Assumption 2 of ASEM:I using type V or type VI classes of nonparametric functions than type III classes, because type III classes allow one to exploit uniform consistency results over fixed sets very effectively. On the other hand, the use of type V or type VI classes allows one to use all of the data to construct MINPIN estimators even when the nonparametric functions have unbounded domain — the trimming of observations vanishes as $T \to \omega$.

Theorems II.4–II.6 can be used to verify Assumption 2(e) or $2^*(e)$ using type V or type VI classes of nonparametric functions in the WLS/PLR and WC–LAD examples discussed above, as well as in some other examples discussed in ASEM:I. For brevity, we do not discuss the details here.

5. STOCHASTIC EQUICONTINUITY VIA SERIES EXPANSIONS

In this section a stochastic equicontinuity result of Andrews (1989a) is presented that is based on series expansions of the functions in $\mathcal{M}$. The method used is quite simple — no chaining argument is needed. The method leads to the most general stochastic
equicontinuity results available with respect to temporal dependence. On the other hand, the classes of functions \( \mathcal{K} \) to which it applies are more restricted than those of Sections 3 and 4 — the functions must be smooth on a bounded set and constant elsewhere. Still, the classes are sufficiently general to cover many examples from ASEM:1.

The method of proof of stochastic equicontinuity via series expansions can be described quite easily. Suppose each function in \( \mathcal{K} \) has a pointwise convergent series expansion of the form

\[
(5.1) \quad m(w, \tau) = \sum_{j=1}^{\infty} c_j(\tau) h_j(w),
\]

where \( \{h_j(w) : j \geq 1\} \) is a set of bounded complex functions and \( \{c_j(\tau) : j \geq 1\} \) is a set of complex series coefficients for each \( \tau \in \mathcal{T} \). Let \( \{v_j : j \geq 1\} \) be a summable sequence of positive real constants. Then,

\[
(5.2) \quad \lim_{T \to \infty} \mathbb{P}^* \left[ \sup_{\rho_T(\tau_1, \tau_2) < \delta} \left| \nu_T(\tau_1) - \nu_T(\tau_2) \right| > \eta \right]
\]

\[
\leq \eta^{-2} \lim_{T \to \infty} \mathbb{E}^* \left[ \sup_{\rho_T(\tau_1, \tau_2) < \delta} \left| \frac{1}{\sqrt{T}} \sum_{j=1}^{\infty} (c_j(\tau_1) - c_j(\tau_2)) (h_j(W_t) - Eh_j(W_t)) \right| \right]^2
\]

\[
= \eta^{-2} \lim_{T \to \infty} \mathbb{E}^* \left[ \sup_{\rho_T(\tau_1, \tau_2) < \delta} \left| \sum_{j=1}^{\infty} \frac{c_j(\tau_1) - c_j(\tau_2)}{\sqrt{v_j}} \frac{1}{\sqrt{T}} h_j(W_t) - Eh_j(W_t) \right| \right]^2
\]

Under fairly general weak dependence conditions, the term in parentheses on the right-hand side is finite. Hence, stochastic equicontinuity holds provided \( \rho_T \) is such that

\[
(5.3) \quad \lim_{\delta \to 0} \sup_{\rho_T(\tau_1, \tau_2) < \delta} \sum_{j=1}^{\infty} \left| c_j(\tau_1) - c_j(\tau_2) \right|^2 / v_j = 0.
\]

With the class of smooth functions, the pseudo--metric \( \rho_T \), and the weak dependence
conditions introduced below, each of the conditions outlined above is satisfied and the argument goes through.

We now introduce the definitions and notation necessary for stating the series expansion stochastic equicontinuity result. As above, $\mathcal{W}$ is the domain of the functions in $\mathcal{M}$. Let $\mathcal{W}^*$ be a bounded subset of $\mathcal{W}$. The functions $m(\cdot, \tau) \in \mathcal{M}$ are taken to be uniformly smooth in $w$ on $\mathcal{W}^*$ in the sense of having a uniformly bounded Sobolev norm over $\mathcal{W}^*$ of some order. On $\mathcal{W} - \mathcal{W}^*$, all of the functions $m(\cdot, \tau) \in \mathcal{M}$ are taken to equal some constant $K$.

As defined, two cases are covered. The first is the case where $\mathcal{W} = \mathcal{W}^*$, the rv's $\{W_{T_t}\}$ are bounded, and the functions in $\mathcal{M}$ are uniformly smooth on their entire domain. The second is the case where $\mathcal{W}^*$ is a proper subset of $\mathcal{W}$, the rv's $\{W_{T_t}\}$ may be unbounded, and the functions in $\mathcal{M}$ are smooth on $\mathcal{W}^*$ and constant elsewhere. The latter case is useful because it covers classes of realizations of nonparametric estimates that are trimmed in order to obtain uniform consistency over $\mathcal{W}^*$ of the estimates and their derivatives.

By definition, the Sobolev norm over $\mathcal{W}^*$ of order $(q, 2)$ of $m(\cdot, \tau)$ is

$$
\|m(\cdot, \tau)\|_{q, 2, \mathcal{W}^*} = \left[ \sum_{|\alpha| \leq q} \int_{\mathcal{W}^*} (D^\alpha m(w, \tau))^2 dw \right]^{1/2},
$$

where $q$ is a non-negative integer, $\alpha = (\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$ has non-negative integer elements, $|\alpha| = \sum_{\ell=1}^{k} \alpha_\ell$, and $D^\alpha m(w, \tau) = \partial |\alpha| m(w, \tau)/(\partial w_1^{\alpha_1} \times \ldots \times \partial w_k^{\alpha_k})$.

Below we assume that $\sup_{\tau \in \mathcal{I}} \|m(\cdot, \tau)\|_{q, 2, \mathcal{W}^*} < \infty$ for some $q > (k+1)/2$. That is, on $\mathcal{W}^*$, the functions $\{m(\cdot, \tau)\}$ must have more derivatives $q$ with respect to $w$ finite than the dimension $k$ of their domain plus one and divided by two.

The set $\mathcal{W}^*$ is assumed to be some open bounded subset of $\mathbb{R}^k$ whose boundary is minimally smooth. Examples of sets in $\mathbb{R}^k$ with minimally smooth boundaries include open bounded sets that are convex or whose boundaries are $C^1$—embedded in $\mathbb{R}^k$. Finite
unions of disjoint sets of the aforementioned type also have minimally smooth boundaries.

Define a pseudo—metric $\rho_T$ on $T$ by

$$\rho_T(\tau_1, \tau_2) = \left[ \int_{\mathcal{W}} (m(w, \tau_1) - m(w, \tau_2))^2 dw \right]^{1/2}.$$  

The rv's $\{W_{Tt}\}$ are assumed to be near—epoch dependent (NED) (also referred to in the literature as "functions of mixing processes"). The NED condition is one of asymptotically weak temporal dependence. It was introduced by Ibragimov (1962) and results utilizing it were developed by Billingsley (1968, p. 182) and McLeish (1975a, b, 1977) among others. The NED condition is quite general. It allows for non—identical distributions and covers (i) square integrable strong mixing rv's, (ii) square integrable general linear processes (with non—identically distributed strong mixing innovations if desired) including autoregressive—moving average processes (which are not necessarily strong mixing, e.g., see Andrews (1985)), and (iii) various nonlinear autoregressions and dynamic simultaneous equations (see Bierens (1981, Ch. 5), Gallant (1987b, pp. 481, 502, 539), and Gallant and White (1988, p. 29)).

We now define strong mixing double arrays and NED triangular arrays of rv's. Let $\{V_{Tt} : t = 0, \pm 1, \pm 2, \ldots; T \geq 1\}$ be a double array of rv's defined on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$. ( $V_{Tt}$ may be $\mathcal{V}_{Tt}$—valued for any measurable space $\mathcal{V}_{Tt}$, but usually $\mathcal{V}_{Tt} = \mathcal{V}$ and $\mathcal{V} \subset \mathbb{R}^j$ for some $j \geq 1$.) Let $\mathcal{F}_{T,s}^t (\subset \mathcal{B})$ denote the $\sigma$—field generated by $(V_{Ts}, \ldots, V_{Tt})$ for $-\infty \leq s \leq t \leq \infty$.

**DEFINITION:** The double array $\{V_{Tt}\}$ of rv's is strong mixing if $\alpha(s) \downarrow 0$ as $s \to \infty$, where

$$\alpha(s) = \sup_{t=0, \pm 1, \ldots; T \geq 1} \sup_{A \in \mathcal{F}_{T,-\infty}^t, B \in \mathcal{F}_{T,t+s}^s} |P(A \cap B) - P(A)P(B)|$$

for $s \geq 1$.

$$\{V_{Tt}\} \text{ is strong mixing of size } -\beta \text{ if } \alpha(s) = O(s^{-\beta-\epsilon}) \text{ for some } \epsilon > 0.$$
DEFINITION: The triangular array \( \{ W_{T_t} : t = 1, \ldots, T; T \geq 1 \} \) of \( R^k \)-valued r.v.'s is near-epoch dependent (NED) on \( \{ V_{T_t} \} \) if \( \mathbb{E} \| W_{T_t} \|^2 < \infty \) \( \forall t \leq T, \forall T \geq 1 \), and \( \eta(s) \downarrow 0 \) as \( s \to \infty \), where

\[
\eta(s) = \sup_{s < t \leq T - s, T \geq 1} \left( \mathbb{E} \| W_{T_t} - \mathbb{E}(W_{T_t} | T_t, t-s) \|^2 \right)^{1/2} \text{ for } s \geq 0.
\]

\( \{ W_{T_t} \} \) is NED of size \(-\beta\) on \( \{ V_{T_t} \} \) if \( \mathbb{E} \| W_{T_t} \|^2 < \infty \) \( \forall t \leq T, \forall T \geq 1 \), and \( \eta(s) = O(s^{-\beta - \epsilon}) \) for some \( \epsilon > 0 \).

The following assumption is sufficient for stochastic equicontinuity of \( \{ \nu_T(\cdot) : T \geq 1 \} \):

ASSUMPTION E: (i) \( \mathcal{H}^* \) is contained in \( \mathcal{H} \) and is an open bounded subset of \( R^k \) with minimally smooth boundary.

(ii) \( \sup_{T \in T} \| m(\cdot, \tau) \|_{q,2,\mathcal{H}^*} < \infty \) for some \( q > (k+1)/2 \).

(iii) For some constant \( K \), \( m(w, \tau) = K \) \( \forall w \in \mathcal{H} - \mathcal{H}^* \), \( \forall \tau \in T \).

(iv) \( \{ W_{T_t} \} \) is a NED triangular array of size \(-1\) on \( \{ V_{T_t} \} \), where \( \{ V_{T_t} \} \) is some strong mixing double array of size \(-2\).

The constancy of \( m(w, \tau) \) except on a bounded set, which is imposed in Assumption E, can be restrictive in some applications. It can be relaxed somewhat at the expense of strengthening the temporal dependence assumption on \( \{ W_{T_t} \} \). Let \( w_a \) be a \( k_a \)-dimensional subvector of \( w \), i.e., \( w = (w_a', w_b')' \). Define \( \mathcal{H}_a = \{ w_a \in R^{k_a} : \exists w_b \) s.t. \( (w_a', w_b')' \in \mathcal{H} \} \). We now consider functions \( m(w, \tau) \) that are the product of two functions, of which the first depends only on \( w_a \), is smooth on a bounded subset \( \mathcal{H}_a^* \) of \( \mathcal{H}_a \), and is constant elsewhere, and of which the second is a function of \( w \) that does not depend on \( \tau \). In the applications of ASEM:I, classes of functions of this sort allow for models with unbounded errors \( \{ U_t : t \geq 1 \} \), because the error \( U_t \) often arises in the functions \( m(W_{T_t}, \tau) \) as a multiplicand that does not depend on \( \tau \).
ASSUMPTION F: (i) \( m(W_{Tt}, \tau) \) is of the form \( m_a(W_{aTt}, \tau)g(W_{Tt}) \), where \( W_{Tt} = (W_{aTt}, W_{bTt})' \).

(ii) \( \mathcal{W}^*_a \) is contained in \( \mathcal{W}_a \) and is an open bounded subset of \( \mathbb{R}^k \) with minimally smooth boundary (where \( k_a \leq k \)).

(iii) \( \sup_{\tau \in \mathcal{T}} \| m_2(\cdot, \tau) \|_{q, 2, \mathcal{W}^*_a} < \infty \) for some integer \( q > k_a / 2 \).

(iv) For some constant \( K \), \( m_2(w_a, \tau) = K \) \( \forall w_a \in \mathcal{W}_a - \mathcal{W}^*_a \), \( \forall \tau \in \mathcal{T} \).

(v) \( \sup_{t \leq T, T \geq 1} \| g(W_{Tt}) \|_{r} < \infty \) for some \( r > 2 \).

(vi) \( \{ W_{Tt} \} \) is a strong mixing triangular array of size \( -2r/(r-2) \) for \( r \) as in (v).

THEOREM II.7: Under Assumption E or F, \( \{ \nu_T(\cdot): T \geq 1 \} \) is stochastically equicontinuous using the pseudo-metric \( \rho_T \) defined in (5.5), where \( m(\cdot, \tau) \) and \( \mathcal{W}_a \) are replaced in the pseudo-metric by \( m_a(\cdot, \tau) \) and \( \mathcal{W}^*_a \), respectively, when Assumption F is used.

EXAMPLES: Theorem II.7 can be applied to any of the examples of ASEM:I for which (a) Assumption 2 is applied and (b) the nonparametric nuisance parameter estimator \( \hat{\tau} \) is a smooth bounded function of its argument and is constant outside a fixed bounded set, with probability \( \rightarrow 1 \). In particular, it can be applied to all of the examples of ASEM:I (under suitable assumptions), except the WC-LAD example. The unobserved risk variables example of ASEM:I is inherently a time series example that usually does not exhibit \( m \)-dependence. Hence, the results of Sections 3 and 4 are not applicable to it. The same is true of the adaptive linear regression example of ASEM:I for the case where the errors are Markov of order \( r \geq 1 \).

We now consider the use of Theorem II.7 in two particular examples:

(1) Regression with unobserved risk variables. As in ASEM:I, we consider a semiparametric instrumental variables (SIV) estimator. The SIV estimator is a MINPIN estimator with \( W_t = (Y_t, X_t', \psi_t, V_{1t}', V_{2t}')' \), \( d(m, \gamma) = m'm/2 \).
\[ m(w, \theta_0, \tau) = \xi(v_1, v_2)(y - x' \theta_{10} - (\psi - \tau_1(v_1))^2 \theta_{20}) \begin{bmatrix} x \\ \tau_2(v_2) \end{bmatrix} \]

\[ = \xi(v_1, v_2)(y - x' \theta_{10} - \psi^2 \theta_{20}) \begin{bmatrix} x \\ \tau_2(v_2) \end{bmatrix} + 2\xi(v_1, v_2)\psi \begin{bmatrix} \tau_1(v_1)x \\ \tau_1(v_1)\tau_2(v_2) \end{bmatrix} \]

(5.8)

\[ - \xi(v_1, v_2) \begin{bmatrix} \tau_1^2(v_1)x \\ \tau_1^2(v_1)\tau_2(v_2) \end{bmatrix} , \]

where \( w = (y, x', \psi, v_1, v_2)' \), \( \theta_0 = (\theta_{10}', \theta_{20})' \), \( \xi(v_1, v_2) = 1(v_j \in \mathcal{V}_j^p \text{ for } j = 1, 2) \), \( \mathcal{V}_j^p \)

is a subset of \( \mathcal{V}_j \), \( \mathcal{V}_j \subset \mathbb{R}^p \) is the domain of \( \tau_j \) and includes the support of \( V_{jt} \) \( \forall t \geq 1 \), for \( j = 1, 2 \). In (5.8), the possibility of trimming is made explicit. If \( \mathcal{V}_j^p = \mathcal{V}_j \), then no trimming occurs and the indicator function \( \xi(v_1, v_2) \) is redundant. If \( \mathcal{V}_j^p \) is a proper subset of \( \mathcal{V}_j \) for \( j = 1 \) or \( j = 2 \), then trimming occurs and the SIV estimator \( \hat{\theta} \) is based on only non-trimmed observations.

Assumption 2(e) can be established using Assumption F and Theorem II.7 by treating each of the three summands of \( m(w, \theta_0, \tau) \) separately. That is, one applies Theorem II.7 three times with \( m(w, \tau) \) equal to one of the three summands each time. The pseudo-metric \( \rho_T \) in this case equals \( \rho_1 + \rho_2 + \rho_3 \), where \( \rho_\ell \) denotes the pseudo-metric of (5.5) with \( m(w, \tau) \) given by the \( \ell \)-th summand of \( m(w, \theta_0, \tau) \) for \( \ell = 1, 2, 3 \). With this pseudo-metric, Assumption 2(e) holds under the following conditions: For some \( r > 2 \),
(i) \( \{W_t : t \geq 1\} \) is a strong mixing sequence of rv's of size \(-2\tau/(r-2)\),

(ii) \( \sup_{t \geq 1} \left[ E \left| Y_t - X_t \theta_0 - \psi_t^2 \theta_{20} \right| \right] + E\left\| \psi_t X_t \right\|^r + E\left\| \psi_t \right\|^r + E\left\| X_t \right\|^r < \infty \),

(iii) \( T = \{ \tau : \tau = (\tau_1, \tau_2), \tau_j \in T_j \text{ for } j = 1, 2 \} \text{ and } 
\[ T_j = \{ \tau_j : \forall \eta_j \in R^p, \left\| \tau_j \right\|_{q,2} \psi_j^* \leq C_j \text{ and } \tau_j(v_j) = 0 \forall v_j \in \psi_j - \psi_j^* \} \]

for some \( q > k_a/2 \) and some \( C_j < \infty \) for \( j = 1, 2 \), where \( k_a \) denotes the number of non-redundant elements of \( (V_{1t}, V_{2t}) \), \( \psi_1^* \), \( \psi_2^* \), and \( \psi_1^* \times \psi_2^* \) are open bounded convex subsets of \( R^1 \), \( R^2 \), and \( R^k \) that are contained in \( \psi_1 \), \( \psi_2 \), and \( \psi_1 \times \psi_2 \), respectively, and \( \psi_1^* \times \psi_2^* \) and \( \psi_1 \times \psi_2 \) denote product sets with redundant elements of \( (V_{1t}, V_{2t}) \) deleted.

Condition (iii) requires \( \tau_j \) to be zero outside a bounded set \( \psi_j^* \) for \( j = 1, 2 \). In consequence, either \( V_{jt} \) is a bounded rv or the nonparametric estimator \( \hat{\tau}_j \) is a trimmed estimator with a single fixed trimming set. In the latter case, all of the data is usually used to compute \( \hat{\tau}_j \), but only the data for which \( V_{jt} \in \psi_j^* \) for \( j = 1, 2 \) is used by the MINPIN estimator \( \hat{\theta} \) of \( \theta_0 \).

Sufficiency of conditions (i)–(iii) for Assumption 2(e) is established as follows: By Theorem II.7, it suffices to establish Assumption F separately for each of the three summands of \( m(w, \theta_0, \tau) \). For the first summand, \( m(w, \tau) \) is a vector with two subvectors that correspond to \( x \) and \( \tau_2(v_2) \). For the subvector that corresponds to \( x \), \( m(w, \tau) \) does not depend on \( \tau \), and hence stochastic equicontinuity holds trivially. For the subvector that corresponds to \( \tau_2(v_2) \), Assumption F(i) holds with \( m_a(W_{at}, \tau) = 1(V_{2t} \in \psi^*) \tau_2(V_{2t}) \) and \( g(W_{Tt}) = \xi(V_{1t}, V_{2t})(Y_t - X_t \theta_0 - \psi_t^2 \theta_{20}) \), F(ii) holds with \( \psi_a^* \) and \( \psi_a \) given by \( \psi_2^* \) and \( \psi_2 \) by (iii), F(iii) and F(iv) hold by (iii), F(v) holds by (ii), and F(vi) holds by (i). The proof for the last two summands of \( m(w, \theta_0, \tau) \) is similar, noting that condition (iii) implies that \( \|\tau_1 \tau_2\|_{q,2} \psi_1^* \psi_2^* \leq C_1^* \) and \( \|\tau_1 \tau_2\|_{q,2} \psi_1^* \psi_2^* \leq C_2^* \).
for all \( \tau_1 \in \mathcal{T}_1 \) and \( \tau_2 \in \mathcal{T}_2 \), for \( q > k_a/2 \), for some finite constants \( C_1^*, C_2^* \), as is required by Assumption F(iii) in these cases.

Next, we consider verifying the condition \( P(\hat{\tau} \in \mathcal{T}) \to 1 \) of Assumption 2(b) of ASEM:I, when \( \mathcal{T} \) is as in condition (iii). For this, it suffices to establish uniform consistency over the bounded set \( \mathcal{V}^*_j \) of \( \hat{\tau}_j \) and its partial derivatives of order \( \leq q \) for \( \tau_{j0}\)
and its corresponding partial derivatives, for some integer \( q > k_a/2 \) and \( j = 1, 2 \), provided the latter are uniformly bounded over \( \mathcal{V}^*_j \). For kernel estimators such uniform consistency results can be established using the results of Sections 7–10 below.

(2) MAD–DUC estimator example of ASEM:I. Here we use Assumption E and Theorem II.7 to verify Assumption 2(e). In this example,

\[
m(w, \theta_0, \tau) = \eta'(y, \tau(\theta_0, h(x, \theta_0))) \frac{\partial}{\partial \theta_0} \tau(\theta_0, h(x, \theta_0))
\]

\[
= \eta'(y, \tau(\theta_0, v))[\tau^{(1)}(\theta_0, v) + d' \tau^{(2)}(\theta_0, v)]
\]

(5.10)

\[\mathcal{M} = \{ m(\cdot, \theta_0, \tau) : \tau \in \mathcal{T} \}, \]

where \( w = (y, v, d)' \), \( v = h(x, \theta_0) \), \( d = \frac{\partial}{\partial \theta_0} h(x, \theta_0) \), and \( \tau^{(j)}(\cdot, \cdot) \) denotes the derivative of \( \tau(\cdot, \cdot) \) with respect to its \( j \)-th argument for \( j = 1, 2 \).

Let \( \Theta \) be a subset of \( \mathbb{R}^p \) that contains a neighborhood of \( \theta_0 \). Let \( \mathcal{V} \) be a subset of \( \mathbb{R}^{k_a} \) that contains the support of \( V_t = h(X_t, \theta_0) \) \( \forall t \geq 1 \). Let \( \tau^{(0)}(\theta_0, \cdot) \) denote \( \tau(\theta_0, \cdot) \).

For Ichimura's (1985) and Ichimura and Lee's (1990) MAD–DUC estimators, \( \eta'(y, \tau) = -(y-\tau) \). In this case, the following conditions are sufficient for Assumption 2(e) with the pseudo–metric \( \rho_T \) given by (5.5) with \( m(\cdot, \theta_0, \tau) \) in place of \( m(\cdot, \tau) \) and with \( \mathcal{W}^* = \mathcal{Y} \times \mathcal{V}^* \times \mathcal{D} \), where \( \mathcal{Y} \), \( \mathcal{V}^* \), and \( \mathcal{D} \) are as defined below:
(i) \(\{(Y_t, h(X_t, \theta_0), \partial_{\theta_0} h(X_t, \theta_0)) : t \geq 1\}\) is a NED sequence of size \(-1\) on some strong mixing double sequence of size \(-2\).

(ii) \(Y_t\) and \(\partial_{\theta_0} h(X_t, \theta_0)\) are bounded rv's uniformly over \(t \geq 1\).

(iii) \(I = \{\tau : \Theta \times \mathcal{Y} \to R | \tau(j)(\theta_0, \cdot) \in T_j \text{ for } j = 0, 1, 2\}\) and

\[
T_j = \{\tau : \mathcal{Y} \to R^j | \|\tau\|_{q, 2, \mathcal{Y}^*} \leq C_j \text{ and } \tau(v) = 0 \text{ for } v \in \mathcal{Y} - \mathcal{Y}^*\} \text{ for some integer } q > k_a + 1, \text{ for some constant } C_j < \infty, \text{ for some set } \mathcal{Y}^* \text{ that is contained in } \mathcal{Y} \text{ and is a finite union of open bounded convex subsets of } R^{k_a}, \text{ for } j = 0, 1, 2, \text{ where } p_0 = 1, p_1 = p, p_2 = k_a, p \text{ is the dimension of } \theta_0, \text{ and } k_a \text{ is the dimension of } h(X_t, \theta_0).\]

By definition, \(\mathcal{Y}\) is an open bounded interval in \(R\) that contains the support of \(Y_t\) \(\forall t \geq 1\) and \(\mathcal{T}\) is an open bounded rectangle in \(R^{k_a \times p}\) that contains the support of \(\partial_{\theta_0} h(X_t, \theta_0)\) \(\forall t \geq 1\).

If necessary, the boundedness of \(Y_t\) and \(\partial_{\theta_0} h(X_t, \theta_0)\) (condition (ii)) can be replaced by moment conditions on these rv's if Assumption F is used in place of Assumption E when verifying Assumption 2(e). The cost of such a change is that condition (i) needs to be strengthened to require the rv's to be strong mixing, rather than NED.

For Klein and Spady's (1987) MAD-DUC estimator, \(\eta'(y, \tau) = -(y - \tau)/(\tau(1 - \tau))\). In this case, conditions (i)–(iii) of (5.11) plus the following condition are sufficient for Assumption 2(e) with \(p_T\) as defined for Ichimura's and Ichimura and Lee's estimators:

\[
(iv) \mathcal{T} \subset \{\tau : \mathcal{Y} \to R | \tau(\epsilon) \leq \tau(w) \leq 1 - \epsilon\} \text{ for some } \epsilon > 0.
\]

For kernel estimators, the results of Sections 7–10 below can be used to obtain conditions under which \(P(\tau \in T) \to 1\) when \(T\) is defined as in condition (iii) of (5.11). Note that condition (iii) requires smoothness in \(v\) of the derivative of the kernel regression estimator of \(Y_t\) on \(h(X_t, \theta_0)\) evaluated at \(h(X_t, \theta_0) = v\) where the derivative is with respect to \(v\) and also when the derivative is with respect to \(\theta\) and \(\theta\) is evaluated at \(\theta_0\).
Sufficiency of (i)–(iii) of (5.11) for Assumption 2(e) for Ichimura's and Ichimura and Lee's estimators is established as follows: We establish stochastic equicontinuity by verifying Assumption E separately for each element of the vector \( m(w, \theta_0, \tau) \). For the \( \ell \)-th element, take \( w = (y, v, d_{\ell}) \), where \( d_{\ell} \) denotes the \( \ell \)-th column of \( d = \frac{\partial}{\partial y} h(x, \theta_0) \). The dimension of \( w \) (i.e., \( k \) in the notation of Assumption E) is \( 1 + 2k_a \). For each \( \ell = 1, \ldots, p \), Assumption E(i) holds by (iii) with \( \Psi = y \times v \times \mathcal{P}_{\ell} \) and \( \Psi^* = y \times v \times \mathcal{P}_{\ell} \), where \( \mathcal{P}_{\ell} \) is an open bounded rectangle in \( \mathbb{R}^{k_a} \) that contains the support of \( \frac{\partial}{\partial y} h(X_t, \theta_0) \) \( \forall t \geq 1 \). \( \Psi^* \) has minimally smooth boundary because it is a finite union of convex sets. For each \( \ell = 1, \ldots, p \), Assumption E(ii) holds, since \( \| (y - \tau)/(\tau^{(1)} + d_{\ell} \tau^{(2)}) \|_{q, 2, \Psi^*} \leq C^* \) \( \forall \tau \in \mathcal{T} \) for some \( C^* < \infty \) by (iii), E(iii) holds by (iii), and E(iv) holds by (i). This completes the proof.

The sufficiency of (i)–(iv) of (5.11) and (5.12) for Assumption 2(e) for Klein and Spady's estimator is established in an analogous fashion.

6. STOCHASTIC EQUICONTINUITY VIA BRACKETING/SERIES EXPANSIONS

The results of this section apply to functions \( \{m(w, \tau)\} \) that are not necessarily smooth in \( w \) or \( \tau \) and to rv's \( \{W_{T_t}\} \) that are strong mixing. In comparison to the results of Sections 3 and 4, those of this section are more general with respect to temporal dependence of \( \{W_{T_t}\} \), but less general in other respects.

First, we present a bracketing result of Andrews and Pollard (1990). This result applies to bounded functions that are indexed by a finite dimensional parameter. It is suitable for application to M-estimators with non-differentiable \( \psi \)-functions (such as the LAD estimator) when the data are temporally dependent. The proof of this result uses a chaining argument that is driven by a moment inequality for the sums of mean zero strong mixing rv's rather than by the Bernstein inequality which is used in the bracketing results of Section 4. See Andrews and Pollard (1990) for details.
Next, this bracketing result is extended by combining it with the series result of Section 5. New stochastic equicontinuity results are obtained for classes of functions that are products of smooth functions from the infinite dimensional classes of Section 5 and non-smooth functions from the finite dimensional classes considered in this section. These results are applicable to weighted censored LAD estimators for censored regression models with strong mixing rv's (see Newey and Powell (1987) for the iid case) and to weighted M-estimators for regression models with strong mixing rv's (see Bates and White (1988)). The proof of the results exploits the simplicity of the series expansion argument outlined in Section 5.

A drawback of the results of this section is that they require the function $m(\cdot, \tau)$ to be bounded. In many applications, a consequence of this is that certain elements of $W_{Tt}$ must be bounded. In particular, in linear regression or simultaneous equations models, it often implies that the exogenous variables must be bounded.

We now introduce the bracketing result.

**ASSUMPTION G:** (i) $\mathcal{M}$ is a type IV class of uniformly bounded functions with index $p = 2$, constant $\psi > 0$, and dimension $d (< \infty)$ of $T$.
(ii) $\{W_{Tt}\}$ is a strong mixing triangular array with mixing numbers $\{\alpha(s)\}$ of size $-(2d + \psi)(d + 2\psi)/\psi^2$.

**THEOREM II.8:** Under Assumption G, $\{\nu_T(\cdot) : T \geq 1\}$ is stochastically equicontinuous with $\rho_T$ given by (4.1).

**COMMENT:** Theorem II.8 follows from Theorem 2.3 of Andrews and Pollard (1990).

**EXAMPLES:** The parametric examples (3) and (4) of Section 3 are not covered by the stochastic equicontinuity results of that section when the rv's $\{W_t\}$ are not $m$-dependent. Theorem II.8 can be used for these examples provided $\{W_t\}$ is strong mixing, the regressors $\{X_t : t \geq 1\}$ in (3) and the IV's $\{g(Z_t, \theta) : \theta \in \Theta, t \geq 1\}$ in (4) are uniformly bounded, and the underlying distribution of $\{W_t\}$ is restricted such that the
$L^p$ continuity condition of type IV classes is satisfied. Note that the latter is slightly less restrictive than the usual conditions imposed in these examples in the temporally independent case that rely on Huber's (1973) results. They are stronger, however, than the conditions used in Section 3 for the $m$-dependent case.

Next we consider functions $m(\cdot, \tau)$ that are given by products of smooth functions from an infinite dimensional class of the type described in Section 5 with bounded functions from a type IV class. Suppose $\tau$ can be partitioned as $\tau = (\tau_a, \tau_b)$. Let $\mathcal{I}_a = \{\tau_a : (\tau_a, \tau_b) \in \mathcal{I} \text{ for some } \tau_b\}$ and $\mathcal{I}_b = \{\tau_b : (\tau_a, \tau_b) \in \mathcal{I} \text{ for some } \tau_a\}$. The function $m(w, \tau)$ is taken to be of the form $m_a(w_a, \tau_a)m_b(w_b, \tau_b)$, where $w = (w_a', w_b')$ and $w_a \in \mathbb{R}^{k_a}$. Define $\mathcal{H}_a = \{w_a \in \mathbb{R}^{k_a} : \exists w_b \text{ s.t. } (w_a', w_b') \in \mathcal{N}\}$. The function $m_a(w_a, \tau_a)$ is taken to be smooth on a subset $\mathcal{H}_a^*$ of $\mathcal{H}_a$ and equal to a constant $K$ elsewhere. The functions $\{m_b(\cdot, \tau_b) : \tau_b \in \mathcal{I}_b\}$ are taken to form a type IV class of uniformly bounded functions.

Define $\rho_\mathcal{I}$ as follows: For $\tau_1 = (\tau_{1a}, \tau_{1b}) \in \mathcal{I}$ and $\tau_2 = (\tau_{2a}, \tau_{2b}) \in \mathcal{I}$,

$$\rho_\mathcal{I}_a(\tau_{1a}, \tau_{2a}) = \left( \int_{\mathcal{H}_a^*} (m_a(w_a, \tau_{1a}) - m_a(w_a, \tau_{2a}))^2 dw \right)^{1/2},$$

(6.1) $$\rho_\mathcal{I}_b(\tau_{1b}, \tau_{2b}) = \left[ E(m_b(W_{bTt}, \tau_{1b}) - m_b(W_{bTt}, \tau_{2b}))^2 \right]^{1/2},$$

and

$$\rho_\mathcal{I}(\tau_1, \tau_2) = \rho_\mathcal{I}_a(\tau_{1a}, \tau_{2a}) \vee \rho_\mathcal{I}_b(\tau_{1b}, \tau_{2b}),$$

where $\vee$ denotes the maximum operator.

ASSUMPTION H: (i) $m(W_{Tt}, \tau)$ is of the form $m_a(W_{aTt}, \tau_a)m_b(W_{bTt}, \tau_b)$, where $W_{Tt} = (W_{aTt}', W_{bTt}')$.

(ii) $\mathcal{H}_a^*$ is contained in $\mathcal{H}_a$ and is an open bounded subset of $\mathbb{R}^{k_a}$ with minimally smooth boundary (where $k_a \leq k$).

(iii) $\sup_{\tau_a \in \mathcal{I}_a} \|m_a(\cdot, \tau_a)\|_{q, \mathcal{H}_a^*} < \infty$ for some $q > k_a/2$. 
(iv) For some constant \( K \), \( m_a(w, \tau_a) = K \forall w, \tau_a \in \mathcal{W}_a - \mathcal{W}_a^* \), \( \forall \tau_a \in \mathcal{T}_a \).

(v) \( \{m_b(\cdot, \tau_b) : \tau_b \in \mathcal{T}_b\} \) is a type IV class of uniformly bounded functions with index \( p = 2 \), constant \( \psi > 0 \), and dimension \( d < \infty \) of \( \mathcal{T}_b \).

(v) \( \{W_{T_t}\} \) is a strong mixing triangular array with mixing numbers of size \(-(2d + \psi)(d + 2\psi)/\psi^2\).

THEOREM II.9: Under Assumption H, \( \{\nu_T(\cdot) : T \geq 1\} \) is stochastically equicontinuous with \( \rho_T \) given by (6.1).


EXAMPLE: WC–LAD estimator of the censored regression model. The WC–LAD example (2) of Section 3 is not covered by the stochastic equicontinuity results of that section when \( \{W_t\} \) is not \( m \)-dependent. It also is not covered by the series expansion results of Section 5, because \( m(w, \theta, \tau) \) is not a smooth function of \( \theta \) for all \( w \), due to the sign and indicator functions. This example is covered by Theorem II.9, however, under the conditions given below.

Let \( \mathcal{X} \subset \mathbb{R}^a \) denote the domain of the functions \( \tau(x) \). \( \mathcal{X} \) is defined to include the support of \( X_t \forall t \geq 1 \). Let \( \Theta \) be a subset of \( \mathbb{R}^p \) that contains a neighborhood of \( \theta_0 \).

The following conditions are sufficient for Assumption 2*(e) of ASEM:I with \( \rho_{\Theta \times I} \) given by \( \rho_I \) of (6.1) with \( I, I_a, I_b, m_a(w, \tau_a), \mathcal{X}_a^* \), and \( m_b(W_{bT_t}, \tau_b) \) of (6.1) replaced by \( \Theta \times I, I, \Theta, \tau(x), \mathcal{X}_* \), and \( 1(X_t^* \theta < C_t)\text{sgn}(Y_t - X_t^* \theta) \) respectively:
(i) \( \{W_t : t \geq 1\} \) is a strong mixing sequence of rv's with mixing numbers of size \(-2(4p+1)(p+1)\), where \( p \) is the dimension of \( \theta_0 \).

(ii) \( T = \{ \tau : \mathcal{X} \to \mathbb{R} \mid \|\tau(\cdot)\|_{q,2,x^*} \leq B \text{ and } \tau(x) = 0 \forall x \in \mathcal{X} - \mathcal{X}^* \} \) for some \( q > \frac{k_a}{2} \) and some constant \( B < \infty \), where \( \mathcal{X}^* \) is contained in \( \mathcal{X} \) and is a k\(^a\)-open bounded subset of \( \mathbb{R}^A \) with minimally smooth boundary.

(iii) \( \{g : g(w) = 1(x' \theta < c)\text{sgn}(y - x' \theta) \mid \text{for some } \theta \in \Theta, \text{ where } w = (y, x', c')\} \)

is a type IV class of functions with index \( p = 2 \) and constant \( \psi = 1/2 \).

Note that a more primitive set of conditions under which condition (iii) holds can be obtained from Powell (1984). (Powell verifies Huber's (1967) Assumption N–3(iii), which is equivalent to condition (iii) above for the given choices of \( p \) and \( \psi \).) Also note that the symbol \( p \) is given dual unconnected roles above. In condition (i), it denotes the dimension of \( \theta_0 \), as in ASEM:I. In condition (iii), it denotes the index of a type IV class, as in Section 4.

Sufficiency of conditions (i)–(iii) of (6.2) for Assumption 2*(e) is established as follows: By Theorem II.9, it suffices to verify Assumption H. Let \( m_a(W_{aT}, \tau_a) \) and \( m_b(W_{bT}, \tau_b) \) of Assumption H(i) equal \( \tau(X_t)X_t \text{ and } 1(X_t \theta < C_t)\text{sgn}(Y_t - X_t \theta) \). Let \( \mathcal{X}_a^* \) and \( \mathcal{X}_b^* \) of Assumption H(ii) equal \( \mathcal{X}^* \) and \( \mathcal{X} \). Assumptions H(ii), H(iii), H(iv), H(v), and H(vi) now hold by conditions (ii), (iii), (ii), (iii), and (i) of (6.2) respectively. The result follows.

Next, we consider the conditions \( P(\hat{\tau} \in \mathcal{T}) \to 1 \) and \( (\hat{\theta}, \hat{\tau}) \mathcal{B}(\theta_0, \tau_0) \) of Assumption 2*(b) of ASEM:I. Suppose \( \hat{\tau}(x) \) is a nonparametric estimator of \( \tau_0(x) \) for \( x \in \mathcal{X} \) which is trimmed to equal zero outside a set \( \mathcal{X}^* \), where \( \mathcal{X}^* \) is a finite union of open, bounded, convex subsets of \( \mathbb{R}^A \). Suppose \( \hat{\tau}(x) \) and its partial derivatives of order \( \leq q \), for some integer \( q > \frac{k_a}{2} \), are uniformly consistent over \( x \in \mathcal{X}^* \) for \( \tau_0(x) \) and its corresponding partial derivatives. Suppose the partial derivatives of \( \tau_0(x) \) of order \( \leq q \) are uniformly bounded over \( x \in \mathcal{X}^* \), say by \( B/2 \). Then, the realizations of \( \hat{\tau}(x) \) lie in the set \( \mathcal{T} \) of
condition (ii) of (6.2) with probability→1. In consequence, the above assumptions,  
\( \hat{\theta} \in \theta_0 \), and conditions (i) and (iii) of (6.2) imply that Assumption 2*(e) and the first two 
parts of Assumption 2*(b) of ASEM:I hold. If  \( \hat{\tau}(x) \) is a kernel estimator, then sufficient 
conditions for the above uniform consistency properties are given in Sections 7–10 below.

7. NONPARAMETRIC KERNEL DENSITY AND REGRESSION ESTIMATION

This section and Sections 8–10 present a number of consistency and rate of converg-
ence results for kernel estimators of density functions, regression functions, and their 
derivatives. The results allow for near epoch dependent non–identically distributed observ-
vations and data–dependent bandwidth parameters. The uniform consistency results that 
are given for kernel regression estimators establish uniform consistency over sets that 
expand with the sample size.  \( L^Q \)–consistency results are also given.

The kernel estimators that are considered here were first introduced by Rosenblatt 
(1956) for density estimation, by Nadaraya (1964) and Watson (1964) for regression esti-
mation, and by Bhattacharya (1967) for derivative estimation. We use higher order bias 
reducing kernels that were first considered by Bartlett (1963). Our uniform consistency 
proofs make use of a Fourier transformation of the kernel. This method is due to Parzen 
(1962) and has been exploited by numerous others including Bierens (1983, 1987) and 

The proof of our results is most similar to that of Bierens (1983) and Newey (1989).
The results given here, however, are more general than those of Bierens and Newey in 
several respects. In particular, the results allow for (i) non–identically distributed rv’s, 
(ii) near–epoch dependent rv’s with a strong mixing base, (iii) estimated as well as 
observed rv’s, (iv) estimation of derivatives as well as the functions themselves (as in 
Newey (1989)), (v) data–dependent bandwidth parameters, (vi) uniform consistency (with
a rate) on sets that expand with the sample size, and (vii) \( L^Q \)-consistency with a rate for \( 0 < Q < \infty \).

Here and in Sections 8–10, we consider a triangular array of rv's \( \{(Y_{Tt}, X_{Tt}) : t \leq T, T \geq 1\} \), where \( Y_{Tt} \in R \) and \( X_{Tt} \in R^k \). We analyze some properties of non-parametric kernel estimators of the densities of \( X_{Tt} \), of the regression functions of \( Y_{Tt} \) given \( X_{Tt} \), and of derivatives of these functions. (Triangular arrays are considered rather than sequences to allow for the application of the results given below in the local power context of ASEM:1.)

The following notation is used below: Let \( \mu \) and \( \lambda \) denote \( k \)-vectors of non-negative integer constants. For such vectors, define (i) \( |\mu| = \sum_{j=1}^{k} \mu_j \), where \( \mu = (\mu_1, \ldots, \mu_k) \), (ii) \( \mu \leq \lambda \) iff \( \mu_j \leq \lambda_j \ \forall j = 1, \ldots, k \), (iii) \( \mu < \lambda \) iff \( \mu \leq \lambda \) and \( \mu_j < \lambda_j \) for some \( j \), (iv) for any function \( c(x) \) on \( R^k \),

\[
D^\mu c(x) = \frac{\partial^{|\mu|}}{\partial x_1^{\mu_1} \cdots \partial x_k^{\mu_k}} c(x), \quad \text{where } x = (x_1, \ldots, x_k), \quad \text{and } (v) \quad x^\mu = \prod_{j=1}^{k} x_j^{\mu_j}.
\]

The estimands of interest here are of the form \( D^\lambda c(x) \) for some density or regression function \( c(x) \). Throughout this section (and in particular in the assumptions), the vector \( \lambda \) denotes the order of differentiation of the estimand of interest. For given \( \lambda \), the rv's are assumed to satisfy the following assumptions when specified:

ASSUMPTION NP1: For some \( \beta \geq 2 \),

(a) \( \sup_{t \leq T, T \geq 1} E|Y_{Tt}|^\beta < \infty \) and 

(b) \( \{(Y_{Tt}, X_{Tt}) : t \leq T, T \geq 1\} \) is NED on some strong mixing triangular array \( \{V_{Tt}\} \) with NED numbers \( \{\eta(s) : s \geq 1\} \) that satisfy \( \eta(s) = O(s^{-\eta}) \) for some \( \eta \in (0, \infty) \) and \( \{V_{Tt}\} \) has strong mixing numbers \( \{\alpha(s) : s \geq 1\} \) that satisfy \( \sum_{s=1}^{\infty} \alpha(s)^\beta/(\beta-2) < \infty \).

(When \( \eta = \infty \), \( \eta(s) \) must be zero for all \( s \) large. When \( \beta = 2 \), \( \alpha(s) \) must be zero for all \( s \) large.)
ASSUMPTION NP2: (a) The distribution of $X_{Tt}$ is absolutely continuous with respect to Lebesgue measure with density $f_{Tt}(x) \forall t \leq T, T \geq 1$.

(b) $f_{Tt}(x)$ is continuously differentiable to order $\omega \geq 1 + |\lambda|$ and

$$\sup_{t \leq T, T \geq 1} \sup_{x \in \mathbb{R}^k} \left| \frac{1}{T} \sum_{i=1}^{T} D^{\mu} f_{Tt}(x) \right| < \infty \ \forall \mu \text{ with } |\mu| \leq \omega.$$

ASSUMPTION NP3: (a) $g_{T}(x) = \mathbb{E}(Y_{Tt} | X_{Tt} = x)$ does not depend on $t \forall t \leq T$.

(b) $g_{T}(x)$ is continuously differentiable to order $\omega \geq 1 + |\lambda|$.

(c) $\sup_{t \leq T, T \geq 1} \sup_{x \in \mathbb{R}^k} \left| \frac{1}{T} \sum_{i=1}^{T} D^{\mu} [g_{T}(x)f_{Tt}(x)] \right| < \infty \ \forall \mu \text{ with } |\mu| \leq \omega.$

Assumptions NP1—NP3 allow for dnd rv's, but require the regressor variables $X_{Tt}$ to have densities with respect to Lebesgue measure. The latter can be circumvented to some extent, as discussed below. Note that the continuity of $f_{Tt}(x)$, which is imposed in Assumption NP2, does not allow $X_{Tt}$ to have density that is bounded away from zero on its support.

We consider the following kernel estimators of $\frac{1}{T} \sum_{i=1}^{T} f_{Tt}(x)$ and $g_{T}(x)$:

$$\hat{f}(x) = \frac{1}{T} \sum_{i=1}^{T} K \left[ \frac{x - X_{Tt}}{\hat{\sigma}_{T}} \right] / \hat{\sigma}_{T}^k \text{ and}$$

$$\hat{g}(x) = \left[ \frac{1}{T} \sum_{i=1}^{T} Y_{Tt} K \left[ \frac{x - X_{Tt}}{\hat{\sigma}_{T}} \right] / \hat{\sigma}_{T}^k \right] / \hat{f}(x),$$

where the kernel $K(\cdot)$ is a non-random real function on $\mathbb{R}^k$ and the bandwidth parameter $\hat{\sigma}_{T}$ is a positive constant or scalar rv. The corresponding kernel estimators of $\frac{1}{T} \sum_{i=1}^{T} D^{\lambda} f_{Tt}(x)$ and $D^{\lambda} g_{T}(x)$ are $D^{\lambda} \hat{f}(x)$ and $D^{\lambda} \hat{g}(x)$ respectively.

The kernel $K$ and bandwidth parameter $\hat{\sigma}_{T}$ are assumed to satisfy the following assumptions when specified:
ASSUMPTION NP4: (a) \( fK(x)dx = 1 \), \( f x^\mu K(x)dx = 0 \) \( \forall 1 \leq |\mu| \leq \omega - |\lambda| - 1 \), \( \int |x^\mu K(x)|dx < \infty \) \( \forall \mu \) with \( |\mu| = \omega - |\lambda| \), and \( D^\mu K(x) \to 0 \) as \( \|x\| \to \infty \) \( \forall \mu \) with \( \mu < \lambda \).

(b) \( D^\mu K(x) \) exists, is absolutely integrable, and has Fourier transform \( \Psi_\mu(r) = (2\pi)^k f \exp(i r^T x) D^\mu K(x)dx \) that satisfies \( f(1 + \|r\|) \sup_{b \geq 1} |\Psi_\mu(br)|dr < \infty \) \( \forall \mu \leq \lambda \), where \( i = \sqrt{-1} \).

ASSUMPTION NP5: The data-dependent bandwidth parameters \( \{\hat{\sigma}_T : T \geq 1\} \) satisfy \( C_1 \sigma_{1T} \leq \hat{\sigma}_T \leq C_2 \sigma_{2T} \) with probability \( \to 1 \) for some sequences of nonrandom bounded positive constants \( \{\sigma_{1T} : T \geq 1\} \) and \( \{\sigma_{2T} : T \geq 1\} \) and some positive finite constants \( C_1 \) and \( C_2 \).

The conditions imposed on the kernel in Assumption NP4(a) are the standard bias reducing conditions first considered by Bartlett (1963). The conditions in Assumption NP4(b) are smoothness conditions on \( K \). The condition on \( \Psi_\mu(r) \) is slightly stronger than the condition used by Bierens (1983, 1987) for the case where \( \mu = 0 \), because we allow for data-dependent bandwidth parameters. (If \( \hat{\sigma}_T \) is a deterministic sequence, then we only need \( f(1 + \|r\|) |\Psi_\mu(r)|dr < \infty \) in Assumption NP4. Also, if \( \{(Y_{T_t}, X_{T_t})\} \) is an independent or strong mixing triangular array, then the multiplicative \( (1 + \|r\|) \) can be deleted whether or not \( \hat{\sigma}_T \) is deterministic.)

Assumption NP4 is satisfied by the multivariate normal–based kernels considered by Bierens (1987):

\[
K(x) = (2\pi)^{-k/2} \det^{-1/2}(\Omega)^{J} \prod_{j=1}^{J} a_j b_j \exp \left[ -\frac{1}{2} x^T \Omega^{-1} x / b_j^2 \right],
\]

where \( J \geq (\omega - |\lambda|)/2 \), \( \Omega \) is a positive definite matrix, and \( \{(a_j, b_j) : j \leq J\} \) are constants that satisfy

\[
\prod_{j=1}^{J} a_j = 1 \quad \text{and} \quad \prod_{j=1}^{J} a_j b_j^{2\ell} = 0 \quad \text{for} \quad \ell = 1, \ldots, J-1.
\]
Assumption NP5 holds in many cases for common data-dependent methods of choosing bandwidth parameters including cross-validation (CV), generalized cross-validation (GCV), \( C_p \), and plug-in procedures. For example, for plug-in procedures, \( \hat{\sigma}_T \) is often of the form \( \hat{\gamma}_{1T}^{-1} \hat{\gamma}_{2T} \) for some positive rv's \( \hat{\gamma}_{1T} \) and \( \hat{\gamma}_{2T} \). In this case, Assumption NP5 holds if \( \hat{\gamma}_{1T} \) and \( \hat{\gamma}_{2T} \) are chosen to be bounded above and away from zero with probability \( \to 1 \). In particular, if \( \hat{\gamma}_{1T} \geq \gamma_{10} \) and \( (\log T)(\hat{\gamma}_{2T} - \gamma_{20}) \geq 0 \) for some \( 0 < \gamma_{10} < \infty \) and \( 0 < \gamma_{20} < \infty \), then Assumption NP5 holds with \( \sigma_{1T} = \sigma_{2T} = T^{-\gamma_{20}} \).

If \( \hat{\sigma}_T \) is determined using a CV, GCV, or \( C_p \) criterion, then it has been shown that \( \hat{\sigma}_T / \sigma^*_T \leq 1 \) for a deterministic sequence \( \sigma^*_T \) (that is optimal in a certain sense) in a variety of iid, inid, and dependent identically distributed contexts under suitable assumptions (e.g., see Li (1987), Andrews (1991), Györfi, Härdle, Sarda, and Vieu (1989, Ch. 6), and references contained therein). In such cases, Assumption NP5 holds with \( \sigma_{1T} = \sigma_{2T} = \sigma^*_T \), \( C_1 = 1-\epsilon \), and \( C_2 = 1+\epsilon \) for arbitrary \( \epsilon > 0 \). Note that \( \sigma^*_T \) is usually of the form \( \sigma^*_T = O(T^{-\psi}) \) and \( \sigma^*_{-1} = O(T^{\psi}) \) for some \( 0 < \psi < \infty \).

Now we state the main results of this section:

**THEOREM II.10:** (a) Under Assumptions NP1, NP2, NP4, and NP5,
\[
\sup_{x \in \mathbb{R}^k} \left| D^\lambda \hat{f}(x) - \frac{1}{T} \sum_{t=1}^T D^\lambda f_T(x) \right| = O_p\left( T^{-\eta/(2\eta+1)} \sigma_{1T}^{-k-1} \log^{-1/(2\eta+1)} + \sigma_{2T}^{\omega-1} \right).
\]

(b) Under Assumptions NP1–NP5, for any sequence of positive constants or rv's \( \{d_T : T \geq 1\} \) such that \( d_T \geq 0 \),
\[
\sup_{\{x : T^{-\sum_{t=1}^T f_T(x) \leq 2d_T}\}} \left| D^\lambda \hat{g}(x) - D^\lambda g_T(x) \right|
\leq O_p\left( T^{-\eta/(2\eta+1)} \sigma_{1T}^{-k-1} \log^{-1/(2\eta+1)} d_T^{-2-1} \right) + O_p\left( \sigma_{2T}^{\omega-1} d_T^{-2-1} \right).
provided the right-hand side is $o_p(1)$. The same result holds with $\frac{1}{T}\sum_{t=1}^{T} f_t(x)$ replaced by $\hat{f}(x)$ in the set above.

**COMMENTS:** 1. Suppose the sequence $\{\sigma_{1T} : T \geq 1\}$ and $\{\sigma_{2T} : T \geq 1\}$ of Assumption NP5 satisfy $\sigma_{1T} = \sigma_{2T}$, $\sigma_{1T} = O(T^{-\psi})$, and $\sigma_{1T}^{-1} = O(T^{\psi})$ for some $\psi > 0$, $\forall T \geq 1$. As discussed above, this is often the case. Then, by Theorem II.10(a), for any $\kappa \geq 0$,

$$T^\kappa \sup_{x \in \mathbb{R}^k} \left| D^\lambda \hat{f}(x) - \frac{1}{T} \sum_{t=1}^{T} D^\lambda f_t(x) \right| \mathbb{P} 0 \text{ provided }$$

$$\psi < [\eta/(2\eta+1) - \kappa]/[k + |\lambda| + 1/(2\eta+1)] \text{ and } \psi > \kappa/[\omega - |\lambda|].$$

Provided $\kappa < \eta/(2\eta+1)$, the two conditions on $\psi$ are compatible for $\omega$ sufficiently large. If the rv's $\{(Y_{Tt}, X_{Tt})\}$ independent or strong mixing, then $\eta = \alpha$ and the first condition on $\psi$ reduces to $\psi < [1/2 - \kappa]/[k + |\lambda|].$

2. Suppose $\{\sigma_{1T}\}$ and $\{\sigma_{2T}\}$ are as in Comment 1. In addition, suppose $d_T \in [L_1 T^{-\delta}, L_2 T^{-\delta}]$ with probability → 1 for some constants $\delta \geq 0$ and $0 < L_1 < L_2 < \omega$. Then, by Theorem II.10(b), for any $\kappa \geq 0$ and $\epsilon > 0$,

$$T^\kappa \sup_{\left\{x : \frac{1}{T} \sum_{t=1}^{T} f_t(x) \geq \epsilon T^{-\delta}\right\}} \left| D^\lambda \hat{g}(x) - D^\lambda g_T(x) \right| \mathbb{P} 0 \text{ provided }$$

$$\psi < [\eta/(2\eta+1) - \delta(2 + |\lambda|) - \kappa]/[k + |\lambda| + 1/(2\eta+1)] \text{ and }$$

$$\psi > [\kappa + \delta(2 + |\lambda|)]/[\omega - |\lambda|].$$

The same result holds with $\frac{1}{T} \sum_{t=1}^{T} f_t(x)$ replaced by $\hat{f}(x)$ in the set above. Provided $\kappa + \delta(2 + |\lambda|) < \eta/(2\eta+1)$, the two conditions on $\psi$ are compatible for $\omega$ sufficiently large.

If the rv's $\{(Y_{Tt}, X_{Tt})\}$ are independent or strong mixing, then $\eta = \alpha$. Furthermore, if one considers the supremum over the set $\left\{x : \frac{1}{T} \sum_{t=1}^{T} f_t(x) \geq \epsilon\right\}$ or $\{x : \hat{f}(x) \geq \epsilon\}$,
then $\delta = 0$. In these cases, the conditions on $\psi$ reduce to $\psi < [1/2 - \kappa]/[k + |\lambda|]$ and $\psi > \kappa/|\omega - |\lambda||$.

3. It should be noted that the rate of convergence obtained in Theorem II.10 is not sharp. In particular, the first term on the right-hand side in parts (a) and (b) of the Theorem (i.e., the bound on the variance of the estimator) is larger than necessary for the case of independent rv’s. Nevertheless, we use the proof of the Theorem that is given, because it is particularly amenable to handling NED rv’s. Also, the given results are good enough to yield consistency when the bandwidth parameters are chosen to yield Stone’s (1980, 1982) optimal rate of convergence.

4. Theorem II.10(b) does not apply directly when estimating a regression function $g_t(x)$ if some elements of the regressor vector $X_{Tt}$ are discrete. It can be applied indirectly, however, if the discrete regressors take on at most a finite number of values. Suppose $X_{Tt} = (X_{1Tt}', X_{2Tt}')'$, $X_{1Tt}$ takes values in a finite set $\{x_{11}, \ldots, x_{1J}\}$, and Assumptions NP1–NP3 hold with $(Y_{Tt}, X_{Tt})$ replaced by $(Y_{Tt}1(X_{1Tt} = x_{1j}), X_{2Tt})$ for $j = 1, \ldots, J$. Then, $g_t(x) = \sum_{j=1}^{J} g_j(x_2)1(x_1 = x_{1j})$, where $x = (x_1', x_2')'$ and $g_j(x_2) = E(Y_{Tt}1(X_{1Tt} = x_{1j})|X_{2Tt} = x_2)$. Let $\hat{g}_j(x_2)$ be the kernel estimator of $g_j(x_2)$ using the dependent variable $Y_{Tt}1(X_{1Tt} = x_{1j})$ and the regressor vector $X_{2Tt}$ for $j = 1, \ldots, J$. Then, take $D^\lambda \hat{g}(x) = \sum_{j=1}^{J} D^\lambda \hat{g}_j(x_2)1(x_1 = x_{1j})$ to be an estimator of $D^\lambda g_t(x)$, where the elements of $\lambda$ that correspond to $x_1$ are zero. The uniform consistency of the kernel estimator $D^\lambda \hat{g}(x)$ can be obtained from Theorem II.10(b) by applying this Theorem to $D^\lambda \hat{g}_j(x)$ for $j = 1, \ldots, J$. (Note that if $g_j(x_2)$ is known to depend on only a subset of elements of $x_2$ for some $j$, then only these elements should be used as regressors when calculating $\hat{g}_j(x_2)$.)

For kernel regression estimators, Theorem II.10 does not deliver uniform consistency over unbounded sets. This is to be expected, because for values of $x$ sufficiently far in the
tail of the density of $X_{T_t}$ one observes relatively few outcomes that are useful in estimating $D^\lambda \hat{g}(x)$. On the other hand, it is possible to obtain $L^Q$-consistency of nonparametric estimators over unbounded sets. In particular, the uniform consistency results of Theorem II.10, which yield uniform consistency over expanding sets, can be used to obtain $L^Q$-consistency of trimmed versions of kernel regression estimators. Trimming is used to eliminate aberrant behavior of kernel regression estimators in regions of $\mathbb{R}^k$ where the densities of $\{X_{T_t}\}$ are small.

We consider a trimming function $\text{tr}_T(\cdot, \cdot, \cdot, \cdot)$ that satisfies

$$
\text{tr}_T(\hat{g}, \hat{G}_T, \lambda, x) \begin{cases} 
D^\lambda \hat{g}(x) & \text{for } x \in \hat{G}_T \\
\leq B & \text{elsewhere}
\end{cases}
$$

for some $B < \infty$, where $\hat{G}_T$ is a random subset of $\mathbb{R}^k$.

The trimming sets $\{\hat{G}_T : T \geq 1\}$ are assumed to satisfy:

**ASSUMPTION NP6:** For some sequences of constants or scalar r.v's $\{d_T : T \geq 1\}$ and $\{d_{2T} : T \geq 1\}$ such that $d_T \leq d_{2T}$ $\forall T \geq 1$ and $d_T = O_p(1)$,

(a) $\hat{G}_T \subseteq \{1^{\Sigma_{T_t}}_t \geq d_T\}$ or $\hat{G}_T \subseteq \{\hat{f}(x) \geq d_T\}$ with probability $\to 1$ and

(b) $\hat{G}_T \supset \{1^{\Sigma_{T_t}}_t \geq d_{2T}\}$ or $\hat{G}_T \supset \{\hat{f}(x) \geq d_{2T}\}$ with probability $\to 1$.

The simplest choice of trimming sets $\hat{G}_T$ and trimming functions $\text{tr}_T(\cdot, \cdot, \cdot, \cdot)$ are

$$
\hat{G}_T = \{\hat{f}(x) \geq d_T\} \quad \text{and} \quad \text{tr}_T(\hat{g}, \hat{G}_T, \lambda, x) = 1(x \in \hat{G}_T)D^\lambda \hat{g}(x),
$$

where $\{d_T : T \geq 1\}$ satisfies $d_T = O_p(1)$. In this case, Assumption NP6 holds with $d_{2T} = d_T$ and the $L^Q$-consistency results of Corollary II.1 below hold.

The above choices, however, are not suitable for use in the examples of ASEM.\(1\). The reason is that Assumption 2(b) and 2*(b) require $P(\tau \in T) \to 1$, where $T$ is chosen such that the stochastic equicontinuity results of Sections 3–5 above can be applied. In particular, for the case of infinite dimensional classes of functions, the stochastic equicontinuity results above require the functions in $T$ to be either smooth everywhere on
their domain (type V classes) or smooth on a single fixed set or on a set taken from a countable collection of sets and constant elsewhere (type III and type VI classes and Assumptions E, F, and H). The set of all possible realizations of 

\[ 1(x \in \{ \hat{f}(x) \geq d_T \})D^\lambda \hat{g}(x) \] does not fit into either of these categories.

Two alternative forms of trimming can be considered that are compatible with the stochastic equicontinuity results above. The first requires \( \hat{G}_T \) to take values in a countable class of sets \( \{ G_j : j \geq 1 \} \), i.e., \( \hat{G}_T = G_j^{\hat{G}_j} \) for some \( \hat{G}_j \), where the class of sets \( \{ G_j \} \) is as in the definition of type VI classes of functions above. One then takes

\[ \text{tr}_T(\hat{g}, G_j^{\hat{G}_j}, \lambda, x) = 1(x \in G_j^{\hat{G}_j})D^\lambda \hat{g}(x). \]

With this form of trimming, one has to ensure that \( \hat{G}_T \) is defined such that Assumption NP6 holds.

A second form of trimming that is compatible with the stochastic equicontinuity results above takes

\[ \hat{G}_T = \{ x \in \mathbb{R}^k : \| x - y \| \geq \delta \forall y \in \{ \hat{f}(y) \leq d_T \} \} \] and

\[ \text{tr}_T(\hat{g}, \hat{G}_T, \lambda, x) = D^\lambda \int_{\hat{G}_T} \hat{g}(z)K\left[ \frac{x-z}{\delta} \right]dz/\delta^k \]

for some \( \delta > 0 \) and some \( \{ d_T : T \geq 1 \} \) such that \( d_T = o_p(1) \), where \( D^\lambda \) denotes partial differentiation with respect to \( x \). Here \( \hat{K}(x) \) is a bounded, smooth, real-valued kernel that integrates to one and is zero for \( \| x \| > 1 \). In this case, \( \text{tr}_T(\hat{g}, \hat{G}_T, \lambda, x) \) inherits the smoothness of \( \hat{K}(x) \), and hence, the set of all of its realizations forms a type V class of functions under suitable smoothness conditions on \( \hat{K}(\cdot) \) and moment conditions on \( \{ X_{Tt} : t \geq 1 \} \). Assumption NP6(a) automatically holds in this case. Assumption NP6(b) holds for some sequence \( \{ d_{2T} : T \geq 1 \} \) for which \( d_{2T} = o_p(1) \) provided

\[ \inf_{x \in C} \lim_{T \to \infty} \sup_{T \to T_n} \| f_T(x) \|_1 \geq \epsilon_C \] for some \( \epsilon_C > 0 \) for each bounded set \( C \subset \mathbb{R}^k \).
In addition to the trimming assumption introduced above, the $L^Q$-consistency results given below rely on a restriction on the tail thickness of the densities of $\{X_{Tt}\}$ and the boundedness of the estimand $D^\lambda g(x)$.

**ASSUMPTION NP7:** (a) For some $0 < a \leq 1$,

$$\lim_{T \to \infty} \sup_{s \leq N, N \geq 1} \int \left( \frac{1}{T^{a/2}} f_{Tt}(x) \right)^{-a} f_N(x) dx < \infty.$$  

(b) $\sup_{x \in X} |D^\lambda g(x)| < \infty$, where $X$ denotes the union of the supports of $X_{Tt}$ $\forall t \leq T, T \geq 1$.

$L^Q$-consistency is established in the following corollary to Theorem II.10:

**COROLLARY II.1:** Under Assumptions NP1–NP7, for any $0 < Q < \infty$,

$$\sup_{t \leq N, N \geq 1} \left( \left[ \left| \text{tr}_T(\hat{g}, \hat{G}_T, \lambda, x) - D^\lambda g_T(x) \right|^{Q} \right]^{1/Q} \right)^{1/Q}$$

$$= O_p(\sigma_1 T^{-\eta/(2\eta+1)} - 1/(2n+1) d_T^{-2-\lambda}) + O_p(\sigma_2 T^{-2-\lambda}) + O_p(d_T^{a}/Q)$$

provided the first two terms on the right-hand side are $O_p(1)$.

**COMMENTS:** 1. Using the Corollary one can see that the conditions given in Comment 2 to Theorem II.10 on $\{\sigma_{1T}\}, \{\sigma_{2T}\}, \{d_T\}$, and $\psi$ for the case of $\kappa = 0$ are sufficient for $L^Q$-consistency of $\text{tr}_T(\hat{g}, \hat{G}_T, \lambda, x)$ for all $0 < Q < \infty$ provided $d_T = o_p(1)$.

2. For some of the examples of ASEM-I, one needs to obtain $L^Q$-consistency of a nonparametric regression estimator with a rate $T^{1/4}$. Corollary II.1 yields such a result if $d_T = o_p(T^{-Q/(4a)})$ and $\{\sigma_{1T}\}, \{\sigma_{2T}\}, \{d_T\}$ satisfy appropriate conditions, such as those given in Comment 2 to Theorem II.10. The above condition on $\{d_T\}$ is quite restrictive. If $Q$ is too large, it is not compatible with the conditions on $\{d_T\}$ and the requirement that $d_T \leq d_T$. In such cases, to apply the results of this paper one needs to estimate the nonparametric regression function with all of the data, but restrict the MINPIN estimator to be based only on $X_{Tt}$ observations that lie in a set over which uniform consistency at rate $T^{1/4}$ can be obtained.
8. KERNEL ESTIMATION WITH ESTIMATED RANDOM VARIABLES

In this section, we consider nonparametric kernel density and regression estimation when the rv's $Y_{Tt}$ and $X_{Tt}$ are not observed, but can be estimated using a finite dimensional parameter estimator. Consistency results for this case are particularly useful for applications in ASEM-I, since nonparametric estimators based on residuals or other estimated rv's often arise in such applications.

Let $\alpha_1$ and $\alpha_2$ be finite dimensional Euclidean-valued parameters that index rv's $Y_{Tt}(\alpha_1)$ and $X_{Tt}(\alpha_2)$. Suppose the rv's $Y_{Tt}$ and $X_{Tt}$ of Section 7 satisfy $Y_{Tt} = Y_{Tt}(\alpha_{10})$ and $X_{Tt} = X_{Tt}(\alpha_{20})$ for some $\alpha_{10}$ and $\alpha_{20}$. Let $\hat{\alpha}_1$ and $\hat{\alpha}_2$ be estimators of $\alpha_{10}$ and $\alpha_{20}$ respectively. Define $\hat{Y}_{Tt} = Y_{Tt}(\hat{\alpha}_1)$ and $\hat{X}_{Tt} = X_{Tt}(\hat{\alpha}_2)$.

When \{\{Y_{Tt} : t \leq T\} and/or \{X_{Tt} : t \leq T\} is not observed, one can form kernel density and regression function estimators using \{\hat{Y}_{Tt} : t \leq T\} and/or \{\hat{X}_{Tt} : t \leq T\}.

In particular, for the case where \{(\hat{Y}_{Tt}, X_{Tt}) : t \leq T\} is observed, let

\begin{equation}
(8.1) \quad g^*(x) = \left[ \frac{1}{T} \sum_{t=1}^{T} \hat{Y}_{Tt} K \left( \frac{x - X_{Tt}}{\hat{\sigma}_T} \right) \right] / \hat{f}(x)
\end{equation}

be an estimator of $g_T(x)$. For the case where only \{(\hat{Y}_{Tt}, \hat{X}_{Tt}) : t \leq T\} is observed, let

\begin{equation}
(8.2) \quad \hat{f}(x) = \frac{1}{T} \sum_{t=1}^{T} K \left( \frac{x - \hat{X}_{Tt}}{\hat{\sigma}_T} \right) / \hat{f}(x)
\end{equation}

be estimators of $\frac{1}{T} \sum_{t=1}^{T} f_{Tt}(x)$ and $g_T(x)$ respectively. Correspondingly, $D^\lambda g^*(x)$, $D^\lambda \hat{f}(x)$, and $D^\lambda \hat{g}(x)$ are estimators of $D^\lambda g_T(x)$, $\frac{1}{T} \sum_{t=1}^{T} D^\lambda f_{Tt}(x)$, and $D^\lambda g_T(x)$.

The following assumptions are used to establish uniform consistency properties of $D^\lambda g^*(x)$, $D^\lambda \hat{f}(x)$, and $D^\lambda \hat{g}(x)$:
ASSUMPTION NP8: (a) $\sqrt{T}(\hat{a}_1 - a_{10}) = O_p(1)$.

(b) $Y_{Tt}(\alpha_1)$ is differentiable in $\alpha_1$ on a neighborhood $A_1$ of $a_{10}$ a.s. and
\[ \sup_{T \geq 1} \frac{1}{T} \sum_{t=1}^{T} \sup_{\alpha_1 \in A_1} \left\| \frac{\partial}{\partial \alpha_1} Y_{Tt}(\alpha_1) \right\| < \omega. \]

(c) $\sup_{\mu} |D^\mu K(x)| < \omega \forall \mu \leq \lambda$, $x \in \mathbb{R}^k$.

ASSUMPTION NP9: (a) $\sqrt{T}(\hat{a}_2 - a_{20}) = O_p(1)$.

(b) $X_{Tt}(\alpha_2)$ is differentiable in $\alpha_2$ on a neighborhood $A_2$ of $a_{20}$ a.s. and
\[ \sup_{T \geq 1} \frac{1}{T} \sum_{t=1}^{T} \sup_{\alpha_2 \in A_2} \left\| \frac{\partial}{\partial \alpha_2} X_{Tt}(\alpha_2) \right\| < \omega. \]

(c) $\sup_{\mu} |D^\mu K(x)| < \omega \forall \mu \leq \lambda + e_j$, $\forall j = 1, \ldots, k$, where $e_j$ is the $j$-th elementary $k$-vector.

ASSUMPTION NP10: $\sup_{T \geq 1} \frac{1}{T} \sum_{t=1}^{T} \sup_{\alpha_1 \in A_1, \alpha_2 \in A_2} \left\| Y_{Tt}(\alpha_1) \frac{\partial}{\partial \alpha_2} X_{Tt}(\alpha_2) \right\| < \omega$ for $A_1$ and $A_2$ as in Assumptions NP8 and NP9.

Using these assumptions and the results of Section 7, we obtain:

THEOREM II.11: (a) Under Assumptions NP1–NP5 and NP8, part (b) of Theorem II.10 holds with $g^*(x)$ in place of $\hat{g}(x)$.

(b) Under Assumptions NP1–NP8, the result of Corollary II.1 holds with $g^*(x)$ in place of $\hat{g}(x)$.

(c) Under Assumptions NP1, NP2, NP4, NP5, NP8, and NP9,
\[ \sup_{x \in \mathbb{R}^k} |D^\lambda \tilde{f}(x) - \frac{1}{T} \sum_{t=1}^{T} D^\lambda f_{Tt}(x)| = O_p(T^{-\eta/(2\eta+1)} \sigma_{1T}^{-k-1} |\lambda|^{-1}) + O_p(\sigma_{2T}^\omega |\lambda|). \]

(d) Under Assumptions NP1–NP5 and NP8–NP10, for any sequence of positive constants or scalar rv's $\{d_T : T \geq 1\}$ such that $d_T = o_p(1)$,
\[\sup_{x : 1/2 < \beta_t \leq d_T} |D^\lambda \hat{g}(x) - D^\lambda g_T(x)|\]

\[= O_p(T^{-\eta/(2\eta+1)} \sigma_1^{-\lambda-1} d_T^{-2-|\lambda|} + O_p(\sigma_2^{-\lambda} d_T^{-2-|\lambda|} + d_2^{\theta/2})\]

provided the right-hand side is \(o_p(1)\). The same result holds with \(1/2 < \beta_t \leq d_T\) replaced by \(\bar{f}(x)\) in the set above.

(e) Under Assumptions NP1–NP10, for any \(0 < Q < \omega\),

\[\sup_{x \in \mathcal{N}, N \geq 1} \left[ \int |\text{tr}_T(\hat{g}, \hat{G}_T, \lambda, x) - D^\lambda g_T(x)| Q dP_N(x) \right]^{1/Q} \]

\[= O_p(T^{-\eta/(2\eta+1)} \sigma_1^{-\lambda-1} d_T^{-2-|\lambda|} + O_p(\sigma_2^{-\lambda} d_T^{-2-|\lambda|} + d_2^{\theta/2})\]

provided the first two terms on the right-hand side are \(o_p(1)\).

COMMENTS: 1. The conditions given for consistency and the corresponding rate of consistency results of Theorems II.10 and II.11 are the same for \(D^\lambda g(x)\) and \(D^\lambda \hat{g}(x)\). For \(D^\lambda \bar{f}(x)\) and \(D^\lambda \hat{f}(x)\), however, the results differ slightly. The results for \(D^\lambda \bar{f}(x)\) are weaker, because \(\sigma_1^{-1}\) appears on the right-hand side rather than \(\sigma_1^{-1/(2\eta+1)}\). The same difference arises in the results for \(D^\lambda \hat{g}(x)\) and \(D^\lambda \hat{g}(x)\).

2. Comments 1 and 2 following Theorem II.10 also hold with \(\bar{f}(x)\) and \(\hat{g}(x)\) replaced by \(\bar{f}(x)\) and \(g^*(x)\) or \(\hat{g}(x)\) under the assumptions of Theorem II.11. The only difference is that \(1/(2\eta+1)\) (but not \(\eta/(2\eta+1)\)) is replaced by 1 everywhere it appears when \(\bar{f}(x)\) or \(\hat{g}(x)\) is being considered.

9. UNIFORM CONSISTENCY OF FAMILIES OF KERNEL ESTIMATORS

In this section we consider nonparametric regressions of a family of dependent variables \(\{Y_{T_t}(\alpha_1) : \alpha_1 \in A_1\}\) on a family of regressor vectors \(\{X_{T_t}(\alpha_2) : \alpha_2 \in A_2\}\). We also consider density estimation for the rv's \(\{X_{T_t}(\alpha_2) : \alpha_2 \in A_2\}\). We derive consistency results that hold uniformly over \(\alpha_1 \in A_1\) and \(\alpha_2 \in A_2\) for kernel estimators of these regression functions, density functions, and their derivatives. Such results are useful in the
MAD–DUC and three–step sample selection estimator examples of ASEM:I. For example, for MAD–DUC estimators, one needs to estimate \( E(Y_t | h(X_t, \theta) = v) \) and some of its derivatives uniformly well over \( \theta \in \Theta \) in order to verify the first part of Assumption 1(b) of ASEM:I. In this example, \( X_{Tt}(\alpha_1) \) of the present section corresponds to the index function \( h(X_t, \theta) \).

The parameters \( \alpha_1 \) and \( \alpha_2 \), referred to above, may be infinite dimensional in this section. In consequence, the results given here can also be used to obtain consistency of kernel estimators when \( Y_{Tt} \) and/or \( X_{Tt} \) are not observed but estimated analogues of them are observed that depend on infinite dimensional preliminary estimators. The results of Section 8 do not apply in this case because they require \( \alpha_1 \) and \( \alpha_2 \) to be finite dimensional. For example, in the three–step sample selection estimator example of ASEM:I, the nonparametric estimator of \( \tau_{30}(\alpha, v) \) could involve infinite dimensional preliminary estimators (although it is possible to avoid such estimators).

The results of this section make use of stochastic equicontinuity conditions. Hence, they provide another context in which the stochastic equicontinuity results of Sections 3–6 can be exploited.

Let \( A_1 \) and \( A_2 \) be pseudo–metric spaces with pseudo–metrics \( \rho_1 \) and \( \rho_2 \) respectively. Let \( \Sigma \) be a bounded subset of \([0, \omega]\). Let \( \mathcal{X}_B \) be a bounded subset of \( \mathbb{R}^k \), where \( k \) is the dimension of \( X_{Tt}(\alpha_2) \). For each \( \mu \leq \lambda \) (where \( \lambda \) is as in Section 7), we define two sequences of stochastic processes \( \{ \xi_{1T}(\mu, \cdot, \cdot, \cdot) : T \geq 1 \} \) and \( \{ \xi_{2T}(\mu, \cdot, \cdot, \cdot) : T \geq 1 \} \), indexed by \( (\alpha_2, \sigma, x) \in A_2 \times \Sigma \times \mathcal{X}_B \) and \( (\alpha_1, \alpha_2, \sigma, x) \in A_1 \times A_2 \times \Sigma \times \mathcal{X}_B \) respectively. By definition,

\[
\xi_{1T}(\mu, \alpha_2, \sigma, x) = \frac{1}{\sqrt{T}} \sum_1^T \left[ \mathcal{D}^{\mu_K} \left( \frac{x-X_{Tt}(\alpha_2)}{\sigma} \right) - ED^{\mu_K} \left( \frac{x-X_{Tt}(\alpha_2)}{\sigma} \right) \right]
\]

and

\[
\xi_{2T}(\mu, \alpha_1, \alpha_2, \sigma, x) = \frac{1}{\sqrt{T}} \sum_1^T \left[ Y_{Tt}(\alpha_1) \mathcal{D}^{\mu_K} \left( \frac{x-X_{Tt}(\alpha_2)}{\sigma} \right) \right]
\]

\[
- EY_{Tt}(\alpha_1) \mathcal{D}^{\mu_K} \left[ \frac{x-X_{Tt}(\alpha_2)}{\sigma} \right],
\]

for \( \mu \leq \lambda \) and \( \lambda \) as in Section 7.
where $K(\cdot)$ is a kernel as in Section 7. Let $\tilde{\rho}$ and $\rho^*$ denote pseudo-metrics on $A_2 \times \Sigma \times \mathcal{X}_B$ and $A_1 \times A_2 \times \Sigma \times \mathcal{X}_B$ respectively. If $A_1$ and $A_2$ are subsets of Euclidean space, as is often the case in examples, then $\tilde{\rho}$ and $\rho^*$ can be taken to be Euclidean metrics.

Below we use the differential operator $D^\mu$ applied to functions of two or three arguments. The arguments include $\alpha_1$ and/or $\alpha_2$ plus an additional variable, such as $x$. In each case, $\alpha_1$ and $\alpha_2$ are fixed and the differential operator applies to the additional variable.

Of the rv's $\{Y_{Tt}(\alpha_1), X_{Tt}(\alpha_2)\}$ and the kernel $K(\cdot)$, we assume:

ASSUMPTION NP1*: For each $\mu \leq \lambda$, $\{\xi_{1T}(\mu, \cdot, \cdot, \cdot) : T \geq 1\}$ is stochastically equicontinuous on $A_2 \times \Sigma \times \mathcal{X}_B$ and $A_2 \times \Sigma \times \mathcal{X}_B$ is totally bounded under the metric $\tilde{\rho}$.

ASSUMPTION NP2*: (a) The distribution of $X_{Tt}(\alpha_2)$ is absolutely continuous with respect to Lebesgue measure with density $f_{Tt}(\alpha_2, x)$, $\forall \alpha_2 \in A_2$, $t \leq T$, $T \geq 1$.

(b) $f_{Tt}(\alpha_2, x)$ is continuously differentiable in $x$ to order $\omega \geq 1 + |\lambda|$ for $\forall \alpha_2 \in A_2$, $t \leq T$, $T \geq 1$, and

$$\sup_{t \leq T, T \geq 1} \sup_{\alpha_2 \in A_2} \sup_{x \in \mathbb{R}^k} \left| \frac{1}{T^{1/2}} D^\mu f_{Tt}(\alpha_2, x) \right| < \infty \forall \mu \text{ with } |\mu| \leq \omega.$$

ASSUMPTION NP3*: (a) $E|Y_{Tt}(\alpha_1)| < \infty \forall \alpha_1 \in A_1$, $t \leq T$, $T \geq 1$.

(b) $g_T(\alpha_1, \alpha_2, x) = E(Y_{Tt}(\alpha_1)|X_{Tt}(\alpha_2) = x)$ does not depend on $t \forall \alpha_1 \in A_1$, $\alpha_2 \in A_2$, $t \leq T$, $T \geq 1$.

(c) $g_T(\alpha_1, \alpha_2, x)$ is continuously differentiable in $x$ to order $\omega \geq 1 + |\lambda|$ for $\forall \alpha_1 \in A_1$, $\alpha_2 \in A_2$, $T \geq 1$.

(d) $\sup_{t \leq T, T \geq 1} \sup_{\alpha_1 \in A_1} \sup_{\alpha_2 \in A_2} \sup_{x \in \mathbb{R}^k} \left| \frac{1}{T^{1/2}} D^\mu [g_T(\alpha_1, \alpha_2, x) f_{Tt}(\alpha_2, x)] \right| < \infty \forall \mu \text{ with } |\mu| \leq \omega$.

(e) For each $\mu \leq \lambda$, $\{\xi_{2T}(\mu, \cdot, \cdot, \cdot, \cdot) : T \geq 1\}$ is stochastically equicontinuous on $A_1 \times A_2 \times \Sigma \times \mathcal{X}_B$ and $A_1 \times A_2 \times \Sigma \times \mathcal{X}_B$ is totally bounded under the pseudo-metric $\rho^*$. 
ASSUMPTION NP4*: (a) \( D^\mu K(x) \) exists \( \forall \mu \leq \lambda \).

(b) Assumption NP4(a) holds.

We consider the following kernel estimators of \( \frac{1}{T} \Sigma_1^{T} f_{T_t}(\alpha_2, x) \) and \( g_{T}(\alpha_1, \alpha_2, x) \) for \( \alpha_1 \in A_1 \) and \( \alpha_2 \in A_2 \):

\[
\hat{f}(\alpha_2, x) = \frac{1}{T} \Sigma_1^{T} K \left[ \frac{x - X_{T_t}(\alpha_2)}{\hat{\sigma}_T} \right] / \hat{\sigma}_T \quad \text{and} \quad (9.2)
\]

\[
\hat{g}(\alpha_1, \alpha_2, x) = \left[ \frac{1}{T} \Sigma_1^{T} Y_{T_t}(\alpha_2) \right] / \hat{\sigma}_T / \hat{\sigma}_T \]

The corresponding kernel estimators of \( \frac{1}{T} \Sigma_1^{T} D^\lambda f_{T_t}(\alpha_2, x) \) and \( D^\lambda g_{T}(\alpha_1, \alpha_2, x) \) are \( D^\lambda \hat{f}(\alpha_2, x) \) and \( D^\lambda \hat{g}(\alpha_1, \alpha_2, x) \) respectively. (As above, \( D^\lambda \) denotes the \( \lambda \)-th derivative with respect to \( x \), not \( \alpha_1 \) or \( \alpha_2 \).)

The bandwidth parameter \( \hat{\sigma}_T \) is assumed to satisfy:

ASSUMPTION NP5*: (a) Assumption NP5 holds.

(b) \( C_2 \sigma_2T \) of Assumption NP5 is in \( \Sigma \) for all \( T \) large.

Note that Assumption NP5*(b) is satisfied for any reasonable choices of \( \hat{\sigma}_T \) and \( \Sigma \).

The consistency results of this section are stated as follows:

THEOREM II.12: (a) Under Assumptions NP1*, NP2*, NP4*, and NP5*,

\[
\sup_{\alpha_2 \in A_2, x \in \mathcal{X}_B} \left| D^\lambda \hat{f}(\alpha_2, x) - \frac{1}{T} \Sigma_1^{T} D^\lambda f_{T_t}(\alpha_2, x) \right| = O_p(T^{-1/2} \sigma_1T^{-k-1} \lambda) + O_p(\sigma_2T^w |\lambda|).
\]

(b) Under Assumptions NP1*–NP5*, for any constant \( d > 0 \),

\[
\sup_{\alpha_1 \in A_1, \alpha_2 \in A_2, x \in \mathcal{X}_B} \left| D^\lambda \hat{g}(\alpha_1, \alpha_2, x) - D^\lambda g_{T}(\alpha_1, \alpha_2, x) \right|
\]

\[
= O_p(T^{-1/2} \sigma_1T^{-k-1} \lambda) + O_p(\sigma_2T^w |-\lambda|).
\]
provided the right-hand side is $o_p(1)$. The same result holds with $\frac{1}{T} \sum T_t \tau_t (\alpha_2, x)$ replaced by $\hat{f}(\alpha_2, x)$ in the set above.

COMMENTS: 1. In contrast to Theorem II.10, Theorem II.12 does not allow for uniform consistency over expanding sets of $x$ values.

2. The rates of convergence provided by Theorem II.12 are identical to those of Theorem II.10 when the latter has $\eta = \infty$ and $d_T = d \forall T \geq 1$.

10. KERNEL ESTIMATION OF DERIVATIVES OF REGRESSIONS ON INDEX FUNCTIONS

In some of the examples of ASEM-I, one needs to estimate derivatives of regressions on index functions, where the derivatives are with respect to the index parameters not the values of the regression function. For example, this nonparametric estimation problem arises with the MAD–DUC estimators of Section 6.3 of ASEM-I, which include Klein and Spady’s (1987), Ichimura’s (1985), and Ichimura and Lee’s (1990) semiparametric estimators. In this section, we briefly show how the results of Sections 7–9 above can be used to establish consistency of kernel estimators of such estimands.

For each $T \geq 1$, suppose $(Y_{T_t}, X_{T_t})$ is identically distributed for $t \leq T$. Let $h(X_{T_t}, \alpha)$ denote the real-valued index function of interest, where $\alpha \in \mathbb{R}^d$ for some $d < \infty$. We assume $h(\cdot, \cdot)$ is known, $\{(Y_{T_t}, X_{T_t}) : t \leq T\}$ is observed, and $\alpha$ is a specified parameter value of interest. (The case of estimated $\alpha$ is also discussed below.) Let

\begin{equation}
(10.1) 
g_{1T}(\alpha, v) = E(Y_{T_t} | h(X_{T_t}, \alpha) = v).
\end{equation}

Suppose $h(X_{T_t}, \tilde{\alpha})$ has a density $f_T(\tilde{\alpha}, v)$ with respect to Lebesgue measure for all $\tilde{\alpha}$ in a neighborhood of the parameter value $\alpha$ of interest.

The estimand of interest is
\[ (10.2) \quad \frac{\partial}{\partial \alpha} \hat{g}_1 T(\alpha, v) \text{ for some } v \in \mathbb{R}. \]

We consider estimating \( \frac{\partial}{\partial \alpha} \hat{g}_1 T(\alpha, v) \) using the derivative \( \frac{\partial}{\partial \alpha} \hat{g}_1 (\alpha, v) \) of the kernel regression estimator \( \hat{g}_1 (\alpha, v) \) of \( Y_T \) on \( h(X_T, \alpha) \). \( \hat{g}_1 (\alpha, v) \) is defined by

\[ \hat{g}_1 (\alpha, v) = \hat{A}_1 (\alpha, v) / \hat{f}(\alpha, v), \text{ where} \]

\[ (10.3) \quad \hat{A}_1 (\alpha, v) = \frac{1}{T \Sigma_1} Y_T K \left[ \frac{v - h(X_{Tt}, \alpha)}{\hat{\sigma}_T} \right] / \hat{\sigma}_T \text{ and} \]

\[ \hat{f}(\alpha, v) = \frac{1}{T \Sigma_1} K \left[ \frac{v - h(X_{Tt}, \alpha)}{\hat{\sigma}_T} \right] / \hat{\sigma}_T. \]

We now show how the results of Sections 7–9 can be used to establish different types of consistency of \( \frac{\partial}{\partial \alpha} \hat{g}_1 (\alpha, v) \) for \( \frac{\partial}{\partial \alpha} \hat{g}_1 T(\alpha, v) \). We use the following notation. Let \( g_2 T(\alpha, v) \) and \( g_3 T(\alpha, v) \) be defined analogously to \( g_1 T(\alpha, v) \) but with \( Y_T \) replaced by \( Y_T \frac{\partial}{\partial \alpha} h(X_{Tt}, \alpha) \) and \( \frac{\partial}{\partial \alpha} h(X_{Tt}, \alpha) \) respectively. Let \( \hat{A}_2 (\alpha, v) \) and \( \hat{A}_3 (\alpha, v) \) be defined analogously to \( \hat{A}_1 (\alpha, v) \) but with \( Y_T \) replaced by \( Y_T \frac{\partial}{\partial \alpha} h(X_{Tt}, \alpha) \) and \( \frac{\partial}{\partial \alpha} h(X_{Tt}, \alpha) \) respectively. We have

\[ \frac{\partial}{\partial \alpha} \hat{A}_1 (\alpha, v) = \frac{1}{T \Sigma_1} Y_T \frac{\partial}{\partial \alpha} h(X_{Tt}, \alpha) D^1 K \left[ \frac{v - h(X_{Tt}, \alpha)}{\hat{\sigma}_T} \right] / \hat{\sigma}_T^2 = D^1 \hat{A}_2 (\alpha, v), \]

\[ (10.4) \frac{\partial}{\partial \alpha} \hat{f}(\alpha, v) = \frac{1}{T \Sigma_1} \frac{\partial}{\partial \alpha} h(X_{Tt}, \alpha) D^1 K \left[ \frac{v - h(X_{Tt}, \alpha)}{\hat{\sigma}_T} \right] / \hat{\sigma}_T^2 = D^1 \hat{A}_3 (\alpha, v), \text{ and} \]

\[ \frac{\partial}{\partial \alpha} \hat{g}_1 (\alpha, v) = \left[ \hat{f}(\alpha, v) D^1 \hat{A}_2 (\alpha, v) - \hat{A}_1 (\alpha, v) D^1 \hat{A}_3 (\alpha, v) \right] / \hat{f}^2 (\alpha, v), \]

where \( D^1 K(v) \) denotes the derivative of \( K(\cdot) \) evaluated at \( v \) and \( D^1 \hat{A}_j (\alpha, v) \) denotes the derivative of \( \hat{A}_j (\alpha, v) \) with respect to \( v \) for \( j = 1, 2, 3 \).

For fixed \( \alpha \), Theorem II.10(a) establishes the consistency of \( \hat{f}(\alpha, v) \) for \( f_T(\alpha, v) \) uniformly over \( v \in \mathbb{R} \) (under suitable conditions). Similarly, for fixed \( \alpha \), Lemma A–1 of the Appendix (which is used to prove Theorem II.10) establishes the consistency of \( D^1 \hat{A}_j (\alpha, v) \) for \( (d^1 / dv^1)[g_j T(\alpha, v)f_T(\alpha, v)] \) uniformly over \( v \in \mathbb{R} \) for \( j = 1, 2, 3 \) and
\( \lambda = 0, 1 \) (under suitable conditions). Since \( \frac{\partial}{\partial \alpha} \hat{\Delta}_1 (\alpha, v) \) is a simple function of \( \hat{f}(\alpha, v) \) and \( D^\lambda \hat{A}_j (\alpha, v) \), this yields consistency of \( \frac{\partial}{\partial \alpha} \hat{\Delta}_1 (\alpha, v) \) for

\[
(10.5) \quad \left[ f_T(\alpha, v) \frac{d}{dv} [g_{2T}(\alpha, v)f_T(\alpha, v)] - g_{1T}(\alpha, v)f_T(\alpha, v) \frac{d}{dv} [g_{3T}(\alpha, v)f_T(\alpha, v)] \right] / f_T^2(\alpha, v),
\]

for fixed \( \alpha \), uniformly over \( \{ v : f_T(\alpha, v) \geq \epsilon \} \) or \( \{ v : \hat{f}(\alpha, v) \geq \epsilon \} \) for arbitrary \( \epsilon > 0 \).

As in Theorem II.10, consistency over expanding sets also can be obtained.

Next, under conditions that permit the interchange of certain integrals and derivatives, one can show that

\[
(10.6) \quad \frac{\partial}{\partial \alpha} [g_{1T}(\alpha, v)f_T(\alpha, v)] = \frac{d}{dv} [g_{2T}(\alpha, v)f_T(\alpha, v)] \quad \text{and}
\]

\[
\frac{\partial}{\partial \alpha} f_T(\alpha, v) = \frac{d}{dv} [g_{3T}(\alpha, v)f_T(\alpha, v)].
\]

Substituting these results into the expression

\[
(10.7) \quad \frac{\partial}{\partial \alpha} g_{1T}(\alpha, v) = \left[ f_T(\alpha, v) \frac{\partial}{\partial \alpha} [g_{1T}(\alpha, v)f_T(\alpha, v)] - g_{1T}(\alpha, v)f_T(\alpha, v) \frac{\partial}{\partial \alpha} f_T(\alpha, v) \right] / f_T^2(\alpha, v),
\]

shows that the "probability limit" of \( \frac{\partial}{\partial \alpha} \hat{\Delta}_1 (\alpha, v) \) given in (10.5) equals \( \frac{\partial}{\partial \alpha} \hat{\Delta}_1 (\alpha, v) \), as desired.

In sum, for fixed \( \alpha \), Theorem II.10 and Lemma A–1 can be used to establish that \( \frac{\partial}{\partial \alpha} \hat{\Delta}_1 (\alpha, v) - \frac{\partial}{\partial \alpha} \hat{\Delta}_1 (\alpha, v) \not\rightarrow 0 \) uniformly over \( \{ v : f_T(\alpha, v) \geq \epsilon \} \), over \( \{ v : \hat{f}(\alpha, v) \geq \epsilon \} \), or over expanding sets. These results can be extended to the case where \( \alpha \) is unknown, \( \hat{\alpha} \) is a consistent estimator of \( \alpha \), and \( \frac{\partial}{\partial \alpha} \hat{\Delta}_1 (\alpha, v) \) replaces \( \frac{\partial}{\partial \alpha} \hat{\Delta}_1 (\alpha, v) \) as the estimator, by employing the results of Theorem II.11 and Lemma A–4 of the Appendix. The above results can be extended to yield consistency of \( \frac{\partial}{\partial \alpha} \hat{\Delta}_1 (\alpha, v) \) for \( \frac{\partial}{\partial \alpha} \hat{\Delta}_1 (\alpha, v) \) uniformly over \( (\alpha, v) \) in \( A \times \{ v : \inf_{\alpha \in A} f_T(\alpha, v) \geq \epsilon \} \) or \( A \times \{ v : \inf_{\alpha \in A} \hat{f}(\alpha, v) \geq \epsilon \} \) for arbitrary \( \epsilon > 0 \),

where \( A \) is some bounded subset of \( \mathbb{R}^d \), by employing the results of Theorem II.12 and equation (B.56) of the proof of Theorem II.12. Results of the latter sort are needed for the MAD–DUC estimators of ASEM:I.
Last, the consistency of \( \frac{\partial^{\ell+1}}{\partial v^\ell \partial \alpha} \hat{\xi}_1(\alpha,v) \) for \( \ell \geq 1 \) and of \( \frac{\partial^2}{\partial \alpha \partial \alpha} \hat{\xi}_1(\alpha,v) \) for \( \frac{\partial^2}{\partial \alpha \partial \alpha} \hat{\xi}_1(\alpha,v) \), etc. and the extension of the results above to vector-valued index functions \( h(X_{\mathbf{T}_t}, \alpha) \) can be established in a manner analogous to that outlined above using the results of Section 7–9 and the Appendix.
APPENDIX

PROOF OF THEOREM II.1: Write $\nu_T(\cdot)$ as the sum of $m$ empirical processes $\{\nu_{Tj}(\cdot) : T \geq 1\}$ for $j = 1, \ldots, m$, where $\nu_{Tj}(\cdot)$ is based on the independent summands $\{m(W_{Tt}, \cdot) : t = j + sm, \ s = 1, 2, \ldots\}$. By standard inequalities it suffices to prove the stochastic equicontinuity of $\{\nu_{Tj}(\cdot) : T \geq 1\}$ for each $j$.

The latter can be proved using Pollard's (1990) proof of stochastic equicontinuity for his functional CLT (Theorem 10.7). We take his functions $f_{ni}(\omega, t)$ to be of the form $m(W_{Tt}, \tau)/\sqrt{T}$. We alter his pseudo-metric from $\lim_{N \to \infty} \left[ \frac{1}{N} \sum_{i=1}^{N} E|m(W_{Tt}, \tau_1) - m(W_{Tt}, \tau_2)|^2 \right]^{1/2}$ to that given in (3.1). Pollard's proof of stochastic equicontinuity relies on conditions (i) and (iii)-(v) of his Theorem 10.7. Condition (ii) of Theorem 10.7 is used only for obtaining convergence of the finite dimensional distributions, which we do not need, and for ensuring that his pseudo-metric is well-defined. Our pseudo-metric does not rely on this condition. Inspection of Pollard's proof shows that any pseudo-metric can be used for his stochastic equicontinuity result (although not for his total boundedness result) provided his condition (v) holds. Thus, it suffices to verify his conditions (i) and (iii)-(v).

Condition (i) requires that the functions $\{m(W_{Tt}, \tau)/\sqrt{T} : t \leq T, T \geq 1\}$ are "manageable." This holds under Assumption A because Pollard's packing numbers satisfy

$$\sup_{\omega \in \Omega, n \geq 1, \alpha \in R^n_+} D(\epsilon | \alpha \otimes F_n(\omega), \alpha \otimes \mathcal{F}_n(\omega)) \leq \sup_{Q \in \mathcal{Q}} N_2(\epsilon / 2, Q, \mathcal{L}).$$

Conditions (iii) and (iv) are implied by Assumption B. Condition (v) holds automatically given our choice of pseudo-metric. \( \square \)

PROOF OF THEOREM II.2: Type I classes of form (a) satisfy Pollard's entropy condition by Lemmas II.28 and II.36(ii) of Pollard (1984, pp. 30 and 34). Type I classes of form (b) satisfy Pollard's entropy condition because (i) they are contained in VC hull classes by
the proof of Proposition 4.4 of Dudley (1987) and the fact that \( \{ f : f(w) = w \cdot \xi \ \forall w \in \mathcal{W}, \xi \in \mathbb{R}^k \} \) is a VC major class, see Pollard (1984, Lemma II.18, p. 20), (ii) VC hull classes are contained in VC subgraph hull classes, and (iii) VC subgraph hull classes satisfy Pollard's entropy condition by Corollary 5.8 of Dudley (1987).

For classes of type II, consider the functions \( f(\cdot, \tau_1), \ldots, f(\cdot, \tau_n) \), where \( \tau_1, \ldots, \tau_n \) are points at the centers of disjoint cubes of diameter \( \epsilon(QF^2)^{1/2}/(QB^2)^{1/2} \) whose union covers \( T \) ( \( \subset \mathbb{R}^s \) for some \( s \geq 1 \)). Since

\[
\min_{j < n} \left[ \min_{j < n} \left[ \left( f(\cdot, \tau_j) - f(\cdot, \tau_j)^2 \right)^{1/2} \right] \right] \leq \min_{j < n} \left[ QB^2 \right]^{1/2} \left\| \tau - \tau_j \right\| \leq \epsilon \left( QF^2 \right)^{1/2},
\]

\( N_2(\epsilon(QF^2)^{1/2}, Q, \mathcal{T}) \) is the number of cubes above. By choice of the envelope \( F(\cdot) = 1 \vee \sup_{f \in \mathcal{T}} |f(\cdot)| \vee B(\cdot), \epsilon(QF^2)^{1/2}/(QB^2)^{1/2} \geq \epsilon \), the number of cubes is \( \leq C \epsilon^{-s} \) for some \( C > 0 \) and all \( Q \in \mathcal{Q} \). Thus, Pollard's entropy condition holds with envelope \( F(\cdot) \).

For classes of type III, Pollard's entropy condition holds because

\[
\sup_{Q \in \mathcal{Q}} N_2(\epsilon(QF^2)^{1/2}, Q, \mathcal{T}) \leq C \exp(-k_2/q) \forall \epsilon \in (0,1]
\]

for some \( C < \infty \) by Kolmogorov and Tihomirov (1961, Thm. XIII, p. 308). Since \( q > k_2/2 \) by assumption, Pollard's entropy condition holds. \( \square \)

**PROOF OF THEOREM II.3:** For \( \mathcal{G} \cup \mathcal{G}^* \), we have

\[
N_2(\epsilon, Q, \mathcal{G} \cup \mathcal{G}^*) \leq N_2(\epsilon, Q, \mathcal{G}) + N_2(\epsilon, Q, \mathcal{G}^*), \text{ and so,}
\]

\[
N_2(\epsilon(Q(G \vee G^*)^2)^{1/2}, Q, \mathcal{G} \cup \mathcal{G}^*) \leq N_2(\epsilon(QG^2)^{1/2}, Q, \mathcal{G}) + N_2(\epsilon(QG^2)^{1/2}, Q, \mathcal{G}^*),
\]

where the second inequality uses the facts that \( N_2(\epsilon, Q, \mathcal{T}) \) is nonincreasing in \( \epsilon \), \( Q(G \vee G^*)^2 \geq QG^2 \), and \( Q(G \vee G^*)^2 \geq QG^2 \). Pollard's entropy condition follows from the second inequality of (B.4).

For \( \mathcal{G} \cap \mathcal{G}^* \), it suffices to suppose that \( r = s = 1 \). As above, Pollard's entropy
condition follows from the inequalities

\[
N_2(\varepsilon, Q, \mathcal{G} \circ \mathcal{G}^*) \leq N_2(\varepsilon/2, Q, \mathcal{G}) N_2(\varepsilon/2, Q, \mathcal{G}^*) ,
\]

(B.5)

\[Q(G + G^*)^2 \geq QG^2, \text{ and } Q(G + G^*)^2 \geq QG^2,\]

where the first inequality holds because \( \min_{j, k \leq n} \left[ \int (g + g^* - g_j - g_k^*)^2 dQ \right]^{1/2} \)

\[\leq \min_{j \leq n} \left[ \int (g - g_j)^2 dQ \right]^{1/2} + \min_{k \leq n} \left[ \int (g^* - g_k^*)^2 dQ \right]^{1/2}.\]

For \( \mathcal{G} \chi \), each element of \( gh \) is a finite union of products of scalar functions, and so, using the result for \( \mathcal{G} \circ \mathcal{G}^* \), it suffices to suppose that \( r = s = u = 1 \). For notational simplicity, assume \( G = G \upharpoonright 1 \) and \( H = H \upharpoonright 1 \). Let \( Q_G(\cdot) = Q(\cdot G^2)/QG^2 \) and \( Q_H(\cdot) = Q(\cdot H^2)/QH^2 \). Note that \( Q_G, Q_H \in \mathcal{G} \). Let \( n = N_2(\varepsilon(Q_H G^2)^{1/2}, Q_H, \mathcal{G}) \) and \( n^* = N_2(\varepsilon(Q_G H^2)^{1/2}, Q_G, \chi) \). Let \( g_1, \ldots, g_n \) and \( h_1, \ldots, h_{n^*} \) denote approximating functions in \( \mathcal{G} \) and \( \chi \), respectively, that correspond to the cover numbers \( n \) and \( n^* \). We use \( g_jh_k \) to approximate \( gh \) for \( g \in \mathcal{G} \) and \( h \in \chi \):

\[
\min_{j \leq n, k \leq n^*} \left[ \int (gh - g_jh_k)^2 dQ \right]^{1/2}
\]

(B.6)

\[\leq \min_{j \leq n} \left[ QH^2 \int (g - g_j)^2 d \left[ \frac{Q(\cdot H^2)}{QH^2} \right] \right]^{1/2} + \min_{k \leq n^*} \left[ QG^2 \int (h - h_k)^2 d \left[ \frac{Q(\cdot G^2)}{QG^2} \right] \right]^{1/2}
\]

\[\leq \left[ QG^2H^2 \right]^{1/2} \varepsilon.
\]

Thus, we get

\[N_2 \left[ \varepsilon(Q_G H^2)^{1/2}, Q, \mathcal{G} \right] \leq N_2 \left[ \varepsilon(Q_H G^2)^{1/2}, Q_H, \mathcal{G} \right] N_2 \left[ \varepsilon(Q_G H^2)^{1/2}, Q_G, \chi \right] \text{ and}
\]

\[
\sup_{Q \in \mathcal{G}} N_2 \left[ \varepsilon(Q_G H^2)^{1/2}, Q, \mathcal{G} \right]
\]

(B.7)

\[\leq \sup_{Q_H \in \mathcal{G}} N_2 \left[ \varepsilon(Q_H G^2)^{1/2}, Q_H, \mathcal{G} \right] \sup_{Q_G \in \mathcal{G}} N_2 \left[ \varepsilon(Q_G H^2)^{1/2}, Q_G, \chi \right]
\]

\[= \sup_{Q \in \mathcal{G}} N_2 \left[ \varepsilon(Q_G^2)^{1/2}, Q, \mathcal{G} \right] \sup_{Q \in \mathcal{G}} N_2 \left[ \varepsilon(Q_H^2)^{1/2}, Q, \chi \right].\]
Pollard's entropy condition follows from the latter inequality.

For $\mathcal{G} \lor \mathcal{G}^*$, it suffices to suppose $r = s = 1$. Pollard’s entropy condition follows from the inequalities

$$N_2(\epsilon, Q, \mathcal{G} \lor \mathcal{G}^*) \leq N_2(\epsilon/2, Q, \mathcal{G})N_2(\epsilon/2, Q, \mathcal{G}^*),$$

$$Q(\mathcal{G} \lor \mathcal{G}^*)^2 \geq QG^2,$$

and

$$Q(\mathcal{G} \lor \mathcal{G}^*)^2 \geq QG^2,$$

where the first inequality uses $|g \lor g^* - g_j \lor g_j^*| \leq |g - g_j| + |g^* - g_j^*|$. The proof for $\mathcal{G} \land \mathcal{G}^*$ is analogous (with the envelope still given by $\mathcal{G} \lor \mathcal{G}^*$ rather than $\mathcal{G} \land \mathcal{G}^*$). The result for $|\mathcal{G}|$ follows because $||g| - |a_j|| \leq |g - a_j|$

Lastly, consider $\mathcal{G}^{-1}$. For $g \in \mathcal{G}$, let $g^\ell$ denote the $\ell$-th element of $g$, where $\ell = 1, \ldots, L$ and $L = r^2$. Let $\mathcal{G}_\ell = \{g^\ell : g \in \mathcal{G}\}$ and $n_\ell = N_2(\epsilon/2, Q, \mathcal{G}_\ell)$ for some $Q \in \mathcal{Q}$. We claim that given any $\epsilon > 0$ and $Q \in \mathcal{Q}$, there exist functions $g_1, \ldots, g_n$ in $\mathcal{G}$ with $n \leq \prod_{\ell=1}^L n_\ell$ such that for all $g \in \mathcal{G}$

$$\min_{j \leq n} \max_{\ell \leq L} \left( Q(g^\ell - g_j^\ell)^2 \right)^{1/2} \leq \epsilon.$$  

To see this, note that by the assumption that $\mathcal{G}$ satisfies Pollard’s entropy condition, for each $\ell$ there exist real functions $g_{\ell1}, \ldots, g_{\ell n_\ell}$ in $\mathcal{G}_\ell$ such that for all $g \in \mathcal{G}$

$$\min_{j \leq n_\ell} \left( Q(g^\ell - g^\ell_{\ell j})^2 \right)^{1/2} < \epsilon/2.$$  

Form the set $\mathcal{G}^+$ of all $R^L$-valued functions whose $\ell$-th element is $g^\ell_{\ell j}$ for some $j = 1, \ldots, n_\ell$ for $\ell = 1, \ldots, L$. The number of such functions is $n^+ = \prod_{\ell=1}^L n_\ell$. The functions in $\mathcal{G}^+$ are not necessarily in $\mathcal{G}$. For each function $g^+$ in $\mathcal{G}^+$ consider the $L^2(Q)$ $\epsilon/2$-ball in $\mathcal{G}$ centered at $g^+$. Take one function from each non-empty ball and let $g_1, \ldots, g_n$ denote the chosen functions. These functions satisfy the claim above.

If $\mathcal{G}$ satisfies Pollard's entropy condition with envelope $G$, it also does so with envelope $G \lor 1$. For notational simplicity, suppose $G = G \lor 1$. Given $Q \in \mathcal{Q}$, let
\( \tilde{Q}(\cdot) = Q(\cdot \bar{G}^4)/Q \bar{G}^4 (\epsilon \mathcal{G}) \). Take \( \epsilon \) and \( Q \) in the claim above to equal \( \epsilon(\bar{Q}G^4)^{1/2}/r^4 \) and \( \bar{Q} \) respectively. Then, there exist functions \( g_1, \ldots, g_n \) in \( \mathcal{G} \) such that

\[
\min_{j \leq n} \max_{\ell \leq L} \left[ \tilde{Q}(\epsilon \mathcal{G}^4)^{1/2}/r^4 \right]^{1/2} \leq \epsilon(\bar{Q}G^4)^{1/2}/r^4 \quad \text{and} \quad n \leq \prod_{\ell=1}^{L} N_2 \left[ \epsilon(\bar{Q}G^4)^{1/2}/r^4, \bar{Q}, \mathcal{G}_\delta \right].
\]

Let \( 1_{\epsilon}^r = (1, \ldots, 1)^r (\epsilon \in \mathbb{R}^r) \) and let \( \| \cdot \| \) denote the matrix of absolute values of the matrix.

For arbitrary unit vectors \( b, c \in \mathbb{R}^r \), we have

\[
\min_{j \leq n} Q \left[ b'g^{-1}c - b'g_j^{-1}c \right]^2 = \min_{j \leq n} Q \left[ b'g^{-1}(g_j - g_j)g_j^{-1}c \right]^2 \leq \min_{j \leq n} r^4 Q \left[ \epsilon \mathcal{G}^4 \right]_{1r} g_j - g^2 \left[ g_j - g \right] ^2 \leq \min_{j \leq n} r^4 \mathcal{G}^4_{1r} \sum_{\ell=1}^{L} \sum_{m=1}^{L} \mathcal{G}^4_{\ell} \mathcal{G}^4_{m} - \mathcal{G}^4_{j} \mathcal{G}^4_{j} \mathcal{G}^4_{m} - \mathcal{G}^4_{j} \mathcal{G}^4_{j} \mathcal{G}^4_{m} \]

Thus, \( N_2(\epsilon(\mathcal{Q}G^4)^{1/2}, Q, \mathcal{G}^{-1}) \leq \prod_{\ell=1}^{L} N_2 \left[ \epsilon(\bar{Q}G^4)^{1/2}/r^4, \bar{Q}, \mathcal{G}_\delta \right] \) and

\[
\sup_{Q \in \mathcal{Q}} N_2 \left[ \epsilon(\mathcal{Q}G^4)^{1/2}, Q, \mathcal{G}^{-1} \right] \leq \sup_{Q \in \mathcal{Q}} \prod_{\ell=1}^{L} N_2 \left[ \epsilon(\bar{Q}G^4)^{1/2}/r^4, \bar{Q}, \mathcal{G}_\delta \right].
\]

The integral over \( \epsilon \in [0,1] \) of the square root of the logarithm of the right-hand side (rhs) of (B.11) is finite since \( \mathcal{G} \) satisfies Pollard's entropy condition with envelope \( \mathcal{G} = \mathcal{G} \lor 1 \).

Thus, \( \mathcal{G}^{-1} \) satisfies Pollard's entropy condition with envelope \( (\mathcal{G} \lor 1)^2 \mathcal{G}^2 \).

PROOF OF THEOREM II.4: By the same argument as in the proof of Theorem II.1, it suffices to prove the result when \( \{W_T : t \leq T\} \) are independent rv's. By Markov's inequality and Theorem 2 of Pollard (1989), we have

\[
\max_{T \to \infty} \mathbb{P}^* \left[ \sup_{\rho_T(\tau_1, \tau_2) < \delta} |\nu_T(\tau_1) - \nu_T(\tau_2)| > \eta \right]
\]

\[
\leq \max_{T \to \infty} \mathbb{E}^* \left[ \sup_{\rho_T(\tau_1, \tau_2) < \delta} |\nu_T(\tau_1) - \nu_T(\tau_2)| / \eta \right]
\]

\[
\leq \max_{T \to \infty} \frac{1}{\sqrt{T}} \mathbb{E}(W_{T}) \mathbb{E}(W_{T}) > \sqrt{T} \mathbb{E} \left[ \epsilon \right] / \eta + C_{\delta} \left( \log N_2(\epsilon, \mathcal{P}, \mathcal{F}) \right)^{1/2} \frac{d\epsilon}{\eta}
\]
for some constant $C < \infty$, where $\xi_\delta > 0$ is a constant that does not depend on $T$.

The second term on the right-hand side of (B.12) can be made arbitrarily small by choice of $\delta$ using Assumption D. The first term is less than or equal to

$$\lim_{T \to \infty} 4 T^{\delta/2} \frac{1}{T} \sum_1^T \text{EM}^{2+\delta}(W_{Tt})1(\text{M}(W_{Tt}) > \sqrt{T\xi_\delta})/\xi_\delta^{1+\delta} = 0$$

using Assumption B. The result follows. ☐

PROOF OF THEOREM II.5: It suffices to prove the result for classes of type III–VI, because a type II class with $\sup_{t \leq T, T \geq 1} (\text{EB}^P(W_{Tt}))^{1/p} < \infty$ is a type IV class under $P$ with index $p$.

First, we consider classes of type III. For given $\epsilon > 0$, define the functions $a_j, b_j$, $j = 1, \ldots, n_\epsilon$ of the definition of $L^P$ bracketing cover numbers as follows: (a) $\forall w \in \mathcal{Y}$ such that $w_a \in \mathcal{Y}_a - \mathcal{Y}_a^*$, let $a_j(w) = K$ and $b_j(w) = 0 \ \forall j$ and (b) $\forall w \in \mathcal{Y}$ such that $w_a \in \mathcal{Y}_a^*$, let $\{a_j(w); j = 1, \ldots, n_\epsilon\}$ be the functions constructed by Kolmogorov and Tihomirov (1961, pp. 312–314) in their proof of Theorem XIV and let $b_j(w) = \epsilon \ \forall j$. These functions satisfy the conditions for $L^P$ bracketing cover numbers for all $p \in [2, \infty]$. Hence, $\text{NB}_{p}(\epsilon, P, \mathcal{Y}) \leq n_\epsilon \ \forall \epsilon \in (0, 1], \forall p \in [2, \infty]$. The number $n_\epsilon$ of such functions is $\leq C \exp(\epsilon^{-k_a/q}) \ \forall \epsilon \in (0, 1]$ for some $C < \infty$ by Kolmogorov and Tihomirov (1961, Thm. XIV). Since $q > k_a/2$ by assumption, Ossiander's entropy condition holds for all $p \in [0, \infty]$.

For a type IV class with index $p$, consider disjoint cubes in $\mathcal{Y}$ of diameter $\delta = (\epsilon/C)^{1/\psi}$. The number $N(\epsilon)$ of such cubes satisfies $N(\epsilon) \leq C^* \epsilon^{-d/\psi}$ for some $C^* < \infty$, where $d$ is the dimension of $\mathcal{T}$. Let $\tau_j$ be some element of the $j$–th cube in $\mathcal{T}$. Let $a_j(\cdot) = f(\cdot, \tau_j)$ and $b_j = \sup_{\|\tau - \tau_j\| < \delta} |f(\cdot, \tau) - a_j(\cdot)|$. By (4.3),

$$\sup_{t \leq T, T \geq 1} \left[\text{EB}^P(W_{Tt})\right]^{1/p} \leq C \delta^\psi = \epsilon. \text{ Thus, } N_{p}^B(\epsilon, P, \mathcal{Y}) \leq N(\epsilon).$$

Since $\int_0^1 (\log N(\epsilon))^{1/2} d\epsilon < \infty$, Ossiander's $L^P$ entropy condition holds.
For a type $\nu$ class with index $p$, let $\mathcal{W}_r = \mathcal{W} \cap \{ w \in \mathbb{R}^k : \| w \| \leq r \}$, let $\mathcal{T}_r$ denote the class of functions $\mathcal{T}$ restricted to $\mathcal{W}_r$, and let $N_\omega(\mathcal{W}_r, \mathcal{F}_r)$ be the minimal number $n$ of real functions $f_1, \ldots, f_n$ on $\mathcal{W}_r$ such that $\min_{j \leq n} \sup_{w \in \mathcal{W}_r} |f(w) - f_j(w)| < \epsilon$ for each $f \in \mathcal{T}_r$. We claim that

$$N_B^B(\epsilon, p, \mathcal{F}) \leq N_\omega(\epsilon/2, \mathcal{W}_r(\epsilon), \mathcal{T}_r(\epsilon)),$$

where $r(\epsilon) = C\epsilon^{-p/\zeta}$ for some constant $C < \omega$ when $p < \omega$ and $r(\epsilon) = \sup\{ \|w\| : w \in \mathcal{W}\} (\leq \omega)$ when $p = \omega$.

Using the proof of Theorem XIV of Kolmogorov and Tihomirov (1961, pp. 312–314), it can be seen that

$$\log N_\omega(\mathcal{W}_r(\epsilon), \mathcal{T}_r(\epsilon)) \leq D r(\epsilon) \leq D^* \epsilon^{-k_0/q} \leq k_0 \left[ \frac{p}{\zeta} + 1 \right]$$

for some constants $D, D^* < \omega$, where the second inequality holds only when $p < \omega$. When $p < \omega$, (B.14) and (B.15) combine to yield Ossiander's $L^p$ entropy condition for $\mathcal{T}$ if $k_0 (p/\zeta + 1/q)/2 < 1$, or equivalently, if $\zeta > pqk_0/(2q - k_0)$ and $q > k_0/2$, as is assumed. When $p = \omega$, (B.14) and the first inequality of (B.15) combine to yield Ossiander's $L^p$ entropy condition for $\mathcal{T}$ provided $q > k_0/2$, as is assumed.

It remains to show (B.14). For $p = \omega$, (B.14) follows immediately from the definition of $N_B^B(\cdot)$ and $N_\omega(\cdot)$, since $\mathcal{W}_r(\epsilon) = \mathcal{W}$ and $\mathcal{T}_r(\epsilon) = \mathcal{T}$ when $p = \omega$. Next, suppose $p < \omega$. For $n = N_\omega(\epsilon/2, \mathcal{W}_r, \mathcal{T}_r)$, define real functions $a_j, b_j, j = 1, \ldots, n$ on $\mathcal{W}$ as follows: On $\mathcal{W}_r$, take $\{a_j(\cdot) : j = 1, \ldots, n\}$ to be the functions constructed by Kolmogorov and Tihomirov (1961, pp. 312–214) in their proof of Theorem XIV and let $b_j(\cdot) = \epsilon/2$ for $j = 1, \ldots, n$. On $\mathcal{W} - \mathcal{W}_r$, take $a_j(\cdot) = 0$ and take $b_j(\cdot) = F$ for $j = 1, \ldots, n$, where $F$ is a constant for which $\sup_{w \in \mathcal{W}} |f(w)| \leq F \forall f \in \mathcal{F}$. Then, for each $f \in \mathcal{F}$, $\min_{j \leq n} |f - a_j| \leq b_j$ and
\[
\sup_{t \leq T, T \geq T_1} \text{E}^P \left( W_{Tt} \right) = \sup_{t \leq T, T \geq T_1} \text{E}^P \left( W_{Tt} \right) \mathbb{1}(W_{Tt} \in \mathbb{H}_r)
\]
\[
+ \sup_{t \leq T, T \geq T_1} \text{E}^P \left( W_{Tt} \right) \mathbb{1}(W_{Tt} \in \mathbb{H} - \mathbb{H}_r)
\]
\[
\leq \left( \varepsilon/2 \right)^P + \left( \varepsilon/2 \right)^P \sup_{t \leq T, T \geq T_1} \text{E} \| W_{aTt} \| \mathbb{1}^{1/\zeta} = \left( \varepsilon/2 \right)^P + C^* \varepsilon^{-\zeta},
\]
where \( C^* \) is defined implicitly. If we let \( r = r(\varepsilon) = \left( 2^{P \cdot C^*/(2^P - 1)} \right)^{1/\zeta} \varepsilon^{-\zeta} \), then
\[\sup_{t \leq T, T \geq T_1} \text{E}^P \left( W_{Tt} \right) \leq \varepsilon^P \] and (B.14) holds.

Last, we consider type VI classes of functions. First, suppose \( p < \infty \). We derive an upper bound on \( N^B_p(\varepsilon; r, P, \mathcal{F}) \) for arbitrary \( \varepsilon > 0 \). Let \( r_\varepsilon = C \varepsilon^{-P/\zeta} \) for some \( C < \infty \) and let \( F \) be a constant for which \( \sup_{w \in \mathbb{H}} | f(w) | \leq F \) \( \forall f \in \mathcal{F} \). Let \( J \) be the index of a set \( \mathbb{H}_a \) that does not include \( \{ w_a \in \mathbb{H}_a : \| w_a \| \leq r_\varepsilon \} \). For functions \( f \in \mathcal{F} \) whose corresponding integer of part (ii) (of the definition of type VI classes) is \( J \), take the centering and \( \varepsilon \)-bracketing functions \( \{ (a_\varepsilon, b_\varepsilon) : \ell = 1, \ldots, n_{\varepsilon J} \} \) (of the definition of \( L^P \) bracketing cover numbers) as follows: (a) \( \forall w \in \mathbb{H} \) such that \( \| w_a \| > r_\varepsilon \), let \( a_\varepsilon(w) = 0 \) and \( b_\varepsilon(w) = F \), (b) \( \forall w \in \mathbb{H} \) such that \( \| w_a \| \leq r_\varepsilon \) and \( w_a \notin \mathbb{H}_a \), let \( a_\varepsilon(w) = K_J \) and \( b_\varepsilon(w) = 0 \), and (c) \( \forall w \in \mathbb{H} \) such that \( \| w_a \| \leq r_\varepsilon \) and \( w_a \in \mathbb{H}_a \), let \( \{ a_\varepsilon(w) : \ell = 1, \ldots, n_{\varepsilon J} \} \) be the functions constructed by Kolmogorov and Tihomirov (1961) in the proof of their Theorem XIV and let \( b_\varepsilon(w) = \varepsilon/2 \) \( \forall \ell \). The number \( n_{\varepsilon J} \) of such functions is \( \leq D_1 \exp \left[ D_2 r_\varepsilon^{-\zeta} \right] \) by Theorem XIV of Kolmogorov and Tihomirov (1961), since \( \{ w : \| w_a \| \leq r_\varepsilon, w_a \in \mathbb{H}_a \} \subset \{ w : \| w_a \| \leq r_\varepsilon \} \).

Next, for all functions \( f \in \mathcal{F} \) whose corresponding integer \( J \) of part (ii) is such that \( \mathbb{H}_a \) contains \( \{ w_a \in \mathbb{H}_a : \| w_a \| \leq r_\varepsilon \} \), take the centering and \( \varepsilon \)-bracketing functions \( \{ (a_\varepsilon, b_\varepsilon) : \ell = 1, \ldots, n_{\varepsilon J} \} \) as follows: (a) \( \forall w \in \mathbb{H} \) such that \( \| w_a \| > r \), let \( a_\varepsilon(w) = 0 \) and \( b_\varepsilon(w) = F \) \( \forall \ell \) and (b) \( \forall w \in \mathbb{H} \) such that \( \| w_a \| \leq r \), let \( \{ a_\varepsilon(w) : \ell = 1, \ldots, n_{\varepsilon J} \} \) be the functions constructed by Kolmogorov and Tihomirov (1961) in the proof of their
Theorem XIV and let \( b(\omega) = \epsilon / 2 \ \forall \ell \). The number of such functions is also is \( D_1 \exp[D_2 r_\epsilon ^{k_a - k_a / q}] \).

Now, the number of indices \( J \) for which \( w_a \in \mathcal{W}_a : \|w_a\| \leq r \) is \( n(r_\epsilon) \). Hence, the total number of centering/\( \epsilon \)-bracketing functions considered above is \( \leq (n(r_\epsilon) + 1)D_1 \exp[D_2 r_\epsilon ^{k_a - k_a / q}] \). Also note that \( \sup_{t \in T, T \geq 1} (E b_\ell^p(W T_t))^1 / p < \epsilon \) for all of the functions \( b_\ell \) introduced above by the same calculations as in (B.16) provided \( C \) (of the definition of \( r_\epsilon \)) is defined suitably. Hence,

\[
N P_\epsilon(p, \mathcal{F}) \leq (n(r_\epsilon) + 1)D_1 \exp[D_2 r_\epsilon ^{k_a - k_a / q}]
\]

(B.17)

\[
\leq (K_1 \exp[K_2 C^{\xi - \xi / \xi}] + 1)D_1 \exp[D_2 C^{\xi / \xi + 1 / q}] \cdot
\]

With this bound, Ossiander's \( L^p \) entropy condition holds provided \( p\xi / (2\zeta) < 1 \) and \( k_a (p / \zeta + 1 / q) / 2 < 1 \), or equivalently, \( \zeta < 2p / q \), \( q > k_a / 2 \), and \( \zeta > p q k_a / (2q - k_a) \), as is assumed.

For the case where \( p = \infty \), take \( r(\epsilon) = \sup\{\|w_a\| : w \in \mathcal{W}\} < \infty \ \forall \epsilon > 0 \) in the argument above. Then, Ossiander's \( L^\infty \) entropy condition holds provided \( q > k_a / 2 \), as is assumed.

PROOF OF THEOREM II.6: For \( \mathcal{G} \cup \mathcal{G}^* \), the result is obvious. For \( \mathcal{G} \cup \mathcal{G}^* \), it suffices to suppose that \( r = s = 1 \). Let \( (g, a_\ell, b_\ell) \) and \( (g^*, a_\ell^*, b_\ell^*) \) for \( g \in \mathcal{G} \) and \( g^* \in \mathcal{G}^* \) be defined analogously to \( (f, a_\ell, b_\ell) \) given in the definition of the \( L^p \) bracketing cover numbers. We have

\[
\left( E (b_\ell + b_\ell^*)^p \right)^{1 / p} \leq \left[ E b_\ell^p \right]^{1 / p} + \left[ E b_\ell^* \right]^{1 / p} \leq 2\epsilon , \text{ and so,}
\]

(B.18)

\[
N P_\epsilon(p, \mathcal{G} \cup \mathcal{G}^*) \leq N P_\epsilon(p, \mathcal{G}) N P_\epsilon(p, \mathcal{G}^*). \]

The result follows.

For \( \mathcal{G} \cup \mathcal{G}^* \), it also suffices to suppose that \( r = s = 1 \). We have
\[ |g \vee g^* - a_j \vee a_j^*| \leq |g - a_j| + |g^* - a_j^*| \leq b_j + b_j^* \text{ and so,} \]
\[ N^B_p(2\epsilon, P, g \vee g^*) \leq N^B_p(\epsilon, P, g) N^B_p(\epsilon, P, g^*). \]

The result for \( G \wedge g^* \) is analogous.

For \( \|g\| \), the result follows from the inequality \( \|g\| - |a_j| \leq |g - a_j| \).

Next consider \( g^{-1} \). For \( g \in G \), let \( g^\ell \) denote the \( \ell \)-th element of \( g \) for \( \ell = 1, \ldots, L \), where \( L = r^2 \). By the same argument as used to prove the claim in the proof of the \( g^{-1} \) result of Theorem II.3, there exist \( r \times r \) matrix functions \( a_1, \ldots, a_n \) and \( b_1, \ldots, b_n \) such that (i) \( a_j \in G \) for all \( j \leq n \), (ii) for all \( g \in G \), \( |g^\ell - a_j^\ell| \leq b_j^\ell \) for all \( \ell = 1, \ldots, L \) for some \( j \leq n \), (iii) \( \left[ E(b_j^\ell)^p \right]^{1/p} < \epsilon \ \forall \ell, \forall j \), and (iv) \( n \leq L \prod_{\ell=1}^L N^B_2(\epsilon/2, P, G^\ell) \).

By an eigenvector/eigenvalue decomposition, we get \( |g^{-1}| \leq 1_{r \times r}(I_r/\lambda_*)^{1_{r \times r}} \) element by element and \( |a_j^{-1}| \leq 1_{r \times r}(I_r/\lambda_*) \). Thus, for arbitrary unit vectors \( b, c \in \mathbb{R}^r \), we have: For any \( g \in G \) there exists \( a_j \) and \( b_j \) for which
\[ |b^\ell g^{-1}c - b^\ell a_j^{-1}c| \leq |b^\ell| |g^{-1}| |a_j - g| |a_j^{-1}| |c| \leq (r^2/\lambda_*)^{1_{r \times r}} \text{ and} \]
\[ \left[ E[(r^4/\lambda_*)^{1_{r \times r}} b^\ell b_j] \right]^{1/p} \leq (r^2/\lambda_*)^{\epsilon}. \]

Thus, \( N^B_p(r^6\epsilon/\lambda_*, P, g^{-1}) \leq n \leq \prod_{\ell=1}^L N^B_2(\epsilon/2, P, G^\ell) \) and the result follows.

To prove part (b) of Theorem II.6 concerning \( G \), note that each element of \( gh \) (for \( g \in G \) and \( h \in X \)) is a finite union of products of scalar functions, and so, using the result for \( G \odot g^* \) it suffices to suppose that \( r = s = u = 1 \). Let \( (g, a_j, b_j) \) and \( (h, a_j^*, b_j^*) \) be defined analogously to \( (f, a_j, b_j) \) given in the definition of the \( L^p \) bracketing-cover numbers, with \( p = \lambda \) and \( p = \mu \) respectively. We have
\[ |gh - a_j a_j^*| \leq |gh - ga_j^*| + |ga_j^* - a_j a_j^*| \]
\[ \leq Gb_j^* + |a_j^* - h + h| b_j \leq Gb_j^* + Hb_j + b_j b_j^* \]
and
\[
\left[ E(Gb_{ij} + Hb_{ij} + b_{ij}b_{ij})^{\frac{1}{\alpha}} \right] \leq \left[ EG^{\alpha}b_{ij}^{\alpha} \right]^{\frac{1}{\alpha}} + \left[ EH^{\alpha}b_{ij}^{\alpha} \right]^{\frac{1}{\alpha}} + \left[ Eb_{ij}^{\alpha}b_{ij}^{\alpha} \right]^{\frac{1}{\alpha}}
\]

\[
\leq \left[ \frac{a_{ij}^{\mu}}{\alpha} \right]^{\frac{\mu-\alpha}{\alpha}} \left[ Eb_{ij}^{\mu} \right]^{\frac{1}{\mu}} + \left[ EH^{\alpha} \right]^{\frac{\lambda-\alpha}{\alpha}} \left[ Eb_{ij}^{\lambda} \right]^{\frac{1}{\lambda}} + \left[ Eb_{ij}^{\mu} \right]^{\frac{\mu-\alpha}{\alpha}} \left[ Eb_{ij}^{\mu} \right]^{\frac{1}{\mu}}
\]

\[
\leq \sup_{t \leq T, T \geq t} \left[ \left[ EG^{\lambda} \right]^{1/\lambda} + \left[ EH^{\mu} \right]^{1/\mu} \right] \epsilon + \epsilon^2
\]

\[
\leq C^* \epsilon
\]

for \( \epsilon \in (0,1) \), where \( C^* \) is defined implicitly and the dependence of each of the functions \( G, b_{ij}^{\mu} \), etc. on \( W_{T_0} \) is suppressed for notational simplicity. The second and third inequalities hold by Hölder's inequality and the fact that \( \frac{\lambda \mu}{\lambda + \mu} \geq \alpha \) implies that \( \frac{a_{ij}^{\mu}}{\alpha} \leq \lambda \) and \( \frac{\alpha \lambda}{\lambda - \alpha} \leq \mu \). Equations (B.21) and (B.22) imply that

\[
N_{\alpha}^B(\epsilon, \mathcal{P}, \mathcal{G}) \leq N_{\lambda}^B(\epsilon, \mathcal{P}, \mathcal{G})N_{\mu}^B(\epsilon, \mathcal{P}, \mathcal{G})
\]

and the desired result follows. Note that using the notational conventions stated in the text, (B.21)–(B.23) hold whether or not \( \alpha = \infty \), \( \lambda = \infty \), or \( \mu = \infty \). \( \square \)

**PROOF OF THEOREM II.7:** For the case where \( \mathcal{W} = \mathcal{W}^* \) or \( \mathcal{W}_a = \mathcal{W}_a^* \), Theorem II.7 follows immediately from Andrews (1989a, Proof of Theorem 4 and Comment 5 following Theorem 4). For the case where Assumption E holds and \( \mathcal{W}^* \) is a proper subset of \( \mathcal{W} \), one only needs to adjust the former proof by changing the series expansion from

\[
m(w, \tau) = \sum_{j=1}^{\infty} c_j(\tau)h_j(w) \to
\]

\[
m(w, \tau) = K1(w \in \mathcal{W} - \mathcal{W}^*) + \sum_{j=1}^{\infty} c_j(\tau)h_j(w)1(w \in \mathcal{W}^*).\]

Since the initial term of the adjusted expansion is the same for each function in \( \mathcal{W} \), it drops out when differences \( m(w, \tau_1) - m(w, \tau_2) \) are considered and has no effect on the argument used in Andrews (1989a). An analogous adjustment is used when Assumption F holds and \( \mathcal{W}_a^* \) is a proper subset of \( \mathcal{W}_a \). \( \square \)
PROOF OF THEOREM II.9: First, we state some relevant results from Andrews (1989a).

By (3.3)—(3.5) of Andrews (1989a) plus the adjustment given in the proof of Theorem II.7, $m_a(w_a, \tau_a)$ has a pointwise convergent series expansion

\[ K_1(w_a \in \mathbb{W} - \mathbb{W}_a^*) + \sum_{j=1}^{\infty} c_j(\tau_a) h_j(w_a) 1(w_a \in \mathbb{W}_a) \forall w_a \in \mathbb{W}_a, \forall \tau_a \in \mathcal{T}_a, \]

where \( \{h_j(\cdot): j \geq 1\} \) are complex exponential functions that are uniformly bounded (over \( j \) and \( w_a \)) and \( \{c_j(\tau_a): j \geq 1\} \) are complex constants. Let \( v_j = j^{\epsilon-2q/k_a} \) for \( j \geq 1 \) for some \( \epsilon \in (0, -1 + 2q/k_a) \). As defined, \( \sum_{j=1}^{\infty} v_j < \infty \). By (3.4), (3.10), (3.11), and (2.6) of Andrews (1989a), \( \{c_j(\tau_a)\} \) are such that

\[
\sup_{\tau_a \in \mathcal{T}_a} \sum_{j=1}^{\infty} |c_j(\tau_a)|^2/v_j \to 0 \text{ as } J \to \infty, \quad \sup_{\tau_a \in \mathcal{T}_a} \sum_{j=1}^{\infty} |c_j(\tau_a)|^2/v_j < \infty, \tag{B.25}
\]

\[
\sum_{j=1}^{\infty} |c_j(\tau_a)| < \infty \forall \tau_a \in \mathcal{T}_a, \text{ and } \lim_{\delta \to 0} \sup_{\tau_1, \tau_2 \in \mathcal{T}_a} \sum_{j=1}^{\infty} |c_j(\tau_1) - c_j(\tau_2)|^2/v_j = 0.
\]

For notational simplicity, let \( m_u(\tau_u) = m_u(W_{uT_t}, \tau_u) \) for \( u = a, b \), let \( h_j = h_j(W_{aT_t}) 1(W_{aT_t} \in \mathbb{W}_a^*) \) for \( j \geq 1 \) and \( h_0 = 1(W_{aT_t} \in \mathbb{W}_a - \mathbb{W}_a^*) \), let \( c_0(\tau_a) = K \forall \tau_a \in \mathcal{T}_a \), and let \( v_0 = 1 \). Then, \( m_a(w_a, \tau_a) = \sum_{j=1}^{\infty} c_j(\tau_a) h_j \) and we have

\[
\sup_{\rho(\tau_1, \tau_2) < \delta} \left| \sqrt{\sum_{j=1}^{\infty} |c_j(\tau_1a) m_{bt}(\tau_1b) - m_{bt}(\tau_2a) m_{bt}(\tau_2b)|} \right|
\]

\[
= \sup_{\rho(\tau_1, \tau_2) < \delta} \left| \sqrt{\sum_{j=1}^{\infty} |c_j(\tau_1a) h_j m_{bt}(\tau_1b) - h_j m_{bt}(\tau_1b)|} \right|
\]

\[
- c_j(\tau_2a)(h_j m_{bt}(\tau_2b) - h_j m_{bt}(\tau_2b)) \right|
\]
Combining (B.25) and (B.26), it suffices to show that

\begin{align}
\lim_{T \to \infty} \sup_{\tau_b \in \mathcal{F}_b} \left| \sup_{\tau_a \in \mathcal{F}_a} \frac{\sum c_i(\tau_a) - c_i(\tau_{2a})}{\sqrt{v}} \right| & < \infty \quad \text{and} \\
\lim_{\delta \to 0} \sup_{T \to \infty} \left| \sup_{\tau_a \in \mathcal{F}_a} \frac{\sum c_i(\tau_a)}{\sqrt{v}} \right| & = \infty
\end{align}

With " \sup " removed, (B.27) and (B.28) hold by Theorem 2.3 of Andrews and Pollard (1990). In fact, the proof of their Theorem goes through with " \sup " added, using the fact that \sup_{j \geq 0} \sup_{w_a \in \mathcal{H}_a} |h_j(w_a)| < \infty. Thus, (B.27) and (B.28) hold. 

The proof of Theorem II.10 uses the following Lemma:
LEMMA A-1: Under Assumptions NP1–NP5,

\[
\sup_{x \in \mathbb{R}^k} \left| T^{-\eta/(2\eta+1)} \sigma_{1T}^{-k-|\lambda|-1/(2\eta+1)} \right| + O_p \left( \sigma_{2T}^{\omega-|\lambda|} \right)
\]

PROOF OF THEOREM II.10: Part (a) follows immediately from Lemma A-1 by taking \( Y_{Tt} = g_T(x) = 1 \) \( \forall x, t, T \) in Lemma A-1.

Next, we establish part (b). Let \( h(x) \) denote \( g_T(x)^{1/2} f_{Tt}(x) \), let \( \hat{h}(x) \) denote \( \hat{g}(x) \hat{f}(x) \), and let \( f(x) \) denote \( \frac{1}{T} \Sigma_{1T}^T f_{Tt}(x) \). Provided \( f(x) \neq 0 \), we have by the chain rule,

\[
(D^\lambda g_T(x) = D^\lambda [h(x)f^{-1}(x)] = \sum_{\mu \leq \lambda} C_{\mu} D^{\lambda-\mu} h(x) D^\mu f^{-1}(x))
\]

for some positive finite constants \( C_{\mu} \) for \( \mu \leq \lambda \), where \( \mu \) denotes a k-vector of non-negative integer constants.

Let \( A_{1T}(\epsilon) = \{ x \in \mathbb{R}^k : f(x) \geq \epsilon d_{T} \} \), \( A_{2T}(\epsilon) = \{ x \in \mathbb{R}^k : \hat{f}(x) \geq \epsilon d_{T} \} \), and \( A_T = A_{1T}(\epsilon) \cap A_{2T}(\epsilon) \) for some \( \epsilon > 0 \). By Assumption NP2,

\[
\sup_{x \in \rho^k} \sup_{T \geq 1} |D^\mu f(x)| < \infty \quad \forall \mu \leq \lambda .
\]

Using this property, some calculations yield

\[
\sup_{x \in A_T} |D^\mu f^{-1}(x)| = O_p \left( d_T^{(1+|\mu|)} \right) \quad \forall \mu \leq \lambda .
\]

By Lemma A-1 with \( Y_{Tt} = g_T(x) = 1 \) and \( \lambda = \mu \), we have

\[
\sup_{x \in \mathbb{R}^k} |D^\mu \hat{f}(x) - D^\mu \hat{f}(x)| = O_p \left( T^{-\eta/(2\eta+1)} \sigma_{1T}^{-k-|\mu|-1/(2\eta+1)} \right) + O_p \left( \sigma_{2T}^{\omega-|\mu|} \right)
\]

\( \forall \mu \leq \lambda \). Since the fact that the rhs of Theorem II.10(b) is \( O_p(1) \) implies that the rhs above is \( O_p(1) \), we have \( \sup_{x \in \mathbb{R}^k} |D^\mu \hat{f}(x)| = O_p(1) \quad \forall \mu \leq \lambda \). Using the latter result and the assumption that \( \sup_{T \geq 1} \sup_{x \in \mathbb{R}^k} |D^\mu f(x)| < \infty \), some tedious calculations yield
\[
\sup_{x \in A_T} |D^{\mu}[\hat{f}^{-1}(x)] - D^{\mu}[\bar{f}^{-1}(x)]| \\
= \frac{|\mu|}{\sum_j 1 |j| = j} \sum_j O_p \left[ d_T^{-1} |\mu|^{-j} \right] \sup_{x \in \mathbb{R}^k} |D^{\mu} \hat{f}(x) - D^{\mu} \bar{f}(x)|
\]

(B.32)

\(\forall \mu \leq \lambda\), where \(\gamma\) is a \(k\)-vector of non-negative integer constants.

By Assumption NP3, \(\sup_{T \geq 1} \sup_{x \in \mathbb{R}^k} |D^{\mu} h(x)| < \infty\; \forall \mu \leq \lambda\). This property and Lemma A-1 give

\[
\sup_{x \in \mathbb{R}^k} |D^{\lambda-\mu} \hat{h}(x) - D^{\lambda-\mu} \bar{h}(x)| = O_p \left[ T^{-\eta/(2\eta+1)} \sigma_{1T}^{-k} |\lambda-\mu|^{-1/(2\eta+1)} \right] \\
\quad + O_p \left[ \sigma_{2T}^{|\lambda-\mu|} \right] \text{ and}
\]

(B.33)

\[
\sup_{x \in \mathbb{R}^k} |D^{\lambda-\mu} \hat{h}(x)| = O_p(1) \; \forall \mu \leq \lambda.
\]

Now, equations (B.29)–(B.33) combine to yield

\[
\sup_{x \in A_T} |D^{\lambda} \hat{g}(x) - D^{\lambda} \bar{g}(x)| \leq \sum \sigma C \sup_{\mu \leq \lambda} \sup_{x \in A_T} |D^{\lambda-\mu} \hat{h}(x)D^{\mu} [\hat{f}^{-1}(x)] - D^{\lambda-\mu} \bar{h}(x)D^{\mu} [\bar{f}^{-1}(x)]| \\
\leq \sum \sigma C \left[ \sup_{\mu \leq \lambda} |D^{\lambda-\mu} \hat{h}(x)| \sup_{x \in A_T} |D^{\mu} [\hat{f}^{-1}(x)] - D^{\mu} [\bar{f}^{-1}(x)]| \\
\quad + \sup_{x \in A_T} |D^{\mu} [\hat{f}^{-1}(x)]| \sup_{x \in \mathbb{R}^k} |D^{\lambda-\mu} \hat{h}(x) - D^{\lambda-\mu} \bar{h}(x)| \right] \\
= \sum \left[ \frac{|\mu|}{\sum_j 1 |j| = j} O_p \left[ T^{-\eta/(2\eta+1)} \sigma_{1T}^{-k-j-1/(2\eta+1)} d_T^{-2+|\mu|-j} \right] \\
\quad + O_p \left[ \sigma_{2T}^{|\lambda-\mu|} d_T^{-1+|\mu|} \right] \right] \\
\quad + O_p \left[ \sigma_{2T}^{|\lambda-\mu|} d_T^{-2+|\lambda|} \right] \\
\quad + O_p \left[ \sigma_{2T}^{|\lambda-\mu|} d_T^{-2+|\lambda|} \right] \\
\leq O_p \left[ T^{-\eta/(2\eta+1)} \sigma_{1T}^{-k} |\lambda|^{-1/(2\eta+1)} d_T^{-2+|\lambda|} \right] + O_p \left[ \sigma_{2T}^{|\lambda|} d_T^{-2+|\lambda|} \right].
\]

(B.34)
It remains to show that (B.34) holds with \( A_T \) replaced by \( A_{1T}(1) \) or \( A_{2T}(1) \). By (B.31) with \( \mu = 0 \), we have

\[
d_T^{-1} \sup_{x \in \mathbb{R}^k} |\hat{f}(x) - f(x)| = O_p \left[ T^{-\eta/(2\eta+1)} \sigma_{1T}^{-k-1/(2\eta+1)} \right] \\
+ O_p \left[ \sigma_{2T}^\omega d_T^{-1} \right] = o_p(1),
\]

(B.35)

where the second equality holds because the rhs of part (b) of the Theorem is assumed to be \( o_p(1) \). In consequence, \( P(A_{1T}(2\epsilon) \subseteq A_{2T}(\epsilon)) \rightarrow 1 \).

Now, since \( A_T = A_T(\epsilon) \cup A_{1T}(2\epsilon) \cap A_{2T}(\epsilon) \), equation (B.34) holds with \( A_T \) replaced by \( A_{1T}(2\epsilon) \cap A_{2T}(\epsilon) \). And since \( P(A_{1T}(2\epsilon) \subseteq A_{2T}(\epsilon)) \rightarrow 1 \), equation (B.34) holds with \( A_T \) replaced by \( A_{1T}(2\epsilon) \) with probability \( \rightarrow 1 \). Setting \( \epsilon = 1/2 \), this gives the first result of part (b) of the Theorem. An analogous argument with the roles of \( A_{1T}(\cdot) \) and \( A_{2T}(\cdot) \) reversed establishes the second result of part (b) of the Theorem. \( \square \)

Lemma A–1 is implied by the following two lemmas:

**LEMMA A–2:** Under Assumptions NP1–NP5,

\[
\sup_{x \in \mathbb{R}^k} \left| \frac{1}{T} \Sigma_1 Y_{iT} D^\lambda K \left[ \frac{x - X_{iT}}{\sigma} \right] / \sigma_T^k + |\lambda| \right| - \left[ \frac{1}{T} \Sigma_1 Y_{iT} D^\lambda K \left[ \frac{x - X_{iT}}{\sigma} \right] / \sigma_T^k + |\lambda| \right]_{\sigma = \hat{\sigma}_T} = O_p \left[ T^{-\eta/(2\eta+1)} \sigma_{1T}^{-k-1/(2\eta+1)} \right].
\]

**LEMMA A–3:** Under Assumptions NP1–NP4,

\[
\sup_{x \in \mathbb{R}^k} \left| \frac{1}{T} \Sigma_1 \mathbb{E} Y_{iT} D^\lambda K \left[ \frac{x - X_{iT}}{\sigma} \right] / \sigma_T^k + |\lambda| \right| - \frac{1}{T} \Sigma_1 D^\lambda \left[ \mathbb{E} f_T(x) f_T(x) \right] = O \left[ \sigma_T^\omega - |\lambda| \right].
\]

**PROOF OF LEMMA A–2:** Let \( \gamma(\sigma, T) = T^{-\eta/(2\eta+1)} \sigma_T^{-k-1/(2\eta+1)} \) and

\[
\tilde{L}_T(\sigma) = \sup_{x \in \mathbb{R}^k} \left| \frac{1}{T} \Sigma_1 Y_{iT} D^\lambda K \left[ \frac{x - X_{iT}}{\sigma} \right] / \sigma_T^k + |\lambda| \right| - \frac{1}{T} \Sigma_1 \mathbb{E} Y_{iT} D^\lambda K \left[ \frac{x - X_{iT}}{\sigma} \right] / \sigma_T^k + |\lambda|.
\]

(B.36)
Below we show that

\[(B.37) \quad E \sup_{\sigma \geq \sigma_1} \hat{L}_T(\sigma) \leq C \gamma(\sigma_1, T) \forall \sigma_1 \in (0, \sigma_{1B}], \quad \forall T \geq 1, \]

for any given constant \( \sigma_{1B} < \infty \) and some constant \( C < \infty \) that depends on \( \sigma_{1B} \). Using this result and Assumption NP5, we obtain: Given any \( \epsilon > 0 \),

\[
P(\hat{L}_T(\hat{\sigma}_T) > M \gamma(\sigma_{1T}, T)) \leq P(\hat{L}_T(\hat{\sigma}_T) > M \gamma(\sigma_{1T}, T), \hat{\sigma}_T \geq C_1 \sigma_{1T}) + P(\hat{\sigma}_T < C_1 \sigma_{1T})
\]

\[
\leq \mathbb{E}(\hat{L}_T(\hat{\sigma}_T)1(\hat{\sigma}_T \geq C_1 \sigma_{1T})/(M \gamma(\sigma_{1T}, T)) + \epsilon/2
\]

\[(B.38) \]

\[
\leq E \sup_{\sigma \geq C_1 \sigma_{1T}} \hat{L}_T(\sigma)/(M \gamma(\sigma_{1T}, T)) + \epsilon/2
\]

\[
\leq C \gamma(C_1 \sigma_{1T}, T)/(M \gamma(\sigma_{1T}, T)) + \epsilon/2 \leq \epsilon
\]

for \( M \) and \( T \) sufficiently large (using the assumption that \( C_1 \sigma_{1T} \leq \sigma_{1B} \forall T \geq 1 \) for some constant \( \sigma_{1B} < \infty \)). This establishes the result of the Lemma.

It remains to show (B.37). By the Fourier inversion formula, \( D^\lambda K(x) = \int \exp(-ir^T x) \Psi \lambda(r) dr \). Thus,

\[
\hat{L}_T(\sigma) = \sup_{x \in \mathbb{R}^k} \left| \frac{1}{T^1} \left[ Y_{T^1} \int \exp(-ir^T (x - X_{T^1})/\sigma) \Psi \lambda(r) dr \right. \right.
\]

\[
- E Y_{T^1} \left| \int \exp(-ir^T (x - X_{T^1})/\sigma) \Psi \lambda(r) dr \right| / \sigma^{k+1} |\lambda|
\]

\[(B.39) \]

\[
\leq \int \sup_{x \in \mathbb{R}^k} \left| \frac{1}{T^1} \left[ Y_{T^1} \exp(ir^T X_{T^1})/\sigma - E Y_{T^1} \exp(ir^T X_{T^1})/\sigma \right] \right|
\]

\[
\times |\exp(-ir^T x/\sigma)| |\Psi \lambda(r)| \sigma^{k-1} |\lambda| dr
\]

\[
\leq \left| \frac{1}{T^1} \left[ Y_{T^1} \exp(ir^T X_{T^1}) - E Y_{T^1} \exp(ir^T X_{T^1}) \right] \right| |\Psi \lambda(\sigma)| \sigma^{-1} |\lambda| dr
\]

by a change of variables.

Below we show that

\[(B.40) \quad E \left| \frac{1}{T^1} \left[ Y_{T^1} \exp(ir^T X_{T^1}) - E Y_{T^1} \exp(ir^T X_{T^1}) \right] \right|
\]

\[
\leq T^{-\eta/(2\eta+1)} \left[ \mathcal{C}_1 (1 + ||r||) \sigma_{1}^2 \eta/(2\eta+1) + \mathcal{C}_2 \sigma_{1}^{-1/(2\eta+1)} \right]
\]
for all \( r \in \mathbb{R}^k \), \( \sigma_1 > 0 \), and \( T \geq 1 \), for some finite constants \( \bar{C}_1 \) and \( \bar{C}_2 \).

Combining (B.39) and (B.40) gives (B.37):

\[
\mathbb{E} \sup_{\sigma \geq \sigma_1} \hat{L}_T(\sigma) \\
\leq T^{-\eta/(2\eta+1)} \left[ \bar{C}_1(1+\|r\|)\sigma_1^{2\eta/(2\eta+1)} + \bar{C}_2\sigma_1^{-1/(2\eta+1)} \right] \sup_{\sigma \geq \sigma_1} |\psi_\lambda(\sigma r)| |\sigma^{-1}| |\lambda| \, dr \\
(B.41) = T^{-\eta/(2\eta+1)} \left[ \bar{C}_1(1+\|r\|)\sigma_1^{2\eta/(2\eta+1)} + \bar{C}_2\sigma_1^{-1/(2\eta+1)} \right] \sup_{\sigma \geq \sigma_1} |\psi_\lambda(r\sigma/\sigma_1)| |\sigma^{-1}| |\lambda|^{-k} \, dr \\
\leq T^{-\eta/(2\eta+1)} \sigma_1^{-k} \left[ 1/(2\eta+1) \right] \sup_{b \geq 1} |\psi_\lambda(br)| \, dr,
\]

where the equality holds by a change of variables.

To establish (B.40), we use a similar argument to that of Bierens (1983, Pf. of Lemma 4). Let \( Y^S_t = E(Y_{Tt} | \mathcal{F}^+_T, t-s) \) and \( X^S_t = E(X_{Tt} | \mathcal{F}^+_T, t-s) \). Note that \( \{(Y^S_t, X^S_t) : t \leq T, T \geq 1\} \) is strong mixing with mixing numbers \( \{\alpha_s(\ell) : \ell \geq 1\} \) defined by \( \alpha_s(\ell) = 1 \) if \( \ell \leq s \) and \( \alpha_s(\ell) = \alpha(\ell-s) \) if \( \ell \geq s \), where \( \{\alpha(\ell) : \ell \geq 1\} \) is as in Assumption NP1. By a well-known strong mixing inequality (e.g., see Hall and Heyde (1980), Cor. A.2)),

\[
(B.42) \quad |\text{Cov}(Y^S_t \cos(r^S X^S_t), Y^S_u \cos(r^S X^S_u))| \leq 8 \sup_{t \geq 1} \left[ E|Y^S_t|^{2/\beta} \right]^{2/\beta} \alpha_s(|t-u|)^{\beta/(\beta-2)}
\]

for all \( s, t, u \geq 1 \). In consequence,

\[
\text{Var}\left[ \frac{1}{T} \sum_{t=1}^{T} Y^S_t \cos(r^S X^S_t) \right] = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{u=1}^{T} \text{Cov}(Y^S_t \cos(r^S X^S_t), Y^S_u \cos(r^S X^S_u)) \\
(B.43) \leq 8 \sup_{t \geq 1} \left[ E|Y^S_t|^{2/\beta} \right]^{2/\beta} \frac{1}{T} \sum_{t=1}^{T} \sum_{u=1}^{T} \alpha_s(|t-u|)^{\beta/(\beta-2)} \leq \bar{C} \frac{s}{T},
\]

for a constant \( \bar{C} < \infty \) that depends on \( \sup_{t \leq T, T \geq 1} E|Y_{Tt}|^{2/\beta} \) and \( \sum_{t=0}^{\infty} \alpha(t)^{\beta/(\beta-2)} \). The same inequality holds with \( \cos(\cdot) \) replaced by \( \sin(\cdot) \).

Equation (B.43) is used in the following calculation: \( \forall s \geq 1 \),
\[
\begin{align*}
E \left| \frac{1}{T} \sum_{t=1}^{T} \left[ Y_{Tt} \cos(r \cdot X_{Tt}) - EY_{Tt} \cos(r \cdot X_{Tt}) \right] \right| \\
= E \left| \frac{1}{T} \sum_{t=1}^{T} \left[ Y_{Tt} \cos(r \cdot X_{Tt}) - Y_{t}^s \cos(r \cdot X_{Tt}) + [Y_{t}^s \cos(r \cdot X_{Tt}) - Y_{t}^s \cos(r \cdot X_{t}^s)] \right. \right. \\
+ \left[ Y_{t}^s \cos(r \cdot X_{t}^s) - EY_{t}^s \cos(r \cdot X_{t}^s) \right] \left. + [EY_{t}^s \cos(r \cdot X_{t}^s) - EY_{t}^s \cos(r \cdot X_{Tt})] \right) \right| \\
\leq \frac{2 \Sigma T}{T^2} \sup_{t \geq 1} \left[ E|Y_{t}^s|^2 \right]^{1/2} \left[ \frac{1}{T} \sum_{t=1}^{T} \left[ E|\sin(z)(r \cdot X_{Tt} - r \cdot X_{t})|^2 \right] \right]^{1/2} + \left( \tilde{C}_{s} \right)^{1/2} \\
\leq 2 \eta(s) + 2 \sup_{t \leq T, T \geq 1} \left[ E|Y_{t}^s|^2 \right]^{1/2} \left[ \frac{1}{T} \sum_{t=1}^{T} \left[ E|\sin(z)(r \cdot X_{Tt} - r \cdot X_{t})|^2 \right] \right]^{1/2} + \left( \tilde{C}_{s} \right)^{1/2}
\end{align*}
\]

where \( z \) lies between \( r \cdot X_{Tt} \) and \( r \cdot X_{t}^s \) (using the mean value theorem). The same inequality holds with \( \cos(\cdot) \) replaced by \( \sin(\cdot) \). Using the bound \( \eta(s) \leq C^s s^{-\eta} \) for some constant \( C^s < \omega \) (Assumption NP1) and the judicious choice of \( s \) as the integer part of \( T^{1/(2\eta+1)} \sigma_1^{-2/(2\eta+1)} \), one obtains (B.40) from (B.44). \( \square \)

**Proof of Lemma A–3:** Let \( h_t(x) = g_T(x) f_{Tt}(x) \). Then, we have

\[
\begin{align*}
\frac{1}{T} \sum_{t=1}^{T} EY_{Tt} D^\lambda K \left[ \frac{x - X_{Tt}}{\sigma} \right] / \sigma^k + |\lambda| = \int \frac{1}{T} \sum_{t=1}^{T} h_t(z) D^\lambda K \left[ \frac{x - z}{\sigma} \right] dz / \sigma^k + |\lambda| \\
= \int \frac{1}{T} \sum_{t=1}^{T} h_t(x - \sigma z) D^\lambda K(z) dz / \sigma^k + |\lambda| \\
= (-1)^{|\lambda|} \int \frac{1}{T} \sum_{t=1}^{T} h_t(x - \sigma z) K(z) dz / \sigma^k + |\lambda| \\
= \int \frac{1}{T} \sum_{t=1}^{T} D^\lambda h_t(x - \sigma z) K(z) dz,
\end{align*}
\]

where the second equality holds by change of variables and the third by integration by parts. The integration by parts boundary terms are zero by Assumptions NP3 and NP4.

Next, a Taylor series expansion in \( \sigma \) about \( \sigma = 0 \) of order \( \omega - |\lambda| - 1 \) gives
\[
\sup_{x \in \mathbb{R}^k} \left| \int_{\mathbb{R}^1} \frac{1}{T^\lambda_1} D^\lambda h_i(x-\sigma z)K(z)dz - \frac{1}{T^\lambda_1} D^\lambda h_i(x) \right|
\]

\[
= \sup_{x \in \mathbb{R}^k} \left| \int \left[ \frac{\omega}{\Sigma} \sum_{r=1}^{\sigma^r} \frac{d^r}{d\omega^r} \left[ \frac{1}{T^\lambda_1} D^\lambda h_i(x-\sigma z) \right] \right]_{\sigma=0} \right| K(z)dz
\]

\[
\geq \sup_{x \in \mathbb{R}^k} \left| \frac{\omega}{\Sigma} \sum_{r=1}^{\sigma^r} \frac{d^r}{d\omega^r} \left[ \frac{1}{T^\lambda_1} D^\lambda h_i(x-\sigma z) \right] \right| \left| (-z)^\mu K(z)dz \right|
\]

\[
= 0 + O(\omega^{-1} \lambda) ,
\]

where \( \sigma \) lies between 0 and \( \sigma \). The first equality uses \( \int K(x)dx = 1 \). The second equality uses \( \sup_{x \in \mathbb{R}^k, T \geq 1} \left| \frac{1}{T^\lambda_1} D^\lambda \mu(g_T(x)T_1(x)) \right| < \infty \) and Assumption NP4(a). The result of the Lemma now follows. \( \Box \)

**PROOF OF COROLLARY II.1:** First we consider the case where \( \hat{G}_T \subset \left\{ \frac{1}{T^\lambda_1} f_{T_1}(x) \geq d_T \right\} \) and \( \hat{G}_T \subset \left\{ \frac{1}{T^\lambda_1} f_{T_1}(x) \geq d_{2T} \right\} \). Let LHS denote the left-hand side expression of Corollary II.1. Using Assumptions NP6 and NP7 and Theorem II.10(b), we have
LHS ≤ sup_{s ≤ N, N ≥ 1} \left[ \int_{\mathbb{R}} \{1(x \not\in \hat{G}_T) \mid D^\lambda \hat{g}(x) - D^\lambda g_T(x) \mid^Q dP_N(x) \}^{1/Q} \right] 
+ sup_{s ≤ N, N ≥ 1} \left[ \int_{\mathbb{R}} \{1(x \not\in \hat{G}_T) \mid B + \mid D^\lambda \hat{g}_T(x) \mid^Q dP_N(x) \} \right]^{1/Q} 
≤ sup_{s ≤ N, N ≥ 1} \left[ \int_{\mathbb{R}} \{1 \mid \frac{1}{T} \Sigma_1^T f_t(x) \mid \ge d_T \} \mid D^\lambda \hat{g}(x) - D^\lambda g_T(x) \mid^Q dP_N(x) \right]^{1/Q} 
+ \bar{B} sup_{s ≤ N, N ≥ 1} \left[ \int_{\mathbb{R}} \{1 \mid \frac{1}{T} \Sigma_1^T f_t(x) \mid \le d_2T \} f_N(x)dx \right]^{1/Q} 
(\text{B.47}) 

\le sup_{x \in \{1 \mid \frac{1}{T} \Sigma_1^T f_t(x) \mid \ge d_T \}} \mid D^\lambda \hat{g}(x) - D^\lambda g_T(x) \mid 
+ \bar{B} sup_{s ≤ N, N ≥ 1} \left[ \int_{\mathbb{R}} \{1 \mid \frac{1}{T} \Sigma_1^T f_t(x) \mid \le d_2T \} f_N(x)dx \right]^{1/Q} 
= O_P \left[ T^{-\eta/(2\eta+1)} \sigma_{T}^{-k-1} \mid \lambda \mid-1/(2\eta+1) d_T^{-2-1} \mid \lambda \mid \right] 
+ O_P \left[ \sigma_{T}^{-k-1} \mid \lambda \mid-1/(2\eta+1) d_T^{-2-1} \mid \lambda \mid \right] 
+ O_P \left[ d_2T^{-2-1} \mid \lambda \mid \right] 

Next, suppose \( \hat{G}_T \subset \{ \hat{f}(x) \ge d_T \} \) and \( \hat{G}_T^c \supset \{ \hat{f}(x) \ge d_2T \} \) with probability → 1. By Theorem II.10(a) and the assumptions that \( d_T = O_p(1) \) and the first two terms of the rhs of the Corollary are \( O_p(1) \), we have \( d_T^{-1} \sup_{x \in \mathbb{R}} \mid \hat{f}(x) - 1 \quad \frac{1}{T} \Sigma_1^T f_t(x) \mid = O_p(1) \). In consequence, \( \hat{G}_T \subset \{ \frac{1}{T} \Sigma_1^T f_t(x) \mid \ge d_T/2 \} \) and \( \hat{G}_T^c \supset \{ \frac{1}{T} \Sigma_1^T f_t(x) \mid \ge 2d_2T \} \) with probability → 1 (where the latter uses the assumption that \( d_2T^{-1} \le d_T^{-1} \)). The proof above with \( (d_T, d_2T) \) replaced by \( (d_T/2, 2d_2T) \) now gives the desired result. □

The proof of Theorem II.11 uses the following lemma:

**Lemma A-4**: (a) Under Assumption NP8,

\[
\sup_{x \in \mathbb{R}} \left| \frac{1}{T} \Sigma_1^T f_t(x) D^k K \left[ \frac{x - X_tT}{\sigma_T} \right] / \sigma_T^k + \mid \lambda \mid - \frac{1}{T} \Sigma_1^T Y_t D^k \left[ \frac{x - X_tT}{\sigma_T} \right] / \sigma_T^k + \mid \lambda \mid \right|
= O_p \left[ T^{-1/2} \sigma_T^{-k-1} \mid \lambda \mid \right].
\]
(b) Under Assumptions NP8–NP10,

\[
\sup_{x \in \mathbb{R}^k} \left| \frac{1}{T} \Sigma_1^{T} Y T T \bar{D}^\lambda K \left[ \frac{x - \bar{x}}{\bar{\sigma}_T} \right] / \bar{\sigma}_T^{k+1} + |\lambda| \right| - \frac{1}{T} \Sigma_1^{T} Y T T \bar{D}^\lambda K \left[ \frac{x - \bar{x}_T}{\bar{\sigma}_T} \right] / \bar{\sigma}_T^{k+1} + |\lambda| \right|
\]

\[= O_p \left( T^{-1/2} \bar{\sigma}_T^{-k-1} |\lambda|^{-1} \right). \]

**Proof of Theorem II.11:** By Lemma A–4(a),

\[
\sup_{\{x: f(x) \geq d_T \}} |D^\lambda g_T(x) - D^\lambda g(x)|
\]

\[= \sup_{\{x: f(x) \geq d_T \}} \left| \frac{1}{T} \Sigma_1^{T} Y T T \bar{D}^\lambda K \left[ \frac{x - \bar{x}_T}{\bar{\sigma}_T} \right] / \bar{\sigma}_T^{k+1} + |\lambda| \right| / |\bar{f}(x)| \]

\[= O_p \left( T^{-1/2} \bar{\sigma}_T^{-k-1} |\lambda|^{-1} \right). \]

Since \( d_T^{-1} \sup_{x \in \mathbb{R}^k} |\bar{f}(x) - \frac{1}{T} \Sigma_1^{T} f T T_T(x)| = o_p(1) \), the same result holds with \( \bar{f}(x) \) replaced by \( \frac{1}{T} \Sigma_1^{T} f T T_T(x) \) in the set above. These results and Theorem II.10(b) combine to establish part (a) of the Theorem.

Part (a) of the Theorem and the proof of Corollary II.1 establish part (b) of the Theorem.

Lemma A–4(b) with \( Y T T_T(\alpha_1) = Y T T_T = 1 \) and \( \hat{\alpha}_1 = \alpha_{10} \) gives

\[
\sup_{x \in \mathbb{R}^k} |D^\lambda \bar{f}(x) - D^\lambda \bar{f}(x)| = O_p \left[ T^{-1/2} \bar{\sigma}_T^{-k-1} |\lambda|^{-1} \right]. \]

This result and Theorem II.10(a) combine to establish part (c) of the Theorem.

Let \( B_{1T}(\epsilon) = \{ x : \bar{f}(x) \geq \epsilon d_T \} \), \( B_{2T}(\epsilon) = \{ x : \bar{f}(x) \geq \epsilon d_T \} \), and \( B_T = B_T(\epsilon) = B_{1T}(\epsilon) \cap B_{2T}(\epsilon) \) for some \( \epsilon > 0 \). By Lemma A–4(a), Lemma A–4(b), and Theorem II.11(b), we have
\[
\sup_{x \in B_T} |D^\lambda \hat{g}(x) - D^\lambda \hat{g}(x)| \\
\leq \sup_{x \in B_T} \left| \frac{1}{T} \Sigma_1 T_T D^\lambda K \left[ \frac{x - X_T T_T}{\hat{\sigma}_T} \right] / \hat{\sigma}_T^{k+|\lambda|} - \frac{1}{T} \Sigma_1 T_T D^\lambda K \left[ \frac{x - X_T T_T}{\hat{\sigma}_T} \right] / \hat{\sigma}_T^{k+|\lambda|} \right| / |f(x)| \\
+ \sup_{x \in B_T} \left| \frac{1}{T} \Sigma_1 T_T D^\lambda K \left[ \frac{x - X_T T_T}{\hat{\sigma}_T} \right] / \hat{\sigma}_T^{k+|\lambda|} \right| |\hat{f}(x) - \hat{f}(x)| / |\hat{f}(x) \hat{f}(x)| \\
= O_p \left( T^{-1/2} \sigma_1 T^{-1} \right) + O_p \left( T^{-1/2} \sigma_1 T^{-1} d_T^{-2} \right) .
\]

By (B.49) and the assumption that the rhs of part (d) of the Theorem is o_p(1), we have

\[
\sup_{x \in R^k} |f(x) - \hat{f}(x)| = O_p \left( T^{-1/2} \sigma_1 T^{-1} d_T^{-2} \right) = o_p(1) .
\]

Similarly, \( d_T^{-1} \sup_{x \in R^k} |f(x) - \frac{1}{T} \Sigma_1 T_T f_T(x)| = o_p(1) \) using Theorem II.10(a). Given these results, the same argument as in the proof of Theorem II.10(b) shows that (B.50) holds with probability \( \to 1 \) when \( B_T \) is replaced by \( B_1 T(2\epsilon) \) or by \( \{x : \frac{1}{T} \Sigma_1 T_T f_T(x) \geq 2\epsilon d_T\} \).

Taking \( \epsilon = 1/2 \) and combining these results with the result of Theorem II.10(b) establishes part (d) of the Theorem.

Part (d) of the Theorem and the proof of Corollary II.1 establish part (e) of the Theorem. \( \Box \)

PROOF OF LEMMA A-4: For part (a), a mean value expansion about \( \alpha_{10} \) gives

\[
\sup_{x \in R^k} \left| \frac{1}{T} \Sigma_1 T_T f_T(\hat{\alpha}_1) D^\lambda K \left[ \frac{x - X_T T_T}{\hat{\sigma}_T} \right] / \hat{\sigma}_T^{k+|\lambda|} - \frac{1}{T} \Sigma_1 T_T D^\lambda K \left[ \frac{x - X_T T_T}{\hat{\sigma}_T} \right] / \hat{\sigma}_T^{k+|\lambda|} \right| \\
= \sup_{x \in R^k} \left| T^{-1/2} \sigma_1 T^{-1} \sqrt{T} (\hat{\alpha}_1 - \alpha_{10}) \right| \frac{1}{T} \Sigma_1 T_T \frac{\partial}{\partial \alpha_1} Y_T T_T (\alpha_{10}) D^\lambda K \left[ \frac{x - X_T T_T}{\hat{\sigma}_T} \right] ,
\]

where \( \alpha_{x1} \) lies on the line segment joining \( \hat{\alpha}_1 \) and \( \alpha_{10} \). By Assumption NP8 and Markov's inequality,
\[
\sup_{x \in \mathbb{R}^k} \left\| \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \alpha_1} Y_{Tt}(\alpha) D^\lambda K \left( \frac{x - X_{Tt}}{\sigma_T} \right) \right\|
\]

(\text{B.53})
\[
\leq \frac{1}{T} \sum_{t=1}^{T} \sup_{\alpha_1 \in A_1} \left\| \frac{\partial}{\partial \alpha_1} Y_{Tt}(\alpha) \right\| \sup_{x \in \mathbb{R}^k} |D^\lambda K(x)| = O_p(1).
\]

Equations (\text{B.52}) and (\text{B.53}) combine to establish part (a) of the Lemma.

For part (b), a mean value expansion about \((\alpha_{10}', \alpha_{20}')\) gives
\[
\sup_{x \in \mathbb{R}^k} \left\| \frac{1}{T} \sum_{t=1}^{T} Y_{Tt}(\alpha) D^\lambda K \left( \frac{x - X_{Tt}(\alpha)}{\hat{\sigma}_T} \right) \right\|
\]

(\text{B.54})
\[
= \sup_{x \in \mathbb{R}^k} \left\| \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \alpha_1} Y_{Tt}(\alpha) D^\lambda K \left( \frac{x - X_{Tt}(\alpha)}{\hat{\sigma}_T} \right) \right\|
\]
\[
+ \sqrt{T}(\hat{\alpha}_2 - \alpha_{20}'), \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \alpha_2} Y_{Tt}(\alpha) D^\lambda K \left( \frac{x - X_{Tt}(\alpha)}{\hat{\sigma}_T} \right)
\]
\[
+ \sqrt{T}(\hat{\alpha}_2 - \alpha_{20}'), \frac{1}{T} \sum_{t=1}^{T} Y_{Tt}(\alpha) \sum_{j=1}^{k} D^{\lambda+e_j} K \left( \frac{x - X_{Tt}(\alpha)}{\hat{\sigma}_T} \right)
\]
\[
\cdot \left[ -\frac{\partial}{\partial \alpha_2} X_{Ttj}(\alpha) / \hat{\sigma}_T \right]
\]

where \((\alpha_{1}', \alpha_{2}')\) lies on the line segment joining \((\hat{\alpha}_1', \hat{\alpha}_2')\) and \((\alpha_{10}', \alpha_{20}')\) and \(X_{Ttj}(\alpha)\) denotes the \(j\)-th element of \(X_{Tt}(\alpha)\). By Assumptions NP9 and NP10 and Markov's inequality,
\[
\sup_{x \in \mathbb{R}^k} \left\| \frac{1}{T} \sum_{t=1}^{T} Y_{Tt}(\alpha) \sum_{j=1}^{k} D^{\lambda+e_j} K \left( \frac{x - X_{Tt}(\alpha)}{\hat{\sigma}_T} \right) \right\|
\]

(\text{B.55})
\[
\leq \frac{1}{T} \sum_{t=1}^{T} \sup_{\alpha_1 \in A_1, \alpha_2 \in A_2} \left\| Y_{Tt}(\alpha) \frac{\partial}{\partial \alpha_2} X_{Ttj}(\alpha) \right\| \sup_{x \in \mathbb{R}^k} |D^{\lambda+e_j} K(x)| = O_p(1).
\]

Equations (\text{B.53}), (\text{B.54}), and (\text{B.55}) combine to establish part (b) of the Lemma. □
PROOF OF THEOREM II.12: The proof is the same as that of Theorem II.10 with $d_T = d \quad \forall T \geq 1, \quad \eta = \omega, \quad$ and reference to Lemma A–1 replaced by reference to the following result. Under Assumptions NP1*–NP5*,

$$ \sup_{(\alpha_1, \alpha_2, x) \in A_1 \times A_2 \times \mathcal{X}_B} \left| \frac{1}{T} \sum_{i=1}^{T} Y_{T_{i}}(\alpha_1) D^\lambda K \left[ \frac{x - X_{T_{i}}(\alpha_2)}{\hat{\sigma}_T} \right] / \hat{\sigma}_T^{k+|\lambda|} \right| $$

(B.56)

$$ - \frac{1}{T} \sum_{i=1}^{T} D^\lambda [g_T(\alpha_1, \alpha_2, x) f_{T_{i}}(\alpha_2, x)] = O \left[ T^{-1/2} \sigma_{1T}^{-k-|\lambda|} \right] + O_p \left[ o_{2T}^{\omega-|\lambda|} \right]. $$

This result follows from analogues of Lemmas A–2 and A–3. The analogue of Lemma A–3 states that under Assumptions NP2*–NP4*,

$$ \sup_{(\alpha_1, \alpha_2, x) \in A_1 \times A_2 \times \mathcal{X}_B} \left| \frac{1}{T} \sum_{i=1}^{T} E Y_{T_{i}}(\alpha_1) D^\lambda K \left[ \frac{x - X_{T_{i}}(\alpha_2)}{\sigma} \right] / \sigma^{k+|\lambda|} \right|_{\sigma = \hat{\sigma}_T} $$

(B.57)

$$ - \frac{1}{T} \sum_{i=1}^{T} D^\lambda [g_T(\alpha_1, \alpha_2, x) f_{T_{i}}(\alpha_2, x)] = O \left[ \sigma_{T}^{\omega-|\lambda|} \right]. $$

The proof of this result is a minor variation of the proof of Lemma A–3.

The analogue of Lemma A–2 states that under Assumptions NP1*–NP5*,

$$ \sup_{(\alpha_1, \alpha_2, x) \in A_1 \times A_2 \times \mathcal{X}_B} \left| \frac{1}{T} \sum_{i=1}^{T} Y_{T_{i}}(\alpha_1) D^\lambda K \left[ \frac{x - X_{T_{i}}(\alpha_2)}{\hat{\sigma}_T} \right] / \hat{\sigma}_T^{k+|\lambda|} \right| $$

(B.58)

$$ - \left[ \frac{1}{T} \sum_{i=1}^{T} E Y_{T_{i}}(\alpha_1) D^\lambda K \left[ \frac{x - X_{T_{i}}(\alpha_2)}{\sigma} \right] / \sigma^{k+|\lambda|} \right]_{\sigma = \hat{\sigma}_T} $$

$$ = O_p \left[ T^{-1/2} \sigma_{1T}^{-k-|\lambda|} \right]. $$

This result holds because the left-hand side is less than or equal to

$$ \frac{1}{\sqrt{TC}} \sigma_{1T}^{k+|\lambda|} \sup_{(\alpha_1, \alpha_2, \sigma, x) \in A_1 \times A_2 \times \Sigma \times \mathcal{X}_B} |\xi_T(\lambda, \alpha_1, \alpha_2, \sigma, x)| = O_p \left[ T^{-1/2} \sigma_{1T}^{-k-|\lambda|} \right] $$

using Assumption NP5*(b), where the equality holds by Assumption NP3*(e).
Note that (B.56) is applied several times in the proof of Theorem II.12 with $\lambda$ replaced by $\mu$ for different $\mu \leq \lambda$, with $Y_{T_0}(\mu_1)$ as given, and with $Y_{T_0}(\mu_1)$ replaced by 1. This explains various conditions in Assumptions NP1*-NP5* that must hold $\forall \mu \leq \lambda$ and it explains the need for Assumption NP1*. \(\square\)
FOOTNOTES

1 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 Rather than considering stochastic equicontinuity of \( \{ \nu_T(\cdot) : T \geq 1 \} \) at a single point \( \tau_0 \in \mathcal{T} \), which is required in Assumptions 2(e) and 2*(e) of ASEM:I, we consider here stochastic equicontinuity uniformly over all points \( \tau \in \mathcal{T} \). The latter is more general, is obtained at little additional cost in terms of assumptions relative to the former, and is useful in other contexts (e.g., for establishing uniform laws of large numbers and functional central limit theorems).

3 To use the results of this paper, the function \( m_{T_0}(\cdot, \cdot, \cdot) \) of ASEM:I cannot depend on \( T \) or \( t \) and the rvs \( \{ W_{T_0} \} \) of ASEM:I must have dimension \( k \) that does not depend on \( T \) or \( t \). For some stochastic equicontinuity results where they may, see Pollard (1989, 1990).

4 As in ASEM:I, the pseudo–metric \( \rho_T(\cdot, \cdot) \) is defined here using a dummy variable \( N \) (rather than \( T \)) to avoid confusion when we consider objects such as \( \lim_{T \to \infty} \rho_T(\tau, \tau_0) \). Note that \( \rho_T(\cdot, \cdot) \) is taken to be independent of the sample size \( T \).

5 If need be, the bound in (4.3) can be replaced by \( C|\log \delta|^{-\lambda} \) for arbitrary constants \( C \in (1, \infty) \) and \( \lambda \geq 1 \) and Theorem II.5 still goes through.

6 The partial derivative \( \partial^q m(w, \tau) \) of \( m(w, \tau) \) can be defined in the usual sense or in a weaker sense (see Stein (1970, p. 180)). The latter allows the partial derivatives of order \( q-1 \) to be Lipschitz (with exponent one) rather than partially differentiable in the usual sense while maintaining \( \|m(\cdot, \tau)\|_{q,2,\mathcal{W}} < \infty \).

7 Let \( \mathcal{W} \) be an open set in \( \mathbb{R}^k \) and let \( \partial \mathcal{W} \) be its boundary. By definition, \( \partial \mathcal{W} \) is minimally smooth if there exists an \( \epsilon > 0 \), an integer \( N \), an \( M > 0 \), and a sequence \( U_1, U_2, \ldots \) of open sets in \( \mathbb{R}^k \) such that (a) if \( \omega \in \partial \mathcal{W} \), then \( S(\omega, \epsilon) \subset U_j \) for some \( j \), where \( S(\omega, \epsilon) \) is the sphere centered at \( \omega \) with radius \( \epsilon \), (b) no point of \( \mathbb{R}^k \) is contained in more than \( N \) of the \( U_j \)'s, and (c) \( \forall j \geq 1, \exists \) a special Lipschitz domain \( D_j \subset \mathbb{R}^k \), whose bound does not exceed \( M \), such that \( U_j \cap \mathcal{W} = U_j \cap D_j \). A set \( D \subset \mathbb{R}^k \) is a special Lipschitz domain if \( D = \{(x,y) \in \mathbb{R}^k : y > \varphi(x) \} \) for some function \( \varphi : \mathbb{R}^{k-1} \to \mathbb{R} \) that satisfies \( |\varphi(x) - \varphi(x')| < M|x - x'| \ \forall x, x' \in \mathbb{R}^{k-1} \). The smallest \( M \) for which the Lipschitz condition holds is called the bound of the special Lipschitz domain.
REFERENCES


