COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1111R

A CONDITIONAL KOLMOGOROV TEST

Donald W. K. Andrews

September 1995
Revised April 1996

A CONDITIONAL KOLMOGOROV TEST

Donald W. K. Andrews[1]

Cowles Foundation for Research in Economics
Yale University

September 1995

Revised: April 1996

ABSTRACT

This paper introduces a conditional Kolmogorov test of model specification for parametric models with covariates (regressors). The test is an extension of the Kolmogorov test of goodness-of-fit for distribution functions. The test is shown to have power against $1/\sqrt{n}$ local alternatives and all fixed alternatives to the null hypothesis. A parametric bootstrap procedure is used to obtain critical values for the test.

Keywords: Bootstrap, consistent test, parametric model, specification test.

JEL Classification Number: C12.

1. INTRODUCTION

This paper introduces a specification test for parametric models for independent observations. The null hypothesis of interest is that the parametric model is correctly specified. The alternative hypothesis is that the parametric model is incorrectly specified. The parametric model we consider is one that specifies the conditional distribution of a vector $Y_i \in R^V$ of response variables given a vector $X_i \in R^K$ of covariates (regressors). The distribution of the covariates is not specified by the parametric model. Many models used in micro-econometric and biometric applications are of this type. For example, see Maddala (1983) and McCullagh and Nelder (1983) for numerous models of this sort.

The test we consider is a generalization of the Kolmogorov (K) test (sometimes called the Kolmogorov–Smirnov test) of goodness-of-fit. We call the test a *conditional Kolmogorov* (CK) test, because it is designed for parametric models for the conditional distribution of Y_i given X_i. The CK test has the following attributes. It is (i) consistent against all alternatives to the null hypothesis H_0, (ii) powerful against $1/\sqrt{n}$ local alternatives to H_0, and (iii) not dependent on any smoothing parameters.

The antecedents to the CK test are numerous. Kolmogorov (1933) introduced the K test for testing whether an independent and identically distributed (iid) sample of random variables (rv's) comes from a given continuous univariate distribution function (df) F. Smirnov (1939) extended the K test to two sample problems for which the null hypothesis is that the two samples are drawn from the same continuous univariate distribution. Doob (1949) provided an illuminating heuristic proof of the asymptotic null distributions of Kolmogorov's and Smirnov's test statistics based on the weak convergence of the empirical df. Donsker (1952) closed a gap in Doob's proof by establishing the first empirical process central limit theorem (CLT), often referred to as Donsker's Theorem. Durbin (1973a, b) established the asymptotic null distribution of K tests for parametric families of univariate continuous distributions in which a parameter vector is estimated. Pollard (1984) extended Durbin's results to univariate distributions that are not necessarily continuous. Beran and Millar (1989) considered K tests for parametric families of multivariate distributions

— not necessarily continuous — in which a parameter vector is estimated. They introduced a bootstrap method for obtaining critical values.

The present paper extends the results above by considering a K test for parametric models that specify a parametric family for the conditional distribution of a (possibly multivariate, possibly non-continuous) response variable given a (possibly multivariate, possibly non-continuous) covariate.

The testing problem considered here is also considered by Zheng (1993, 1994) and Stinchcombe and White (1993). None of their tests is of the Kolmogorov type. Zheng's tests are consistent against all alternatives to the null, but are not powerful against $1/\sqrt{n}$ local alternatives and are dependent on smoothing parameters. Stinchcombe and White's tests do not suffer from the latter problems, but they rely on either the law of the iterated logarithm to obtain an upper bound on the asymptotic critical value, which seems overly conservative, or the simulation of the supremum or integral of a Gaussian process indexed by a multidimensional parameter, which can be difficult to carry out. (The use of the nonparametric bootstrap, which is mentioned by Stinchcombe and White as another way of obtaining critical values, is problematic because it yields tests with no power.) The CK test avoids the above problems by using a parametric bootstrap procedure to obtain critical values.

The bulk of the econometrics literature on consistent tests, e.g., see Bierens (1990), considers tests of the specification of a parametric regression function. The present paper differs from this literature in that it considers tests of the specification of a wide variety of conditional parametric models.

The remainder of the paper is organized as follows. The conditional Kolmogorov test is introduced in Section 2. The asymptotic null distribution of the test statistic is established in Section 3. A parametric bootstrap procedure for obtaining critical values and p–values is introduced and justified asymptotically in Section 4. Consistency of the test is established in Section 5. The power of the test against $1/\sqrt{n}$ local alternatives and its asymptotic local unbiasedness are shown in Section 6. An Appendix contains proofs of results stated in the text.

All limits below are as $n \to \infty$.

2. DEFINITION OF THE CONDITIONAL KOLMOGOROV TEST

The observed sample consists of the n rv's $\{Z_i : i \leq n\}$, where $Z_i = (Y_i', X_i')' \in R^{V+K}$. We assume the sample comes from a sequence of rv's that satisfies:

Assumption D: *$\{Z_i : i \geq 1\}$ are* iid *with conditional* df $H(\cdot|X_i)$ *of Y_i given X_i and marginal* df $G(\cdot)$ *of X_i.*

(See the Appendix for results that are applicable to the case of independent non-identically distributed (inid) rv's.)

The parametric model considered here consists of a parametric family of conditional distributions of the response variable Y_i given the covariate X_i. In particular, the parametric family is

$$\{f(y|x, \theta) : \theta \in \Theta\} , \tag{2.1}$$

where $f(y|x, \theta)$ is a density with respect to a σ-finite measure μ and $\Theta \subset R^L$ is the parameter space. Since μ need not be Lebesgue measure, Y_i may be discrete, continuous, or mixed.

The parametric conditional df of Y_i given $X_i = x$ is denoted

$$F(y|x, \theta) = \int (y^* \leq y) f(y^*|x, \theta) d\mu(y^*) \tag{2.2}$$

for $y \in R^V$, $x \in R^K$, and $\theta \in \Theta$. Here, $(y^* \leq y)$ denotes the indicator function of the event $y^* \leq y$. That is, $(y^* \leq y) = 1$ if $y^* \leq y$ and $(y^* \leq y) = 0$ otherwise.

The null hypothesis of interest is

$$H_0 : H(\cdot|\cdot) = F(\cdot|\cdot, \theta) \ \text{ for some } \ \theta \in \Theta . \tag{2.3}$$

The alternative hypothesis H_1 of interest is the negation of H_0.

We now define the CK test statistic. Let $\widehat{H}_n(z)$ denote the empirical df of $\{Z_i : i \leq n\}$:

$$\widehat{H}_n(z) = \tfrac{1}{n} \textstyle\sum_{i=1}^n (Z_i \leq z) \ \text{ for } \ z \in R^{V+K} . \tag{2.4}$$

Let $\widehat{G}_n(\cdot)$ denote the empirical df of $\{X_i : i \leq n\}$:

$$\widehat{G}_n(x) = \frac{1}{n}\sum_{i=1}^n (X_i \leq x) \ \text{ for } \ x \in R^K \ . \tag{2.5}$$

Let $\widehat{\theta}$ be an estimator of θ. When H_0 is true, we let θ_0 denote the true value of θ. Below we assume that under H_0, $\widehat{\theta}$ is a $\sqrt{n}$–consistent estimator of θ_0.

Let $\widehat{F}_n(z, \theta)$ denote the semi-parametric/semi-empirical df of $\{Z_i : i \leq n\}$ based on the parametric conditional df $F(\cdot|\cdot,\theta)$ and the empirical df $\widehat{G}_n(\cdot)$:

$$\widehat{F}_n(z, \theta) = \frac{1}{n}\sum_{i=1}^n F(y|X_i, \theta)(X_i \leq x) \ \text{ for } \ z = (y', x')' \in R^{V+K} \ . \tag{2.6}$$

Note that under the null hypothesis $E(\widehat{H}_n(z)|X) = \widehat{F}_n(z, \theta_0) \ \forall z \in R^{V+K}$, where $E(\cdot|X)$ denotes conditional expectation given $\{X_i : i \geq 1\}$. On the other hand, under the alternative hypothesis, $E(\widehat{H}_n(z)|X) \neq \widehat{F}_n(z,\theta)$ for some $z \in R^{V+K}$, for n large, for all $\theta \in \Theta$ (see Section 5 below). Hence, one can construct a model specification test based on the difference between $\widehat{H}_n(\cdot)$ and $\widehat{F}_n(\cdot,\widehat{\theta})$.

Define the CK test statistic as

$$\begin{aligned} CK_n &= \sqrt{n}\max_{j \leq n} |\widehat{H}_n(Z_j) - \widehat{F}_n(Z_j, \widehat{\theta})| \\ &= \max_{j \leq n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n [(Y_i \leq Y_j) - F(Y_j|X_i, \widehat{\theta})](X_i \leq X_j) \right| \ . \end{aligned} \tag{2.7}$$

Note that CK_n differs from a traditional K test statistic in two ways. First, it is not based on the difference between an empirical df and a parametric df, as K statistics are. Rather, it is based on the difference between the empirical df $\widehat{H}_n(\cdot)$ and the semi-parametric/semi-empirical df $\widehat{F}_n(\cdot, \widehat{\theta})$. The reason, of course, is that the parametric model does not specify the df of Z_i up to an unknown parameter. It only specifies the *conditional* df of Y_i given X_i up to an unknown parameter.

Second, the CK_n statistic is not defined by taking the supremum over all points z in R^{V+K}, as K statistics are. Rather, CK_n is defined by taking the maximum over points z in the sample $\{Z_i : i \leq n\}$. The reason for defining CK_n in this way is computational. Maximizing a statistic over an unbounded high dimensional space is very difficult and time consuming. Maximizing over $z \in \{Z_i : i \leq n\}$ is straightforward and not very time consuming unless n is quite large.

The asymptotic null distribution of CK_n is determined in Section 3 below. It turns out to depend on θ_0 as well as the df $G(\cdot)$ of the covariates. In consequence, we obtain critical values and p–values for the CK_n statistic by a parametric bootstrap procedure. This procedure and its properties are discussed in Section 4 below. For now, let $c_{\alpha nB}(\widehat{\theta})$ denote the bootstrap critical value for significance level $\alpha \in (0,1)$, where B denotes the number of bootstrap repetitions. The CK_n test rejects H_0 if

$$CK_n > c_{\alpha nB}(\widehat{\theta}) \ . \tag{2.8}$$

As defined, the CK_n test depends on the signs of the elements of the rv's $Z_i = (Y_i', X_i')'$. That is, if one changes the j-th element, Z_{ij}, of Z_i to $-Z_{ij}$ for all $i = 1, ..., n$, then the value of the CK test statistic changes in general. To obtain a sign invariant CK test statistic, one can define CK_n as in (2.7) for each of the possible sign permutations of the rv's $\{Z_i : i \leq n\}$ and define the sign invariant CK test statistic to be the maximum of these statistics. The resultant CK_n test statistic is sign invariant, though it is more burdensome computationally than the statistic of (2.7). The sign invariant CK statistic can be analyzed in exactly the same way as the statistic of (2.7) with some increase in notational complexity. Its asymptotic properties are analogous to those established below for the statistic of (2.7). For simplicity, then, we only consider formally the statistic of (2.7) below, though it may be desirable to use the sign invariant statistic in many applications.

We note that the CK_n statistic of (2.7) depends on the ordering of the values that any given element of Z_i can take on. In some cases, this is undesirable. For example, in a trinary discrete response model, suppose Y_i equals 0, 1, or 2 if individual i takes the car, bus, or train respectively. Then, the arbitrary ordering of (car, bus, train) as (0, 1, 2) affects the value of the CK statistic. This undesirable feature can be circumvented by taking the response variable Y_i to be multivariate with Y_i equal to $(1, 0, 0)'$, $(0, 1, 0)'$, or $(0, 0, 1)'$ depending on whether individual i takes the car, bus, or train respectively. This method can be applied more generally.

The results given below can be adapted to establish analogous results for "conditional Cramer–von Mises tests." For brevity, we do not do so in this paper.

3. THE ASYMPTOTIC NULL DISTRIBUTION OF THE CK_n TEST STATISTIC

In this section, we determine the asymptotic null distribution of the CK_n test statistic. First, we introduce some notation. Let

$$\nu_n(z,\theta) = \sqrt{n}(\widehat{H}_n(z) - \widehat{F}_n(z,\theta)) \quad \text{for} \quad z \in R^{V+K} \ . \tag{3.1}$$

Note that if Y_i given X_i has df $F(\cdot|\cdot,\theta)$, then $\nu_n(\cdot,\theta)$ is a conditional empirical process, as defined in Andrews (1988), for the iid rv's $\{Z_i : i \geq 1\}$, since $E(\widehat{H}_n(z)|X) = \widehat{F}_n(z,\theta)$.

For notational convenience, we switch between variables Z_i and (Y_i, X_i), z and (y,x), z^* and (y^*, x^*), z_1 and (y_1, x_1), etc., without comment. Thus, $\nu_n(z,\theta) = \nu_n(y,x,\theta)$, etc. Often, we use y^* and x^* as dummy variables of integration.

Below we establish asymptotic results that hold conditional on $\{X_i : i \geq 1\}$ with $\{X_i : i \geq 1\}$ probability one. Such results are stronger than the corresponding unconditional asymptotic results.[2] They are needed to justify the parametric bootstrap procedure that is introduced in Section 4 below.

For brevity, we let "cond'l on X wp1" abbreviate "conditional on $\{X_i : i \geq 1\}$ with $\{X_i : i \geq 1\}$ probability one." We also let "wp1" abbreviate "with $\{X_i : i \geq 1\}$ probability one." We let $P(\cdot|X)$ denote probability conditional on $\{X_i : i \geq 1\}$.

The estimator $\widehat{\theta}$ is assumed below to have a linear expansion of the form

$$\sqrt{n}(\widehat{\theta} - \theta_0) = \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n D_0 \psi(Z_i, \theta_0) + o_p(1) \quad \text{cond'l on } X \text{ wp1} \tag{3.2}$$

when the sample is generated by the null df $F(\cdot|\cdot, \theta_0)$, where D_0 is a non-random $L \times L$ matrix, $\psi(z,\theta)$ is a measurable function from $R^{V+K} \times \Theta$ to R^L, and $E(\psi(Z_i, \theta_0)|X) = 0 \ \forall i \geq 1$. Let

$$\overline{\psi}_n(\theta) = \tfrac{1}{n} \textstyle\sum_{i=1}^n \psi(Z_i, \theta) \ . \tag{3.3}$$

The asymptotic null distribution of the test statistic CK_n depends on that of $(\nu_n(\cdot,\theta), \sqrt{n}\,\overline{\psi}_n(\theta)')'$. The covariance matrix of the latter is defined as follows. Let G^* be any df on R^K. Define

$$(3.4) \qquad C(z_1,\, z_2,\, \theta,\, G^*) \;=\; \iint \begin{pmatrix} (z \le z_1) - F(y_1|x,\theta)(x \le x_1) \\ \psi(z,\theta) \end{pmatrix} \times \begin{pmatrix} (z \le z_2) - F(y_2|x,\theta)(x \le x_2) \\ \psi(z,\theta) \end{pmatrix}' f(y|x,\theta)d\mu(y)dG^*(x)\;.$$

Then,

$$(3.5) \qquad C(z_1,\, z_2,\, \theta,\, G) = \mathrm{Cov}_{F(\cdot|\cdot,\theta)}\left(\left(\nu_n(z_1,\theta),\, \sqrt{n}\,\overline{\psi}_n(\theta)'\right)',\, \left(\nu_n(z_2,\theta),\, \sqrt{n}\,\overline{\psi}_n(\theta)'\right)' \right)\;.$$

Weak convergence of $\left(\nu_n(\cdot,\,\theta_0),\, \sqrt{n}\,\overline{\psi}_n(\theta_0)'\right)'$, which is used to obtain the asymptotic null distribution of CK_n, requires the specification of a pseudometric ρ on R^{V+K}. Given θ_0 and G, we define ρ as follows: For $z_1,\, z_2 \in R^{V+K}$,

$$(3.6) \qquad \rho(z_1,\, z_2) = \left(\iint [(z \le z_1) - (z \le z_2)]^2 f(y|x,\, \theta_0)d\mu(y)dG(x) \right)^{1/2}\;.$$

Given any df F, let $\mathrm{supp}(F)$ denote the support of F.

We now specify assumptions on the parametric model (M1), $\{f(y|x,\theta)\;:\;\theta \in \Theta\}$, and on the estimator (E1), $\widehat{\theta}$, under which the CK_n statistic has the asymptotic null distribution given below.

Assumption M1: (i) *$F(y|X_i,\,\theta)$ is differentiable in θ on a neighborhood N_1 of θ_0 $\forall i \ge 1$.*

(ii) $\sup_{z \in R^{V+K}} \sup_{\theta:\|\theta-\theta_0\| \le r_n} \left\| \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial\theta} F(y|X_i,\,\theta)(X_i \le x) - \Delta_0(z) \right\| \;\to\; 0$ wp1 *for all sequences of positive constants $\{r_n\;:\;n \ge 1\}$ such that $r_n \to 0$ when $\{X_i\;:\;i \ge 1\}$ are* iid *with* df *$G(\cdot)$, where $\Delta_0(z) = \int \frac{\partial}{\partial\theta} F(y|x^*,\,\theta_0)(x^* \le x)dG(x^*)$.*

(iii) $\sup_{z \in R^{V+K}} \|\Delta_0(z)\| < \infty$ *and $\Delta_0(\cdot)$ is uniformly continuous on R^{V+K} (with respect to ρ).*

Assumption E1: (i) $\sqrt{n}(\widehat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n D_0\psi(Z_i,\,\theta_0) + o_p(1)$ *cond'l on X* wp1 *when the sample is generated by the null* df *$F(\cdot|\cdot,\,\theta_0)$, where D_0 is a non-random $L \times L$ matrix that may depend on θ_0.*

(ii) *$\psi(z,\theta)$ is a measurable function from $R^{V+K} \times \Theta$ to R^L that satisfies* (a) $\int \psi(z,\,\theta_0)f(y|x,\,\theta_0)d\mu(y) = 0$ $\forall x \in \mathrm{supp}(G)$ *and* (b) $\int \psi_0^*(x)dG(x) < \infty$, *where* $\psi_0^*(x) = \int \|\psi(z,\,\theta_0)\|^{2+\varepsilon} f(y|x,\,\theta_0)d\mu(y)$ *for some $\varepsilon > 0$.*

Assumption M1 requires that the conditional parametric df is differentiable in θ. This is a weaker assumption than the requirement that the conditional parametric *density* $f(y|x,\theta)$ is differentiable in θ, since the integration of $f(y|x,\theta)$ to obtain $F(y|x,\theta)$ is a smoothing operation. For example, a simple model for which $f(y|x, \theta)$ is not differentiable in θ, but for which Assumption M1 holds, is a linear regression model with double exponential errors. On the other hand, for most models used in practice, $f(y|x, \theta)$ is differentiable in θ. In such cases, it is easy to verify Assumption M1 using the following sufficient condition:

ASSUMPTION M1′: (i) *$f(y|x,\theta)$ is twice continuously differentiable in θ on a neighborhood N_1 of θ_0 $\forall z \in R^{V+K}$.*

(ii) *The score function $s(y|x,\theta) = \frac{\partial}{\partial\theta}\log f(y|x,\theta)$ satisfies $\int s_k^*(x)dG(x) < \infty$ for $k = 1, 2$, where*

$$
\begin{aligned}
s_1^*(x) &= \int \sup_{\theta\in N_1} \|s(y|x,\theta)\|^2 f(y|x,\theta)d\mu(y) \quad \textit{and} \\
s_2^*(x) &= \int \sup_{\theta\in N_1} \left\|\frac{\partial}{\partial\theta'}s(y|x,\theta)\right\| f(y|x,\theta)d\mu(y) \; .
\end{aligned}
$$

LEMMA 1: *Assumption* M1′ *implies Assumption* M1.

COMMENTS: 1. Under Assumption M1′, $\Delta_0(z)$ can be written as

$$\Delta_0(z) = \int\int (z^* \le z)s(y^*|z^*, \theta_0)f(y^*|x^*, \theta_0)d\mu(y^*)dG(x^*) \; .$$

2. Assumption M1′ is satisfied by most generalized linear models (as defined by McCullagh and Nelder (1983)) including probit, logit, and Poisson regression models under moment condition on the covariates, because they are constructed from differentiable link functions and exponential densities, which are differentiable in their parameters. Similarly, tobit (i.e., censored regression), truncated regression, and sample selection models satisfy Assumption M1′ under a moment condition on the covariates.

Assumption E1 requires the estimator $\widehat{\theta}$ to be $\sqrt{n}$–consistent and to have a linear expansion cond'l on X wp1. This is not restrictive for samples with non-trending independent observations.

In the case of the maximum likelihood (ML) estimator, the function $\psi(z,\theta)$ is the conditional score function $\frac{\partial}{\partial\theta}\log f(y|x,\theta)$ and the matrix D_0 is the inverse of the asymptotic information matrix $\int\int \frac{\partial}{\partial\theta}\log f(y|x,\theta_0)\left(\frac{\partial}{\partial\theta'}\log f(y|x,\theta_0)\right) f(y|x,\theta_0)d\mu(y)dG(x)$.[3] The ML estimator satisfies Assumption E1 in all the models mentioned in Comment 2 following Lemma 1 under a moment condition on the covariates.

Under Assumptions D, M1, and E1, the test statistic CK_n has a limit distribution under the null hypothesis. To specify that limit distribution requires the introduction of some additional notation. Let

$$\begin{aligned} H(z) &= \int H(y|x^*)(x^* \leq x)dG(x^*) \ \text{ and} \\ F(z,\theta) &= \int F(y|x^*,\theta)(x^* \leq x)dG(x^*) \ \text{ for } \ z \in R^{V+K} . \end{aligned} \tag{3.7}$$

$H(\cdot)$ is the unconditional df of Z_i. $F(\cdot,\theta)$ is the unconditional parametric df of Z_i. Under the null hypothesis, $F(\cdot,\theta_0) = H(\cdot)$. Under the alternative hypothesis, $F(\cdot,\theta) \neq H(\cdot)$ for any $\theta \in \Theta$, see Section 5 below.

Let

$$\mathcal{Z} = \text{supp}(H) \subset R^{V+K} . \tag{3.8}$$

We show below that CK_n has the same asymptotic distribution as $CK_n(\mathcal{Z})$, where

$$CK_n(\mathcal{Z}) = \sqrt{n}\sup_{z\in\mathcal{Z}}|\widehat{H}_n(z) - \widehat{F}_n(z,\widehat{\theta})| . \tag{3.9}$$

The latter, in turn, has an asymptotic distribution that depends on that of the conditional empirical process $\nu_n(\cdot,\theta_0)$.

Let $\Rightarrow$ denote weak convergence (as defined in Pollard (1984, Ch. IV)). We show that under the null hypothesis

$$\begin{pmatrix} \nu_n(\cdot,\theta_0) \\ \sqrt{n}\,\overline{\psi}_n(\theta_0) \end{pmatrix} \Rightarrow \begin{pmatrix} \nu(\cdot) \\ \nu_0 \end{pmatrix} \ \text{ cond'l on } X \text{ wp1} \tag{3.10}$$

as a sequence of processes indexed by $z \in \mathcal{Z}$. Here, $(\nu(\cdot),\nu_0')'$ is a mean zero Gaussian process with covariance function defined by

$$E\begin{pmatrix} \nu(z_1) \\ \nu_0 \end{pmatrix}\begin{pmatrix} \nu(z_2) \\ \nu_0 \end{pmatrix}' = C(z_1,z_2,\theta_0,G) . \tag{3.11}$$

The sample paths of $\nu(\cdot)$ are uniformly continuous with respect to ρ (defined in (3.6)) on $\mathcal{Z}$ with probability one. This allows one to apply the continuous mapping theorem (see Pollard (1984, Thm. IV.12)) to obtain the asymptotic distribution of functions of $\nu_n(\cdot, \widehat{\theta})$, such as $CK_n(\mathcal{Z})$. Due to the estimation of θ_0 by $\widehat{\theta}$, the asymptotic null distribution of CK_n also depends on $\Delta_0(z)$ and D_0 defined in Assumptions M1 and E1 respectively.

The asymptotic null distribution of CK_n is given in the following theorem.

THEOREM 1: *Suppose Assumptions* D, M1, *and* E1 *hold. Then, under the null hypothesis,*

(a) $CK_n(\mathcal{Z}) \xrightarrow{d} \sup_{z\in\mathcal{Z}} |\nu(z) - \Delta_0(z)'D_0\nu_0|$ *cond'l on* X wp1 *and*

(b) $CK_n \xrightarrow{d} \sup_{z\in\mathcal{Z}} |\nu(z) - \Delta_0(z)'D_0\nu_0|$ *cond'l on* X wp1.

COMMENTS: 1. The asymptotic null distributions of $CK_n(\mathcal{Z})$ and CK_n are the same because $\{Z_i : i \geq 1\}$ is a dense subset of $\mathcal{Z}$ with probability one.

2. The asymptotic null distribution of CK_n is nuisance parameter dependent, since it depends on θ_0 and $G(\cdot)$. In consequence, asymptotic critical values for CK_n cannot be tabulated. Instead, one can obtain critical values and p–values using the parametric bootstrap procedure described in the next section.

4. BOOTSTRAP CRITICAL VALUES

4.1. Definition and Asymptotic Properties of Parametric Bootstrap Critical Values

When bootstrapping any test statistic, one is faced with the task of finding a bootstrap distribution that mimics the null distribution of the data, even though the data may be generated by an alternative distribution. For this reason, the empirical df $\widehat{H}_n(z)$ is not suitable for use as a bootstrap distribution. On the other hand, the conditional df $F(\cdot|\cdot, \widehat{\theta})$ is in the null hypothesis and it mimics $F(\cdot|\cdot, \theta_0)$ or $F(\cdot|\cdot, \theta_1)$, both of which are in the null, depending on whether or not the null hypothesis is true. For this reason, we consider the parametric bootstrap distribution of the data in which the covariates are the same as in the observed sample and the response variables are independent across i with df's $F(\cdot|X_i, \widehat{\theta})$ for $i \leq n$.

The idea of the bootstrap is to pretend that the null df of the data is the bootstrap distribution and to obtain critical values by taking the appropriate quantile from the distribution of CK_n when CK_n is computed from a sample that has the bootstrap distribution. This bootstrap distribution of CK_n is intractable. Thus, one has to approximate it by Monte Carlo simulation.

The bootstrap simulation is carried out as follows. One simulates B bootstrap samples each of size n. The b–th bootstrap sample is denoted $\{Z^*_{ib} : i \leq n\}$ for $b = 1, ..., B$. The b–th bootstrap sample contains the same covariate vectors $\{X_i : i \leq n\}$ as the original sample. In consequence, $Z^*_{ib} = (Y^*_{ib}, X_i)$ for $i \leq n$. Given X_i, one simulates Y^*_{ib} using the parametric conditional density $f(y|X_i, \widehat{\theta})$ (or df $F(y|X_i, \widehat{\theta})$). This is repeated (independently) for $i = 1, ..., n$ to give $\{Y^*_{ib} : i \leq n\}$. Since $Z^*_{ib} = (Y^*_{ib}, X_i)$, this yields the b–th bootstrap sample $\{Z^*_{ib} : i \leq n\}$. One repeats this procedure for $b = 1, ..., B$.

Next, one computes the b–th bootstrap value of CK_n, call it CK^*_{nb}, by applying the definition of CK_n to the b–th bootstrap sample $\{Z^*_{ib} : i \leq n\}$ in place of the original sample $\{Z_i : i \leq n\}$. Repeating this for $b = 1, ..., B$ gives a sample $\{CK^*_{nb} : b = 1, ..., B\}$ of CK_n values. This sample mimics a random sample of draws of CK_n under a parametric null distribution. Thus, its $(1-\alpha)$–th sample quantile, denoted $c_{\alpha nB}(\widehat{\theta})$, yields a critical value of significance level α for CK_n. This critical value is valid asymptotically provided $B \to \infty$ as $n \to \infty$ by the results given below. The p–value of the CK_n test is obtained from the bootstrap values $\{CK^*_{nb} : b = 1, ..., B\}$ by computing the fraction of CK^*_{nb} values that are greater than the observed value of CK_n.

Note that the parametric bootstrap procedure defined above uses the observed covariate values $\{X_i : i \leq n\}$ in each bootstrap sample that is simulated — only the response variables differ across bootstrap samples. This makes the bootstrap samples mimic the actual sample as closely as possible. It promises better finite sample performance of the bootstrap critical values than if the covariate values for each bootstrap sample are simulated as being iid with df $\widehat{G}_n(\cdot)$. (Limited Monte Carlo simulation of the bootstrap procedure substantiates quite clearly the better finite sample performance of the "fixed covariates" bootstrap.) To justify the use of the "fixed covariates" parametric bootstrap, it is necessary to establish asymptotic results that hold not only unconditionally, but also conditionally on $\{X_i : i \geq 1\}$ with $\{X_i : i \geq 1\}$ probability one, as is done below.

We now provide the asymptotic justification of the parametric bootstrap. Let

$$\mathcal{L}_\theta(CK_n|X) \tag{4.1}$$

denote the conditional distribution (or law) of CK_n given the covariates $\{X_i : i \geq 1\}$ when the sample is generated according to the null hypothesis with conditional df $F(\cdot|\cdot,\theta)$. Let $P_\theta(\cdot|X)$ denote conditional probability given $\{X_i : i \geq 1\}$ under the null with conditional df $F(\cdot|\cdot,\theta)$.

By Theorem 1 above,

$$\mathcal{L}_{\theta_0}(CK_n|X) \xrightarrow{d} \sup_{z\in\mathcal{Z}} |\nu(z) - \Delta_0(z)'D_0\nu_0| \quad \text{wp1}. \tag{4.2}$$

We show below that, under suitable assumptions, for any sequence of non-random parameters $\{\theta_n : n \geq 1\}$ such that $\theta_n \to \theta_0$, we have

$$\mathcal{L}_{\theta_n}(CK_n|X) \xrightarrow{d} \sup_{z\in\mathcal{Z}} |\nu(z) - \Delta_0(z)'D_0\nu_0| \quad \text{wp1}. \tag{4.3}$$

That is, $\mathcal{L}_{\theta_n}(CK_n|X)$ has the same limit as $\mathcal{L}_{\theta_0}(CK_n|X)$.

For fixed θ_n, let $c_{\alpha n}(\theta_n)$ denote the level α critical value obtained from $\mathcal{L}_{\theta_n}(CK_n|X)$ for $\alpha \in (0,1)$. That is, $P_{\theta_n}(CK_n > c_{\alpha n}(\theta_n)|X) = \alpha$. Note that $c_{\alpha n}(\theta_n)$ depends on $\{X_i : i \leq n\}$. Let $c_\alpha(\theta_0)$ denote the level α critical value obtained from the limit distribution of CK_n. That is, $P(\sup_{z\in\mathcal{Z}} |\nu(z) - \Delta_0(z)'D_0\nu_0| > c_\alpha(\theta_0)) = \alpha$. Now, $P_{\theta_n}(CK_n > c_{\alpha n}(\theta_n)|X) = \alpha\ \forall n$, $P_{\theta_n}(CK_n > c_\alpha(\theta_0)|X) \to \alpha$ wp1, and absolute continuity of the limit distribution of CK_n (which holds because it is the supremum of a Gaussian process whose covariance function is nonsingular, see Lifshits (1982)) imply that

$$c_{\alpha n}(\theta_n) \to c_\alpha(\theta_0) \quad \text{wp1}. \tag{4.4}$$

That is, for any sequence $\{\theta_n : n \geq 1\}$ for which $\theta_n \to \theta_0$ and (4.3) holds, we have $c_{\alpha n}(\theta_n) \to c_\alpha(\theta_0)$ wp1. Note that this holds no matter what value θ_0 takes on within Θ provided Assumptions M1 and E1 hold for all values θ_0 in Θ.

This result and the Skorokhod representation theorem (e.g., see Billingsley (1979, Thm. 25.6, p. 287)) imply that if $\widehat{\theta} \xrightarrow{p} \theta_1$ cond'l on X wp1, then

$$c_{\alpha n}(\widehat{\theta}) \xrightarrow{p} c_{\alpha}(\theta_1) \text{ cond'l on } X \text{ wp1} \tag{4.5}$$

whether or not the null hypothesis is true.[4] By assumption, if the null is true, θ_1 equals the true value θ_0. This implies that when the null is true, we have

$$P_{\theta_0}(CK_n > c_{\alpha n}(\widehat{\theta})|X) = P_{\theta_0}(CK_n + o_p(1) > c_{\alpha}(\theta_0)|X) \to \alpha \text{ wp1}, \tag{4.6}$$

as desired. In addition, this implies that the asymptotic unconditional rejection probability of the CK_n test is α (see footnote 2):

$$\lim_{n\to\infty} P_{\theta_0}(CK_n > c_{\alpha n}(\widehat{\theta})) = \alpha\ . \tag{4.7}$$

Provided Assumptions M1 and E1 and (4.3) hold for any $\theta_0 \in \Theta$, this gives

$$\sup_{\theta_0 \in \Theta} \lim_{n\to\infty} P_{\theta_0}(CK_n > c_{\alpha n}(\widehat{\theta})) = \alpha\ . \tag{4.8}$$

That is, the asymptotic significance level of the CK_n test with critical value $c_{\alpha n}(\widehat{\theta})$ is α, as desired.

Equations (4.5)–(4.8) justify the use of the bootstrap critical value $c_{\alpha n}(\widehat{\theta})$. The latter corresponds to the case where the number of bootstrap repetitions B equals ∞. If $B < \infty$, but $B \to \infty$ as $n \to \infty$, the differences between the approximate bootstrap critical values $c_{\alpha nB}(\widehat{\theta})$ and the bootstrap critical values $c_{\alpha n}(\widehat{\theta})$ go to zero in probability and almost surely (a.s.) with respect to the bootstrap simulation randomness, because sample quantiles of iid rv's converge a.s. to population quantiles. In consequence, the $B < \infty$ approximate bootstrap critical values $c_{\alpha nB}(\widehat{\theta})$ are asymptotically valid provided $B \to \infty$ as $n \to \infty$.

It remains to show that (4.3) holds. To do this, we augment Assumptions M1 and E1 by the following Assumptions M2 and E2. For brevity, we say that the sample is distributed "under $\{\theta_n : n \geq 1\}$" when the covariates $\{X_i : i \geq 1\}$ are iid with df $G(\cdot)$ and the response variables form a triangular array of rv's $\{Y_i : i \leq n,\ n \geq 1\}$ that are independent across observations in each row with the df of Y_i $(= Y_{ni})$ given by the parametric conditional df $F(\cdot|X_i,\ \theta_n)$ for $i \leq n$.

ASSUMPTION M2: (i) $C(z_1, z_2, \theta, G)$ *is continuous in* θ *at* θ_0 $\forall z_1, z_2 \in \mathcal{Z}$.

(ii) $\int\int |f(y|x, \theta) - f(y|x, \theta_0)| d\mu(y) dG(x) \to 0$ *as* $\theta \to \theta_0$.

ASSUMPTION E2: (i) *For all non-random sequences* $\{\theta_n : n \geq 1\}$ *for which* $\theta_n \to \theta_0$, *we have* $\sqrt{n}(\widehat{\theta} - \theta_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} D_0 \psi(Z_i, \theta_n) + o_p(1)$ *under* $\{\theta_n : n \geq 1\}$ *cond'l on* X wp1 *for* D_0 *and* $\psi(z, \theta)$ *as in Assumption* E1.

(ii) $\int \psi_1^*(x) dG(x) < \infty$, *where* $\psi_1^*(x) = \sup_{\theta \in N_1} \int \|\psi(z, \theta)\|^{2+\varepsilon} f(y|x, \theta) d\mu(y)$ *for some* $\varepsilon > 0$.

Assumption M2 does not require $f(y|x, \theta)$ and $\psi(z, \theta)$ to be continuous in θ. Nevertheless, if they are, Assumption M2 is implied by the following:

ASSUMPTION M2′: (i) $f(y|x, \theta)$ *and* $\psi(z, \theta)$ *are continuous at* θ_0 $\forall z \in \mathcal{Z}$.

(ii) $\int \psi_2^*(x) dG(x) < \infty$, *where* $\psi_2^*(x) = \int \sup_{\theta \in N_1} (\|\psi(z, \theta)\|^2 + 1) f(y|x, \theta) d\mu(y)$.

LEMMA 2: *Assumption* M2′ *implies Assumption* M2.

COMMENT: Assumption M2′ and, hence, Assumption M2 hold for all models mentioned above and below Lemma 1 except for the double exponential regression model (since $\psi(z, \theta)$ is not continuous at θ_0 for all $z \in \mathcal{Z}$). Assumption M2 holds for the latter model by direct verification.

Assumption E2 is not overly restrictive. Most proofs of asymptotic normality of parametric estimators can be altered straightforwardly to yield the triangular array linear expansion of Assumption E2(i). For example, Assumption E2 holds in all the models mentioned above and below Lemma 1.

The validity of (4.3) is established in the following Theorem.

THEOREM 2: *Suppose Assumptions* M1, M2, *and* E2 *hold. Then, for any non-random sequence* $\{\theta_n : n \geq 1\}$ *for which* $\theta_n \to \theta_0$, *we have*

(a) $CK_n(\mathcal{Z}) \xrightarrow{d} \sup_{z \in \mathcal{Z}} |\nu(z) - \Delta_0(z)' D_0 \nu_0|$ *under* $\{\theta_n : n \geq 1\}$ *cond'l on* X wp1 *and*

(b) $CK_n \xrightarrow{d} \sup_{z \in \mathcal{Z}} |\nu(z) - \Delta_0(z)' D_0 \nu_0|$ *under* $\{\theta_n : n \geq 1\}$ *cond'l on* X wp1.

Assumptions E1 and E2 specify the behavior of $\widehat{\theta}$ under $F(\cdot|\cdot, \theta_0)$ and $F(\cdot|\cdot, \theta_n)$ respectively.

Its behavior for an arbitrary conditional distribution $H(\cdot|\cdot)$ is given by

ASSUMPTION E3: *$\widehat{\theta} \xrightarrow{p} \theta_1$ cond'l on X* wp1 *for some $\theta_1 \in \Theta$ under Assumption* D.

If $H(\cdot|\cdot)$ equals $F(\cdot|\cdot, \theta_0)$, then by Assumptions E1 and E3, θ_1 equals the true value θ_0. When $H(\cdot|\cdot)$ is not in the parametric model, θ_1 is often referred to as a "pseudo-true" value. If $H(\cdot|\cdot)$ is not in the parametric model and $\widehat{\theta}$ is the ML estimator, θ_1 is the value that maximizes $\int\int \log f(y|x, \theta) dH(y|x) dG(x)$ over $\theta \in \Theta$. That is, θ_1 is the value that minimizes the Kullback–Leibler information distance between the true distribution of the data and the parametric model.

Now, Theorem 2 and (4.5)–(4.8) combine to give the desired asymptotic justification of the bootstrap.

COROLLARY 1: (a) *Suppose Assumptions* D, M1, M2, *and* E2 *hold,* $H(\cdot|\cdot)$ *of Assumption* D *equals* $F(\cdot|\cdot, \theta_0)$ *(i.e., $H(\cdot|\cdot)$ is in the null hypothesis), and $B \to \infty$ as $n \to \infty$. Then,*

$$c_{\alpha nB}(\widehat{\theta}) \xrightarrow{p} c_\alpha(\theta_0) \text{ cond'l on } X \text{ wp1}, \; P_{\theta_0}(CK_n(\mathcal{Z}) > c_{\alpha nB}(\widehat{\theta})|X) \to \alpha \text{ wp1, and}$$

$$P_{\theta_0}(CK_n > c_{\alpha nB}(\widehat{\theta})|X) \to \alpha \text{ wp1}.$$

(b) *Suppose Assumption* D *holds, Assumptions* M1, M2, *and* E2 *hold for any value of* $\theta_0 \in \Theta$, *and $B \to \infty$ as $n \to \infty$. Then,*

$$\sup_{\theta_0 \in \Theta} \lim_{n \to \infty} P_{\theta_0}(CK_n > c_{\alpha nB}(\widehat{\theta})) = \alpha \; .$$

(c) *Suppose Assumptions* D *and* E3 *hold, Assumptions* M1, M2, *and* E2 *hold for any value* $\theta_0 \in \Theta$, *and $B \to \infty$ as $n \to \infty$. Then,*

$$c_{\alpha nB}(\widehat{\theta}) \xrightarrow{p} c_\alpha(\theta_1) \text{ cond'l on } X \text{ wp1}.$$

COMMENTS: 1. The randomness of parts (a), (b), and (c) includes both the randomness of the sample and the independent randomness of the bootstrap simulations.

2. Part (c) shows that even when the null hypothesis fails, the bootstrap critical value still converges in probability to a finite constant $c_\alpha(\theta_1)$ under Assumption E3. This ensures that the CK_n has power (see Section 5 below).

4.2. Simulation Performance of the Parametric Bootstrap

In this section, the performance of the parametric bootstrap is evaluated in a small simulation experiment. We take the parametric family to be a trivariate logit model with

$$(4.9) \qquad P(Y_i = e_j) = \exp(X_{ij}^{*\prime}\theta_0)/\textstyle\sum_{k=1}^{3}\exp(X_{ik}^{*\prime}\theta_0) \quad \text{for} \quad j = 1, 2, 3\ ,$$

where $e_1 = (1,\ 0,\ 0)'$, $e_2 = (0,\ 1,\ 0)'$, $e_3 = (0,\ 0,\ 1)'$, $X_{i1}^* = (1,\ 0,\ X_{i1},\ X_{i2},\ 0,\ 0)'$, $X_{i2}^* = (0, 1, X_{i3}, 0, X_{i4}, 0)'$, $X_{i3}^* = (0, 0, X_{i5}, 0, 0, X_{i6})'$, $X_i = (X_{i1}, ..., X_{i6})' \sim$ iid $N(0, \sigma_X^2 I_6)$, and $\theta_0 = (1, ..., 1)' \in R^6$. Three values of σ_X^2 are considered: 1/2, 1, and 3. Four sample sizes are considered: 25, 50, 100, and 250. The number of bootstrap repetitions is 299. The number of Monte Carlo simulation repetitions is 4,000 for the smaller sample sizes and 2,000 for the largest. Computational requirements prevented us from considering larger sample sizes and more bootstrap repetitions. Such cases certainly are feasible in applications, however, since one need not compute 2,000 or 4,000 simulation repetitions in any given application.

Table 1 provides the results. The numbers given in parentheses are asymptotic (as the number of simulation repetitions goes to infinity) standard errors of the simulated rejection probabilities. The results indicate over-rejection for most cases when the sample size is less than or equal to 100. For sample size 250, the results are quite good — the simulated true sizes are all within two simulation standard errors of the nominal sizes.

5. CONSISTENCY OF THE CONDITIONAL KOLMOGOROV TEST

In this section, we show that the CK test is consistent against any conditional df $H(\cdot|\cdot)$ in the alternative hypothesis H_1. By definition, $H(\cdot|\cdot)$ is in the alternative hypothesis if it satisfies

Assumption H1: *For each $\theta \in \Theta$, there exists $y \in R^V$ for which*

$$P_G(H(y|X) \neq F(y|X,\theta)) > 0\ ,$$

where $X \sim G$ and y may depend on θ.

Assumption H1 requires that $H(\cdot|x)$ differs from each parametric conditional df $F(\cdot|x,\theta)$ for some value(s) x of the covariates that occur(s) with positive probability. Note that Assumption

H1 is equivalent to:

(5.1) *For each* $\theta \in \Theta$, $H(z) \neq F(z, \theta)$ *for some* $z \in R^{V+K}$ *(where* z *may depend on* θ*).*

Consistency of the conditional Kolmogorov test is established in the following theorem:

THEOREM 3: *Under Assumptions* D, M1, E3, *and* H1, *for all sequences of* rv's $\{c_n : n \geq 1\}$ *with* $c_n = O_p(1)$ *cond'l on* X wp1, *we have*

$$\lim_{n\to\infty} P(CK_n(\mathcal{Z}) > c_n|X) = 1 \ \text{ wp1 and } \ \lim_{n\to\infty} P(CK_n > c_n|X) = 1 \ \text{ wp1.}$$

COMMENTS: 1. The bootstrap critical values discussed in Section 4 converge in probability to constants cond'l on X wp1 under the null and under the alternative, so they satisfy the requirements of Theorem 3 on $\{c_n : n \geq 1\}$.

2. There may be some combinations of conditional df's $H(\cdot|\cdot)$ and estimators $\widehat{\theta}$ for which Assumption E3 does not hold. For example, suppose $\widehat{\theta}$ is the ML estimator and the Kullback–Leibler information distance between the true df of the data $H(z)$ and the parametric df's $\{F(\cdot, \theta) : \theta \in \Theta\}$ is minimized not at a unique value θ_1, but rather, at a set of values $\Theta_1 \subset \Theta$. In this case, $\widehat{\theta}$ will not converge in probability to a constant Θ_1. But, the distance between $\widehat{\theta}$ and Θ_1 typically will converge in probability to zero. In consequence, $c_{\alpha n B}(\widehat{\theta})$ will be $O_p(1)$ provided $\sup_{\theta\in\Theta_1} c_\alpha(\theta) < \infty$. Theorem 3 can be extended to include conditional df's $H(\cdot|\cdot)$ and estimators $\widehat{\theta}$ that exhibit such behavior. For brevity and because such behavior is relatively rare, however, we do not do so here.

6. LOCAL POWER OF THE CONDITIONAL KOLMOGOROV TEST

In this section, we determine the power of the CK test against contiguous local alternatives to the null hypothesis. The alternatives we consider are of distance $1/\sqrt{n}$ from the null hypothesis.

Suppose one is interested in the power of the CK test for sample size n_0 against an alternative conditional distribution of Y_i given X_i defined by the density $q(y|x)$ with respect to the σ–finite measure μ. Since $q(\cdot|\cdot)$ is an alternative density, $q(\cdot|\cdot) \notin \{f(\cdot|\cdot, \theta) : \theta \in \Theta\}$. Let $Q(\cdot|\cdot)$ denote the df corresponding to $q(\cdot|\cdot)$.

Let

$$d(z) = \sqrt{n_0}(q(y|x) - f(y|x,\, \theta_0))\ . \tag{6.1}$$

Define the following sequence of local alternative conditional densities:

$$q_n(y|x) = f(y|x,\, \theta_0) + d(z)/\sqrt{n} \ \text{ for } \ n = 1, 2, \ldots\ . \tag{6.2}$$

Note that $q_n(y|x)$ is a proper density (at least) for all n greater than or equal to n_0. Let $Q_n(\cdot|\cdot)$ denote the df corresponding to $q_n(\cdot|\cdot)$. The sequence of local alternative distributions that we consider is that of $\{Z_i : i \leq n\}$ when $\{Z_i : i \leq n\}$ are distributed independently with $Y_i \sim Q_n(\cdot|X_i)$ and $X_i \sim G(\cdot)$ for $i \leq n$ and $n \geq 1$. In this case, we say that the observations are distributed "under $\{Q_n(\cdot|\cdot) : n \geq 1\}$." Note that the n_0–th distribution in the sequence is that of $\{Z_i : i \leq n_0\}$ when $\{Z_i : i \leq n_0\}$ is distributed independently with $Y_i \sim Q(\cdot|X_i)$ and $X_i \sim G(\cdot)$, as desired.

To ensure that $\{Q_n(\cdot|\cdot) : n \geq 1\}$ generates contiguous alternatives (CA) to $F(\cdot|\cdot,\, \theta_0)$, we assume:

Assumption CA: *$\int\int h_j(y,x)d\mu(y)dG(x) < \infty$ for $j = 1, 2,$ where $h_1(y,x) = \|\psi(y,\, x,\, \theta_0)\|^{2+\varepsilon}$ $q(y|x)$ for some constant $\varepsilon > 0$ and $h_2(y,x) = \sup_{\lambda \in [0,\delta]} |d(y,x)/(f(y|x,\, \theta_0)+\lambda d(y,x))|^3(f(y|x,\, \theta_0)$ $+\, q(y|x))$ for some constant $\delta > 0$.*

The asymptotic distribution of CK_n under local alternatives is given in the following theorem.

Theorem 4: *Suppose Assumptions* M1, E1, *and* CA *hold. Then,*

(a) $CK_n(\mathcal{Z}) \xrightarrow{d} M$ *under* $\{Q_n(\cdot|\cdot) : n \geq 1\}$ *cond'l on* X wp1 *and*

(b) $CK_n \xrightarrow{d} M$ *under* $\{Q_n(\cdot|\cdot) : n \geq 1\}$ *cond'l on* X wp1,

where

$$\begin{aligned} M &= \sup_{z \in \mathcal{Z}} |\nu(z) - \Delta_0(z)' D_0 \nu_0 + \mu(z)|, \\ \mu(z) &= \sqrt{n_0}\left[\int (Q(y|x^*) - F(y|x^*,\, \theta_0))(x^* \leq x) dG(x^*)\right. \\ &\quad - \left. \Delta_0(z)' D_0 \int\int \psi(z^*,\, \theta_0) q(y^*|x^*) d\mu(y^*) dG(x^*)\right], \end{aligned}$$

and $(\nu(\cdot),\, \nu_0)$ *is the Gaussian process defined in Section* 3.

COMMENTS: 1. In Theorem 4, we create local alternatives by starting with an alternative df $Q(\cdot|\cdot)$ and shrinking it to a df $F(\cdot|\cdot,\, \theta_0)$ in the conditional parametric model. What is the most appropriate choice of parameter value θ_0 to shrink $Q(\cdot|\cdot)$ towards? Note that the choice of this value θ_0 affects the asymptotic result. It makes no sense to choose θ_0 to be the true value of θ_0 under the null, because the latter is undefined when null is not true, as is the case with local alternatives. Instead, it seems natural to choose θ_0 to be the value in Θ such that $F(\cdot|\cdot,\, \theta_0)$ mimics $Q(\cdot|\cdot)$ most closely in terms of the behavior of $\widehat{\theta}$. Let θ_1 denote the probability limit of $\widehat{\theta}$ under the fixed alternative $Q(\cdot|\cdot)$. If we take $\theta_0 = \theta_1$, then $Q(\cdot|\cdot)$ and $F(\cdot|\cdot,\, \theta_0)$ are close in the sense that $\widehat{\theta}$ has the same probability limit under both distributions and $\sqrt{n}(\widehat{\theta} - \theta_0)$ has the same asymptotic distribution under the local alternatives as under $F(\cdot|\cdot,\, \theta_0)$. Furthermore, when $\widehat{\theta}$ is the (quasi-) ML estimator, the choice of $\theta_0 = \theta_1$ corresponds to taking θ_0 to be the value in Θ that minimizes the Kullback–Leibler information distance between $Q(\cdot|\cdot)$ and the conditional parametric model (when the covariates have df $G(\cdot)$).

If one takes $\theta_0 = \theta_1$, as suggested above, then $\int\int \psi(z,\, \theta_0) q(y|x) d\mu(y) dG(x) = 0$, because the latter are the asymptotic first order conditions of $\widehat{\theta}$ under the fixed alternative $Q(\cdot|\cdot)$. In this case, the asymptotic distribution of CK_n under the local alternatives $\{Q_n(\cdot|\cdot) \,:\, n \geq 1\}$ simplifies, because $\mu(z)$ simplifies to

$$(6.3) \qquad \mu(z) = \sqrt{n_0} \int (Q(y|x^*) - F(y|x^*,\, \theta_0))(x^* \leq x) dG(x^*)\ .$$

2. Theorem 4 and Anderson's Lemma (e.g., see Ibragimov and Has'minski (1981, Lemma 10.1, p. 155)) can be combined to show that the CK_n test is *asymptotically locally unbiased.* By definition, the latter holds when

$$(6.4) \qquad \lim_{n\to\infty} P_{Q_n}(CK_n > c_{\alpha n B}(\widehat{\theta})) \geq \alpha \quad \left(= \sup_{\theta_0 \in \Theta} \lim_{n\to\infty} P_{\theta_0}(CK_n > c_{\alpha n B}(\widehat{\theta})) \right)$$

for all local alternatives $\{Q_n(\cdot|\cdot) \,:\, n \geq 1\}$.

The proof is as follows. By Anderson's Lemma, since $\nu(z) - \Delta_0(z)' D_0 \nu_0$ has mean zero $\forall z \in \mathcal{Z}$, for any finite set $\mathcal{Z}_j = \{z_1, ..., z_j\} \subset \mathcal{Z}$,

$$\begin{aligned} \text{(6.5)} \qquad & P(\sup_{z\in\mathcal{Z}_j} |\nu(z) - \Delta_0(z)'D_0\nu_0 + \mu(z)| > c_\alpha(\theta_0)) \\ \geq\ & P(\sup_{z\in\mathcal{Z}_j} |\nu(z) - \Delta_0(z)'D_0\nu_0| > c_\alpha(\theta_0)) \ . \end{aligned}$$

$\mathcal{Z}$ is totally bounded under the metric ρ (see the proof of Lemma A.5). Hence, $\mathcal{Z}$ is separable and there exists a sequence of increasing finite sets $\{\mathcal{Z}_j : j \geq 1\}$ whose union is dense in $\mathcal{Z}$. In consequence, the uniform continuity of $\Delta_0(\cdot)$ and of the sample paths of $\nu(\cdot)$ with probability one, the monotone convergence theorem, and (6.5) yield

$$\begin{aligned} \text{(6.6)} \qquad & P(\sup_{z\in\mathcal{Z}} |\nu(z) - \Delta_0(z)'D_0\nu_0 + \mu(z)| > c_\alpha(\theta_0)) \\ =\ & P(\sup_{z\in\cup_{j=1}^\infty \mathcal{Z}_j} |\nu(z) - \Delta_0(z)'D_0\nu_0 + \mu(z)| > c_\alpha(\theta_0)) \\ =\ & \lim_{j\to\infty} P(\sup_{z\in\mathcal{Z}} |\nu(z) - \Delta_0(z)'D_0\nu_0 + \mu(z)| > c_\alpha(\theta_0)) \\ \geq\ & \lim_{j\to\infty} P(\sup_{z\in\mathcal{Z}_j} |\nu(z) - \Delta_0(z)'D_0\nu_0| > c_\alpha(\theta_0)) \\ =\ & P(\sup_{z\in\cup_{j=1}^\infty \mathcal{Z}_j} |\nu(z) - \Delta_0(z)'D_0\nu_0| > c_\alpha(\theta_0)) \\ =\ & P(\sup_{z\in\mathcal{Z}} |\nu(z) - \Delta_0(z)'D_0\nu_0| > c_\alpha(\theta_0)) \\ =\ & \alpha \ . \end{aligned}$$

Since the left-hand sides of (6.4) and (6.6) are equal by Theorem 4, asymptotic local unbiasedness of the CK_n test is established.

3. A stronger property than asymptotic local unbiasedness is asymptotic local *strict* unbiasedness. This holds if (6.4) holds with a strict inequality. This property is more difficult to establish than asymptotic local unbiasedness (because even if the inequality in (6.6) is strict for any fixed j, which it is when $\mu(z) \neq 0$ for some $z \in \mathcal{Z}_j$, it is not necessarily strict in the limit as $j \to \infty$). Nevertheless, one can conjecture that if $\mu(\cdot) \neq 0$, $\mu(\cdot)$ is uniformly continuous with respect to ρ, and $\mathrm{Var}(\nu(z) - \Delta_0(z)'D_0\nu_0) > 0$ for some $z \in \mathcal{Z}$ for which $\mu(z) \neq 0$, then

$$\text{(6.7)} \quad P(\sup_{z\in\mathcal{Z}} |\nu(z) - \Delta_0(z)'D_0\nu_0 + \mu(z)| > c_\alpha(\theta_0)) > P(\sup_{z\in\mathcal{Z}} |\nu(z) - \Delta_0(z)'D_0\nu_0| > c_\alpha(\theta_0)) \ .$$

If this conjecture is true and we choose $\theta_0 = \theta_1$ as discussed in Comment 1 above, then the CK_n test is asymptotically locally strictly unbiased against all local alternatives $\{Q_n(\cdot|\cdot) : n \geq 1\}$ that are based on a conditional df $Q(\cdot|\cdot)$ that satisfies Assumption H1 (which implies $\mu(\cdot) \neq 0$) and that satisfies Assumption CA (which ensures that $\mu(\cdot)$ is uniformly continuous with respect to ρ).

4. The asymptotic local power of the CK_n test against $\{Q_n(\cdot|\cdot) : n \geq 1\}$ is $P(M > c_\alpha(\theta_0))$. In addition, $Q_{n_0}(\cdot|\cdot) = Q(\cdot|\cdot)$. This suggests approximating the power of the CK test against $Q(\cdot|\cdot)$ when the sample size is n_0 by $P(M > c_\alpha(\theta_0))$.

5. In contrast to Zheng's (1993, 1994) tests, the CK_n test has non-trivial power against $1/\sqrt{n}$ local alternatives.

APPENDIX

In Section A.1 below, we introduce assumptions such that for a fixed (non-random) sequence $\{X_i : i \geq 1\}$ the desired asymptotic results stated in the text hold. In Section A.2, we show that these assumptions hold wp1 if the covariates are iid with df $G(\cdot)$ and the corresponding assumptions of the text hold. This establishes the results stated in the text. Lastly, in Section A.3, we prove the fixed covariates results of Section A.1.

Note that the results of Section A.1 can be used to establish the asymptotic validity of the CK_n test when the covariates are inid. In this case, the limit df $G(\cdot)$ of Section A.1 equals $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} G_i(\cdot)$, where $X_i \sim G_i(\cdot)\ \forall i \geq 1$.

A.1. Fixed Covariate Assumptions and Results

Definitions and notation used here are as in the text, except where stated otherwise.

For $z \in R^{V+K}$ and $\varepsilon > 0$, let $B(z, \varepsilon) = \{z_1 \in R^{V+K} : \rho(z_1, z) \leq \varepsilon\}$.

First, we provide assumptions under which CK_n has the desired asymptotic null distribution for the case of fixed (F) covariates:

ASSUMPTION F.D: (i) $\{X_i : i \geq 1\}$ *are fixed, i.e., non-random.*

(ii) $\{Y_i : i \geq 1\}$ *are independent with conditional* df $H(\cdot|X_i)$ *of* Y_i *given* X_i *for all* $i \geq 1$ *for some conditional* df $H(\cdot|\cdot)$.

ASSUMPTION F.C1: (i) $\widehat{G}_n(x) \to G(x)\ \forall x \in R^K$ *for some* df $G(\cdot)$.

(ii) $\mathrm{supp}(\widehat{G}_n) \subset \mathrm{supp}(G)\ \forall n \geq 1$.

(iii) $\sup_{z\in R^{V+K}} \left| \int \int (z^* \leq z) f(y^*|x^*, \theta_0) d\mu(y^*)(d\widehat{G}_n(x^*) - dG(x^*)) \right| \to 0$.

(iv) $C(z_1, z_2, \theta_0, \widehat{G}_n) \to C(z_1, z_2, \theta_0, G)\ \forall z_1, z_2 \in R^{V+K}$.

(v) $\int \int (z^* \in B(z, 1/k)) f(y^*|x^*, \theta_0) d\mu(y^*)(d\widehat{G}_n(x^*) - dG(x^*)) \to 0$ *for all integers* $k \geq 1$, *for all* z *in a countably dense subset* $\mathcal{Z}_d$ *of* R^{V+K} *(with respect to the metric* ρ*).*

ASSUMPTION F.M1: (i) $F(y|X_i, \theta)$ *is differentiable in* θ *on a neighborhood* N_1 *of* $\theta_0\ \forall i \geq 1$.

(ii) $\sup_{z\in R^{V+K}} \sup_{\theta:\|\theta-\theta_0\|\leq r_n} \left\| \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta} F(y|X_i, \theta)(X_i \leq x) - \Delta_0(z) \right\| \to 0$ *for all sequences of* *positive constants* $\{r_n : n \geq 1\}$ *such that* $r_n \to 0$, *where* $\Delta_0(z) = \int \frac{\partial}{\partial\theta}F(y|x^*, \theta_0)(x^* \leq x)dG(x^*)$.

(iii) $\sup_{z \in R^{V+K}} \|\Delta_0(z)\| < \infty$ *and* $\Delta_0(\cdot)$ *is uniformly continuous on* R^{V+K} *(with respect to* ρ*).*

ASSUMPTION F.E1: (i) $\sqrt{n}(\widehat{\theta} - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n D_0\psi(Z_i, \theta_0) + o_p(1)$ *when the sample is generated by the null* df $F(\cdot|\cdot, \theta_0)$, *where* D_0 *is a non-random* $L \times L$ *matrix that may depend on* θ_0.

(ii) $\psi(z, \theta)$ *is a measurable function from* $R^{V+K} \times \Theta$ *to* R^L *that satisfies* (a) $\int \psi(z, \theta_0)f(y|x, \theta_0) d\mu(y) = 0 \ \forall x \in \mathrm{supp}(G)$ *and* (b) $\overline{\lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \psi_0^*(X_i) < \infty$, *where* $\psi_0^*(x) = \int \|\psi(z, \theta_0)\|^{2+\varepsilon} f(y|x, \theta_0)d\mu(y)$ *for some* $\varepsilon > 0$.

ASSUMPTION F.M1′: (i) $f(y|x, \theta)$ *is twice continuously differentiable in* θ *on a neighborhood* N_1 *of* θ_0 $\forall z \in R^{V+K}$.

(ii) $\overline{\lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^n s_k^*(X_i) < \infty$ *for* $k = 1, 2$ *and* $\int s_1^*(x)dG(x) < \infty$.

(iii) $\sup_{z \in R^{V+K}} |\int \int (z^* \leq z)s(y^*|x^*, \theta_0)f(y^*|x^*, \theta_0)d\mu(y^*)(d\widehat{G}_n(x^*) - dG(x^*))| \to 0$.

LEMMA A.1: *Assumption* F.M1′ *implies Assumption* F.M1.

THEOREM A.1: *Suppose Assumptions* F.D, F.C1, F.M1, *and* F.E1 *hold. Then, under the null,* (a) $CK_n(\mathcal{Z}) \xrightarrow{d} \sup_{z \in \mathcal{Z}} |\nu(z) - \Delta_0(z)' D_0 \nu_0|$ *and* (b) $CK_n \xrightarrow{d} \sup_{z \in \mathcal{Z}} |\nu(z) - \Delta_0(z)' D_0 \nu_0|$.

COMMENT: The convergence in distribution in Theorem A.1 is with respect to the randomness in $\{Y_i : i \geq 1\}$ only, since $\{X_i : i \geq 1\}$ are fixed.

Next, we provide bootstrap results for the fixed covariate case:

ASSUMPTION F.C2: *For all non-random sequences* $\{\theta_n : n \geq 1\}$ *for which* $\theta_n \to \theta_0$, *we have*

(i) $C(z_1, z_2, \theta_n, \widehat{G}_n) - C(z_1, z_2, \theta_n, G) \to 0 \ \forall z_1, z_2 \in \mathcal{Z}$ *and*

(ii) $\int \int (z^* \in B(z, 1/k))f(y^*|x^*, \theta_n)d\mu(y^*)(d\widehat{G}_n(x^*) - dG(x^*)) \to 0 \ \forall z \in \mathcal{Z}_d$, $\forall k \geq 1$, *where* $\mathcal{Z}_d$ *is as in Assumption* F.C1.

ASSUMPTION F.E2: (i) *For all non-random sequences* $\{\theta_n : n \geq 1\}$ *for which* $\theta_n \to \theta_0$, *we have* $\sqrt{n}(\widehat{\theta} - \theta_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^n D_0\psi(Z_i, \theta_n) + o_p(1)$ *under* $\{\theta_n : n \geq 1\}$ *for* D_0 *and* $\psi(z, \theta)$ *as in Assumption* F.E1.

(ii) $\overline{\lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \psi_1^*(X_i) < \infty$.

In Sections A.1 and A.3, we say that the sample is distributed "under $\{\theta_n : n \geq 1\}$" ("under $\{Q_n(\cdot|\cdot) : n \geq 1\}$") when the covariates are fixed and the response variables form a triangular array of rv's $\{Y_i : i \leq n,\ n \geq 1\}$ that are independent across observations in each row with the df of Y_i given by the parametric conditional df $F(\cdot|X_i, \theta_n)$ $(Q_n(\cdot|X_i))$ for $i \leq n$.

THEOREM A.2: *Suppose Assumptions* F.C1, F.C2, F.M1, M2, *and* F.E2 *hold. Then, for any non-random sequence* $\{\theta_n : n \geq 1\}$ *for which* $\theta_n \to \theta_0$, *we have*

(a) $CK_n(\mathcal{Z}) \xrightarrow{d} \sup_{z \in \mathcal{Z}} |\nu(z) - \Delta_0(z)' D_0 \nu_0|$ *under* $\{\theta_n : n \geq 1\}$ *and*

(b) $CK_n \xrightarrow{d} \sup_{z \in \mathcal{Z}} |\nu(z) - \Delta_0(z)' D_0 \nu_0|$ *under* $\{\theta_n : n \geq 1\}$.

The behavior of $\widehat{\theta}$ for arbitrary conditional df $H(\cdot|\cdot)$ is given by

ASSUMPTION F.E3: $\widehat{\theta} \xrightarrow{p} \theta_1$ *for some* $\theta_1 \in \Theta$ *under Assumption* F.D.

Theorem A.2 and the discussion of Section 4 now yield:

COROLLARY A.1: (a) *Suppose Assumptions* F.D, F.C1, F.C2, F.M1, F.M2, *and* F.E2 *hold,* $H(\cdot|\cdot)$ *of Assumption* F.D *equals* $F(\cdot|\cdot, \theta_0)$, *and* $B \to \infty$ *as* $n \to \infty$. *Then,* $c_{\alpha n B}(\widehat{\theta}) \xrightarrow{p} c_\alpha(\theta_0)$, $P_{\theta_0}(CK(\mathcal{Z}) > c_{\alpha n B}(\widehat{\theta})) \to \alpha$, *and* $P_{\theta_0}(CK_n > c_{\alpha n B}(\widehat{\theta})) \to \alpha$.

(b) *Suppose Assumption* F.D *holds, Assumptions* F.M1, F.M2, *and* F.E2 *hold for any value* $\theta_0 \in \Theta$, *and* $B \to \infty$ *as* $n \to \infty$. *Then,* $\sup_{\theta_0 \in \Theta} \lim_{n \to \infty} P_{\theta_0}(CK_n > c_{\alpha n B}(\widehat{\theta})) = \alpha$.

(c) *Suppose Assumptions* F.D *and* F.E3 *hold, Assumptions* F.M1, F.M2, *and* F.E2 *hold for any value* $\theta_0 \in \Theta$, *and* $B \to \infty$ *as* $n \to \infty$. *Then,* $c_{\alpha n B}(\widehat{\theta}) \xrightarrow{p} c_\alpha(\theta_1)$.

To obtain consistency of CK_n in the fixed covariates case, we assume:

ASSUMPTION F.C3: (i) $\sup_{z \in \mathcal{Z}} |\int H(y|x^*)(x^* \leq x)(d\widehat{G}_n(x^*) - dG(x^*))| \to 0$.

(ii) $\int \int (|H(z) - F(z, \theta_1)| > 1/k) dH(y|x)(d\widehat{G}_n(x) - dG(x)) \to 0\ \forall k \geq 1$.

(iii) $\sup_{z \in R^{V+K}} |\int \int (z^* \leq z) f(y^*|x^*, \theta_1) d\mu(y^*)(d\widehat{G}_n(x^*) - dG(x^*)) \to 0$.

THEOREM A.3: *Under Assumptions* F.D, F.C3, F.M1, F.E3, *and* H1, *for all sequences of* rv's $\{c_n : n \geq 1\}$ *with* $c_n = O_p(1)$, *we have* $\lim_{n \to \infty} P(CK_n(\mathcal{Z}) > c_n) = 1$ *and* $\lim_{n \to \infty} P(CK_n > c_n) = 1$.

The local power results when the covariates are fixed assumes:

ASSUMPTION F.CA: (i) $\overline{\lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int h_j(y, X_i)d\mu(y) < \infty$ *for* $j = 1, 2$.

(ii) $\int\int(d(z)/f(y|x, \theta_0))^2 f(y|x, \theta_0)d\mu(y)(d\widehat{G}_n(x) - dG(x)) \to 0$.

(iii) $\sup_{z\in\mathcal{Z}}|\int\int(z^* \leq z)q(y^*|x^*)d\mu(y^*)(d\widehat{G}_n(x^*) - dG(x^*)) \to 0$.

(iv) $\int\int\psi(z, \theta_0)q(y|x)d\mu(y)(d\widehat{G}_n(x) - dG(x)) \to 0$.

THEOREM A.4: *Suppose Assumptions* F.C1, F.M1, F.E1, *and* F.CA *hold. Then,*

(a) $CK_n(\mathcal{Z}) \xrightarrow{d} M$ *under* $\{Q_n(\cdot|\cdot) : n \geq 1\}$ *and* (b) $CK_n \xrightarrow{d} M$ *under* $\{Q_n(\cdot|\cdot) : n \geq 1\}$.

A.2. Proofs of Theorems 1–4 and Lemmas 1 and 2

To prove Theorems 1–4 and Lemma 1, we use the following:

LEMMA A.2: (a) *Under Assumptions* D, M1, *and* E1, $P_G(\{X_i : i \geq 1\}$ *satisfies Assumptions* F.C1, F.M1, *and* F.E1$) = 1$. (b) *Under Assumptions* D, M1, *and* E2, $P_G(\{X_i : i \geq 1\}$ *satisfies Assumptions* F.C1, F.C2, F.M1, *and* F.E2$) = 1$. (c) *Under Assumptions* D, M1, *and* E3, $P_G(\{X_i : i \geq 1\}$ *satisfies Assumptions* F.C3, F.M1, *and* F.E3$) = 1$. (d) *Under Assumptions* M1, E1, *and* CA, $P_G(\{X_i : i \geq 1\}$ *satisfies Assumptions* F.C1, F.M1, F.E1, *and* F.CA$) = 1$. (e) *Under Assumption* M1$'$, $P_G(\{X_i : i \geq 1\}$ *satisfies Assumption* F.M1$'$$) = 1$.

Theorems 1–4 now follow from Theorems A.1–A.4 and Lemma A.2(a)–A.2(d) respectively. Lemma 1 follows from Lemma A.1 and Lemma A.2(e).

The proof of Lemma A.2 uses the following uniform strong law of large numbers:

LEMMA A.3: *Suppose* $\{X_i : i \geq 1\}$ *are* iid *with* df $G(\cdot)$. *Let* $\mathcal{D}_n$ *be a class of functions from* $\text{supp}(G) \subset R^K$ *to* R *of the form* $\mathcal{D}_n = \{d : d(x) = \int a_n(z)(z \leq z_1)dJ_n(y|x)$ *for some* $z_1 \in \overline{R}^{V+K}\}$ *for some function* $a_n(z) : R^{V+K} \to R$ *and some conditional* df $J_n(\cdot|\cdot)$, *where* $\overline{R} = R \cup \{\infty\}$.

(a) *If* $\int[\sup_{n\geq 1}\int|a_n(z)|dJ_n(y|x)]dG(x) < \infty$, *then* $\sup_{d\in\mathcal{D}_n}\left|\frac{1}{n}\sum_{i=1}^{n}d(X_i) - Ed(X_i)\right| \to 0$ wp1.

(b) *If* $\int [\sup_{n\geq 1} \int |a_n(z)| dJ_n(y|x)]^2 dG(x) < \infty$, *then*

$$\sup_{d_1, d_2 \in \mathcal{D}_n} \left| \frac{1}{n} \sum_{i=1}^n d_1(X_i) d_2(X_i) - E d_1(X_i) d_2(X_i) \right| \to 0 \ \text{ wp1.}$$

Proof of Lemma A.3: Let

$$\text{(A.1)} \qquad \begin{aligned} \overline{d}(x) &= \sup_{n\geq 1} \int |a_n(z)| dJ_n(y|x) \,, \\ \mathcal{B}_n &= \{b : b(z) = a_n(z)(z \leq z_1) \text{ for some } z_1 \in \overline{R}^{V+K}\} \,, \quad \text{and} \\ J\widehat{G}_n(z) &= \int\!\!\int (z^* \leq z) dJ_n(y^*|x^*) d\widehat{G}_n(x^*) \,. \end{aligned}$$

Note that $\overline{d}(x)$ is an envelope of $\mathcal{D}_n$ $\forall n \geq 1$ and $J\widehat{G}_n(z)$ is a df. For $\varepsilon > 0$, let $N_1(\varepsilon, \widehat{G}_n, \mathcal{D}_n)$ and $N_1(\varepsilon, J\widehat{G}_n, \mathcal{B}_n)$ denote the $L^1(\widehat{G}_n)$ and $L^1(J\widehat{G}_n)$ cover numbers of $\mathcal{D}_n$ and $\mathcal{B}_n$, respectively, as defined in Pollard (1984, p. 25). If $\int \overline{d}(x) dG(x) < \infty$ and

$$\text{(A.2)} \qquad E N_1^r(\varepsilon, \widehat{G}_n, \mathcal{D}_n) \leq C_\varepsilon \quad \forall \varepsilon > 0$$

for some positive finite constants r and C_ε that do not depend on $\widehat{G}_n$ or n, then the result of part (a) holds by an extension of Thm. II.24 of Pollard (1984, p. 25). The extension is to let the class of functions depend on n. Pollard's proof goes through unchanged except for the added subscript n on the class of functions and the failure of his reverse sub-martingale argument. The latter can be replaced by an application of the Borel–Cantelli Lemma using the last inequality in his proof and the summability of its right-hand side over $n \geq 1$ by (A.2) and the following argument:

$$\text{(A.3)} \qquad \begin{aligned} \sum_{n=1}^\infty P(\log N_1(\varepsilon, \widehat{G}_n, \mathcal{D}_n) > nK) &= \sum_{n=1}^\infty P(N_1^r(\varepsilon, \widehat{G}_n, \mathcal{D}_n) > e^{nrK}) \\ &\leq \sum_{n=1}^\infty E N_1^r(\varepsilon, \widehat{G}_n, \mathcal{D}_n)/e^{nrK} < \infty \end{aligned}$$

for any constant $0 < K < \infty$ using Markov's inequality.

To show (A.2), we establish that $\forall \varepsilon > 0$

$$\text{(A.4)} \qquad N_1(\varepsilon, \widehat{G}_n, \mathcal{D}_n) \leq N_1(\varepsilon, J\widehat{G}_n, \mathcal{B}_n) \,.$$

The inequality holds because a function b_2 approximates $b_1 \in \mathcal{B}_n$ in $L^1(J\widehat{G}_n)$ metric within ε implies that $d_2(x) = \int b_2(z) dJ_n(y|x)$ approximates $d_1(x) = \int b_1(z) dJ_n(y|x) \in \mathcal{D}_n$ in $L^1(\widehat{G}_n)$

metric within ε:

$$\text{(A.5)} \qquad \int |d_1(x) - d_2(x)| d\widehat{G}_n(x) = \int \left| \int (b_1(z) - b_2(z)) dJ_n(y|x) \right| d\widehat{G}_n(x)$$
$$\leq \iint |b_1(z) - b_2(z)| dJ_n(y|x) d\widehat{G}_n(x) = \int |b_1(z) - b_2(z)| dJ\widehat{G}_n(z) < \varepsilon \; .$$

Next, we show that $\forall \varepsilon > 0$

$$\text{(A.6)} \qquad N_1(\varepsilon, J\widehat{G}_n, \mathcal{B}_n) \leq N_1\left(\varepsilon / \int |a_n(z)| dJ\widehat{G}_n(z),\ J\widehat{G}_n^1,\ \mathcal{C}\right)$$
$$\leq A\left(\varepsilon / \int |a_n(z)| dJ\widehat{G}_n(z)\right)^{-W} \leq A\varepsilon^{-W} \left(\tfrac{1}{n} \textstyle\sum_{i=1}^n \overline{d}(X_i)\right)^W \; , \quad \text{where}$$
$$\mathcal{C} = \{(z \leq z_1) \; : \; \text{for some } z_1 \in \overline{R}^{V+K}\} \; ,$$
$$J\widehat{G}_n^1(z) = \int (z^* \leq z) |a_n(z^*)| dJ\widehat{G}_n(z^*) / \int |a_n(z^*)| dJ\widehat{G}_n(z^*) \; ,$$

and A, W are positive finite constants. Note that $J\widehat{G}_n^1(\cdot)$ is a df. The third inequality of (A.6) holds by definition of $J\widehat{G}_n(z)$ and $\overline{d}(x)$. The second inequality of (A.6) holds because $\mathcal{C}$ contains indicator functions of a Vapnik–Cervonenkis class of sets. The first inequality of (A.6) holds because if $c_1(z)$ approximates $c_2(z) \in \mathcal{C}$ within $\varepsilon / \int |a_n(z)| dJ\widehat{G}_n(z)$ in $L^1(J\widehat{G}_n^1)$ metric then $b_1(z) = a_n(z) c_1(z)$ approximates $b_2(z) = a_n(z) c_2(z) \in \mathcal{B}_n$ within ε in $L^1(J\widehat{G}_n)$ metric by the following inequality:

$$\text{(A.7)} \qquad \int |b_1(z) - b_2(z)| dJ\widehat{G}_n(z) = \int |a_n(z)| \cdot |c_1(z) - c_2(z)| dJ\widehat{G}_n(z)$$
$$= \int |a_n(z)| dJ\widehat{G}_n(z) \cdot \int |c_1(z) - c_2(z)| dJ\widehat{G}_n^1(z) < \varepsilon \; .$$

Combining (A.4) and (A.6) gives (A.2) with $r = 1/W$ using the moment condition of part (a).

Next, we establish part (b). Let $\mathcal{D}_n\mathcal{D}_n = \{d_1 d_2 \; : \; d_1 \in \mathcal{D}_n, \ d_2 \in \mathcal{D}_n\}$. By the generalization of Thm. II.24 of Pollard (1984) referred to above and the moment condition of part (b), it suffices to show that for some positive finite constants r and C_ε,

$$\text{(A.8)} \qquad EN_1^r(\varepsilon, \widehat{G}_n, \mathcal{D}_n\mathcal{D}_n) \leq C_\varepsilon \quad \forall \varepsilon > 0 \; .$$

We claim that $\forall \varepsilon > 0$

$$\text{(A.9)} \qquad N_1(\varepsilon, \widehat{G}_n, \mathcal{D}_n\mathcal{D}_n) \leq N_1^2\left(\varepsilon / \int 2\overline{d}(x) d\widehat{G}_n(x),\ D\widehat{G}_n,\ \mathcal{D}_n\right) \; , \quad \text{where}$$
$$D\widehat{G}_n(x) = \int (x^* \leq x) \overline{d}(x^*) d\widehat{G}_n(x^*) / \int \overline{d}(x^*) d\widehat{G}_n(x^*) \; .$$

Note that $D\widehat{G}_n(\cdot)$ is a df. Equation (A.9) holds because if $d_1(x)$ and $d_1^*(x)$ approximate $d_2(x) \in \mathcal{D}_n$ and $d_2^*(x) \in \mathcal{D}_n$, respectively, within $\varepsilon / \int 2\overline{d}(x)d\widehat{G}_n(x)$ in $L^1(D\widehat{G}_n)$ metric, then $d_1(x)d_1^*(x)$ approximates $d_2(x)d_2^*(x) \in \mathcal{D}_n\mathcal{D}_n$ within ε in $L^1(\widehat{G}_n)$ metric by the following inequality:

$$
\text{(A.10)} \qquad \begin{aligned}
& \int |d_1(x)d_1^*(x) - d_2(x)d_2^*(x)|d\widehat{G}_n(x) \\
\leq\ & \int \overline{d}(x)|d_1(x) - d_2(x)|d\widehat{G}_n(x) + \int \overline{d}(x)|d_1^*(x) - d_2^*(x)|d\widehat{G}_n(x) \\
\leq\ & \int \overline{d}(x)d\widehat{G}_n(x) \int |d_1(x) - d_2(x)|dD\widehat{G}_n(x) \\
& + \int \overline{d}(x)d\widehat{G}_n(x) \int |d_1^*(x) - d_2^*(x)|dD\widehat{G}_n(x) < \varepsilon \ .
\end{aligned}
$$

Next, we have $\forall \varepsilon > 0$

$$
\text{(A.11)} \qquad \begin{aligned}
& N_1^2\left(\varepsilon / \int 2\overline{d}(x)d\widehat{G}_n(x),\, D\widehat{G}_n,\, \mathcal{D}_n\right) \leq N_1^2\left(\varepsilon / \int 2\overline{d}(x)d\widehat{G}_n(x),\, JD\widehat{G}_n,\, \mathcal{B}_n\right) \\
\leq\ & N_1^2\left(\varepsilon / \left(\int 2\overline{d}(x)d\widehat{G}_n(x) \int |a_n(z)|dJD\widehat{G}_n(z)\right),\, JD\widehat{G}_n^1,\, \mathcal{C}\right) \\
\leq\ & A^2(\varepsilon/2)^{-2W}\left(\int \overline{d}(x)d\widehat{G}_n(x) \int |a_n(z)|dJD\widehat{G}_n(z)\right)^{2W} \\
\leq\ & A^2(\varepsilon/2)^{-2W}\left(\tfrac{1}{n}\textstyle\sum_{i=1}^n \overline{d}^2(X_i)\right)^{2W} \quad , \text{ where} \\
& JD\widehat{G}_n(z) = \int (z^* \leq z)dJ_n(y^*|x^*)dD\widehat{G}_n(x^*) \ \text{ and} \\
& JD\widehat{G}_n^1(z) = \int (z^* \leq z)|a_n(z)|dJD\widehat{G}_n(z^*) / \int |a_n(z^*)|dJD\widehat{G}_n(z^*) \ .
\end{aligned}
$$

Note that $JD\widehat{G}_n(\cdot)$ and $JD\widehat{G}_n^1(\cdot)$ are df's. The first three inequalities of (A.11) hold by the arguments establishing (A.4), (A.6), and (A.6) respectively. Combining (A.9) and (A.11) yields (A.8) with $r = 1/(2W)$ using the moment condition of part (b). □

PROOF OF LEMMA A.2: First, we prove part (a). Assumption F.C1(i) holds wp1 by the Glivenko–Cantelli Theorem. Assumption F.C1(ii) holds wp1 because

$$
\text{(A.12)} \quad P_G(\operatorname{supp}(\widehat{G}_n) \subset \operatorname{supp}(G)) = P_G(X_i \subset \operatorname{supp}(G)\ \forall i \leq n) = P_G(X_i \subset \operatorname{supp}(G))^n = 1 \ .
$$

Assumption F.C1(iii) holds wp1 by Lemma A.3(a) with $J_n(y|x) = F(y|x, \theta_0)$ and $a_n(z) = 1$. (When $\psi(z, \theta_0)$ is vector-valued, as it usually is, Lemma A.3(a) is applied repeatedly element by element.) Assumption F.C1(iv) holds wp1 by several applications of Lemma A.3(a) with

$J_n(y|x) = F(y|x, \theta_0)$, $a_n(z) = 1$, $a_n(z) = \psi(z, \theta_0)$, $a_n(z) = \int \psi(y^*, x, \theta_0) f(y^*|x, \theta_0) d\mu(y^*)$, and $a_n(z) = \int \psi(y^*, x, \theta_0) f(y^*|x, \theta_0) d\mu(y^*) \int \psi(y^*, x, \theta_0)' f(y^*|x, \theta_0) d\mu(y^*)$, and by one application of Lemma A.3(b) with $J_n(y|x) = F(y|x, \theta_0)$ and $a_n(z) = 1$. Note that a *uniform* SLLN is used here, even though Assumption F.C1(iv) only requires pointwise convergence, because it ensures that pointwise convergence holds for all z_1, z_2 in a (possibly) uncountable set $\mathcal{Z} \times \mathcal{Z}$ on a single set of realizations of $\{X_i : i \geq 1\}$ that has probability one. Assumption F.C1(v) holds wp1 by the SLLN for bounded iid rv's. Assumption F.M1 holds wp1 by Assumption M1. Assumption F.E1 holds wp1 by Assumption E1 and the SLLN applied to $\frac{1}{n}\sum_{i=1}^{n} \psi_0^*(X_i)$.

Next, we prove part (b). Assumptions F.C1 and F.M1 hold wp1 by the argument of part (a). Assumption F.C2(i) holds wp1 by the same argument as for Assumption F.C1(iv) above, but with θ_0 replaced by θ_n throughout. Assumption F.C2(ii) holds by a countable number of applications of the SLLN for triangular arrays of bounded inid rv's. Assumption F.E2 holds wp1 by Assumption E2 and the SLLN applied to $\frac{1}{n}\sum_{i=1}^{n} \psi_1^*(X_i)$.

We now prove part (c). Assumption F.C3(i) holds wp1 by Lemma A.3(a) with $J_n(y|x) = H(y|x)$ and $a_n(z) = 1$. Assumption F.C3(ii) holds wp1 by a countable number of applications of the SLLN for bounded iid rv's. Assumption F.C3(iii) holds wp1 by Lemma A.3(a) with $J_n(y|x) = F(y|x, \theta_1)$ and $a_n(z) = 1$. Assumptions F.M1 and F.E3 hold wp1 by Assumptions M1 and E3.

For part (d), Assumptions F.C1, F.M1, and F.E1 hold by the argument above for part (a) using Assumptions M1 and E1. Assumptions F.CA(i), F.CA(ii), and F.CA(iv) hold wp1 by Kolmogorov's SLLN given the moment conditions specified in Assumption CA. Assumption F.CA(iii) holds by Lemma A.3(a) with $J_n(y|x) = Q(y|x)$ and $a_n(z) = 1$.

For part (e), Assumption F.M1′(i) holds by Assumption M1′(i), Assumption F.M1′(ii) holds by Assumption M1′(ii) and Kolmogorov's SLLN, and Assumption F.M1′(iii) holds by Lemma A.3(a) with $J_n(y|x) = F(y|x, \theta_0)$ and $a_n(z) = s(y|x, \theta_0)$. □

PROOF OF LEMMA 2: Assumptions M2(i) and M2(ii) hold by the dominated convergence theorem given the continuity at θ_0 of $f(y|x, \theta_0)$, $\psi(z, \theta_0)$, and $F(y|x, \theta_0)$ $\forall z \in \mathcal{Z}$ and the moment condition of Assumption M2′(ii). □

A.3. Proofs of Lemma A.1 and Theorems A.1–A.4

PROOF OF LEMMA A.1: Assumption F.M1(i) holds with $\frac{\partial}{\partial\theta}F(y|X_i,\theta) = \int (y^* \le y) s(y^*|X_i,\theta)$ $f(y^*|X_i,\theta)d\mu(y^*)$ by a standard argument using a mean-value expansion of $f(y^*|X_i,\theta)$ and finiteness of $s_1^*(X_i)$ $\forall i \ge 1$. Assumption F.M1(ii) follows from

$$
\begin{aligned}
\text{(A.13)} \qquad & \sup_{z\in R^{V+K}} \sup_{\theta:\|\theta-\theta_0\|\le r_n} \left| \tfrac{1}{n}\sum_{i=1}^n \tfrac{\partial}{\partial\theta}F(y|X_i,\theta)(X_i \le x) - \Delta_0(x)\right| \\
\le\ & \sup_{z\in R^{V+K}} \sup_{\theta:\|\theta-\theta_0\|\le r_n} \left| \iint (z^* \le z)[s(y^*|x^*,\theta)f(y^*|x^*,\theta) \right. \\
& \left. - s(y^*|x^*,\theta_0)f(y^*|x^*,\theta_0)]d\mu(y^*)d\widehat{G}_n(x^*)\right| \\
& + \sup_{z\in R^{V+K}} \left| \iint (z^* \le z)s(y^*|x^*,\theta_0)f(y^*|x^*,\theta_0)d\mu(y^*)(d\widehat{G}_n(x^*) - dG(x^*))\right| \\
\le\ & \tfrac{1}{n}\textstyle\sum_{i=1}^n (s_1^*(X_i) + s_2^*(X_i))r_n + o(1) = o(1)\ ,
\end{aligned}
$$

where the first inequality holds by the triangle inequality, the second inequality holds using a mean-value expansion and Assumption F.M1$'$(iii), and the equality holds by Assumption F.M1$'$(ii) and the definition of r_n.

The first part of Assumption F.M1(iii) holds because $\sup_{z\in R^{V+K}} \|\Delta_0(z)\| \le (\int s_1^*(x)dG(x))^{1/2}$ $< \infty$ by Assumption F.M1$'$(ii). The second part of Assumption F.M1(iii) (i.e., uniform continuity of $\Delta_0(\cdot)$) holds, because

$$
\begin{aligned}
\text{(A.14)} \qquad & \|\Delta_0(z_1) - \Delta_0(z_2)\| = \left\| \iint [(z \le z_1) - (z \le z_2)]s(y|x,\theta_0)f(y|x,\theta_0)d\mu(y)dG(x)\right\| \\
\le\ & \left(\iint [(z \le z_1) - (z \le z_2)]^2 f(y|x,\theta_0)d\mu(y)dG(x)\right)^{1/2} \\
& \times \left(\iint \|s(y|x,\theta_0)\|^2 f(y|x,\theta_0)d\mu(y)dG(x)\right)^{1/2}
\end{aligned}
$$

by the Cauchy–Schwarz inequality. The first multiplicand on the right-hand side equals $\rho(z_1, z_2)$ and the second multiplicand is finite by Assumption F.M1$'$(ii). □

PROOF OF THEOREM A.1: The proof of Theorem A.1 is the same as that of Theorem A.2 below with θ_n replaced by θ_0 throughout, which allows Assumptions F.C2(i) and F.M2(i), F.C2(ii) and F.M2(ii), and F.E2 to be replaced by Assumptions F.C1(iv), F.C1(v), and F.E1 respectively. □

The proof of Theorem A.2 uses the following Lemmas.

LEMMA A.4: *Suppose Assumptions* F.D, F.M1, *and* F.E2 *hold. Then, for any non-random sequence* $\{\theta_n : n \geq 1\}$ *for which* $\theta_n \to \theta_0$, *we have*

$$\sqrt{n} \sup_{z \in \mathcal{Z}} |\widehat{F}_n(z, \widehat{\theta}) - \widehat{F}_n(z, \theta_n) - \Delta_0(z)' D_0 \overline{\psi}_n(\theta_n)| = o_p(1) \text{ under } \{\theta_n : n \geq 1\} .$$

LEMMA A.5: *Suppose Assumptions* F.D, F.C1, F.C2, F.M1, F.M2, *and* F.E2 *hold. Then, for any non-random sequence* $\{\theta_n : n \geq 1\}$ *for which* $\theta_n \to \theta_0$, *we have* $\binom{\nu_n(\cdot, \theta_n)}{\sqrt{n}\,\overline{\psi}_n(\theta_n)} \Rightarrow \binom{\nu(\cdot)}{\nu_0}$ *under* $\{\theta_n : n \geq 1\}$, *where* $(\nu_n(\cdot, \theta_n), \sqrt{n}\,\overline{\psi}_n(\theta_n)')$ *is a stochastic process indexed by* $z \in \mathcal{Z}$ *and the pseudometric* ρ *on* $\mathcal{Z}$ *is as defined in Section* 3.

LEMMA A.6: *Suppose* (a) $P(Z \in A) > 0$ *for some* $A \subset R^{V+K}$ *and* (b) $\frac{1}{n}\sum_{i=1}^{n} P(Z_{ni} \in A) \to P(Z \in A)$ for some triangular array $\{Z_{ni} : i \leq n, n \geq 1\}$ of rv's *that are independent within rows. Then,* $P(Z_{ni} \in A$ *for some* $i \leq n) \to 1$.

PROOF OF THEOREM A.2: Theorem A.2(a) follows from Lemma A.4, Lemma A.5, and the continuous mapping theorem, see Pollard (1984, Thm. IV.12, p. 70). The latter applies because the function $h(\nu(\cdot), \lambda) = \sup_{z \in \mathcal{Z}} |\nu(z) - \Delta_0(z)'\lambda|$ is a continuous function of $(\nu(\cdot), \lambda)$, where $\lambda \in R^L$ and $\nu(\cdot)$ is an element of the space of uniformly continuous functions on $\mathcal{Z}$ equipped with the sup norm, using the fact that $\sup_{z \in R^{V+K}} \|\Delta_0(z)\| < \infty$ by Assumption F.M1(iii).

Next, we prove Theorem A.2(b). Since $P_{\theta_n}(Z_i \in \mathcal{Z}\ \forall i \leq n) = 1$ by Assumption F.C1(ii),

$$\text{(A.15)} \qquad \begin{aligned} \overline{\lim_{n \to \infty}}\, P_{\theta_n}(\max_{i \leq n} |\nu_n(Z_i, \widehat{\theta})| > c) &\leq \underline{\lim}_{n \to \infty} P_{\theta_n}(\sup_{z \in \mathcal{Z}} |\nu_n(z, \widehat{\theta})| > c) \\ &= P(\sup_{z \in \mathcal{Z}} |\nu(z) - \Delta_0(z)' D_0 \nu_0| > c) \end{aligned}$$

for all $c \in R$, where the equality holds by Theorem A.2(a). It remains to show that (A.15) holds with the inequality reversed.

We use Lemma A.6 to show that $\forall k \geq 1$, $\forall z \in \mathcal{Z}_d$ (defined in Assumption F.C1(v)),

$$\text{(A.16)} \qquad \lim_{n \to \infty} P_{\theta_n}(Z_i \in B(z, 1/k) \text{ for some i} \leq \text{n}) = 1 .$$

Condition (a) of Lemma A.6 hold since $P_{\theta_0}(Z \in B(z, 1/k)) > 0$. Condition (b) of Lemma A.6 holds by Assumption F.C2(ii) and F.M2(ii) and the triangle inequality: $\forall z \in \mathcal{Z}_d$, $\forall k \geq 1$,

$$\text{(A.17)}\qquad \begin{aligned} & \left|\tfrac{1}{n}\textstyle\sum_{i=1}^{n} P_{\theta_n}(Z_i \in B(z,\, 1/k)) - P_{\theta_0}(Z \in B(z,\, 1/k))\right| \\ \le\ & \left|\tfrac{1}{n}\textstyle\sum_{i=1}^{n} P_{\theta_n}(Z_i \in B(z,\, 1/k)) - P_{\theta_n}(Z \in B(z,\, 1/k))\right| \\ & + |P_{\theta_n}(Z \in B(z,\, 1/k)) - P_{\theta_0}(Z \in B(z,\, 1/k))| \to 0 \ , \end{aligned}$$

where $Z \sim F(\cdot,\, \theta_0)$ or $Z \sim F(\cdot,\, \theta_n)$ as indicated.

Equation (A.16) implies that for all $k \geq 1$ and for any $z_j \in \mathcal{Z}_d$ for $j = 1, \ldots, J$,

$$\text{(A.18)}\qquad \begin{aligned} & \lim_{n\to\infty} P_{\theta_n}(\max_{i\le n} |\nu_n(Z_i,\, \widehat{\theta})| > c) \\ =\ & \lim_{n\to\infty} P_{\theta_n}(\max_{i\le n} |\nu_n(Z_i,\, \widehat{\theta})| > c,\ \cap_{j=1}^{J}(Z_i \in B(z_j,\, 1/k) \text{ for some } i \le n)) \\ \ge\ & \overline{\lim_{n\to\infty}}\ P_{\theta_n}(\max_{j\le J}\ \inf_{z\in B(z_j,1/k)} |\nu_n(z, \widehat{\theta})| > c) = P(\max_{j\le J}\ \inf_{z\in B(z_j,1/k)} |\nu^*(z)| > c) \ , \end{aligned}$$

where $\nu^*(z) = \nu(z) - \Delta_0(z)' D_0 \nu_0$ and the last equality uses the continuous mapping theorem and Lemma A.5. Thus,

$$\text{(A.19)}\qquad \begin{aligned} & \lim_{n\to\infty} P_{\theta_n}(\max_{i\le n} |\nu_n(Z_i,\, \widehat{\theta})| > c) \ge D \ , \quad \text{where} \\ & D = \sup_{J\ge1}\ \sup_{k\ge1}\ \max_{\substack{\{z_1,\ldots,z_J\}:\\ z_j\in\mathcal{Z}_d \forall j\le J}} P(\max_{j\le J}\ \inf_{z\in B(z_j,1/k)} |\nu^*(z)| > c) \ . \end{aligned}$$

It remains to show that $D = P(\sup_{z\in\mathcal{Z}} |\nu^*(z)| > c)$. We have

$$\text{(A.20)}\qquad D \le \sup_{J\ge1}\ \max_{\substack{\{z_1,\ldots,z_J\}:\\ z_j\in\mathcal{Z}_d \forall j\le J}} P(\max_{j\le J} |\nu^*(z_j)| > c) \le P(\sup_{z\in\mathcal{Z}} |\nu^*(z)| > c) \ .$$

To obtain the reverse inequality, we use the fact that R^{V+K}, and hence, $\mathcal{Z}$ is totally bounded under the metric ρ (see the proof of Lemma A.5 below). In consequence, given any $k \geq 1$, $\mathcal{Z}$ can be covered by a finite number of balls, say J balls, of radius $1/k$ centered at points $z_1, \ldots, z_J$ in the countably dense subset $\mathcal{Z}_d$ of $\mathcal{Z}$. The sample paths of $\nu^*(\cdot)$ are uniformly continuous (with respect to ρ) almost everywhere. This follows because the same is true for $\nu(\cdot)$ by the functional CLT of Pollard (1990, Thm. 10.6) applied in Lemma A.5 below and $\Delta_0(\cdot)$ is uniformly continuous by Assumption F.M1(iii). In consequence, by the bounded convergence theorem, given any $\eta > 0$ there exists a $k_\eta \geq 1$ such that $P(\sup_{z_1,z_2:\rho(z_1,z_2)<1/k_\eta} |\nu^*(z_1) - \nu^*(z_2)| > \eta) < \eta$. Thus, we get

$$\text{(A.21)} \qquad P(\sup_{z\in\mathcal{Z}} |\nu^*(z)| > c) = \lim_{\eta\to 0} P(\sup_{z\in\mathcal{Z}} |\nu^*(z)| > c+\eta)$$
$$= \lim_{\eta\to 0} P(\max_{j\le J} \sup_{z\in B(z_j,1/k_\eta)} |\nu^*(z)| > c+\eta)$$
$$\le \lim_{\eta\to 0}[P(\max_{j\le J} \sup_{z\in B(z_j,1/k_\eta)} |\nu^*(z)| > c+\eta, \quad \sup_{\rho(z_1,z_2)<1/k_\eta} |\nu^*(z_1)-\nu^*(z_2)| < \eta) + \eta]$$
$$\le \lim_{\eta\to 0} P(\max_{j\le J} \inf_{z\in B(z_j,1/k_\eta)} |\nu^*(z)| > c) \le D\ ,$$

where the first equality holds because $\sup_{z\in\mathcal{Z}} |\nu^*(z)|$ is a continuous rv (see Lifshits (1982)). □

Proof of Lemma A.4: By Assumption F.E2, $\sqrt{n}(\widehat{\theta} - \theta_n) = O_p(1)$, because the mean-squared error of $\frac{1}{n}\sum_{i=1}^n D_0\psi(Z_i, \theta_n)$ is $O(1)$ under $\{\theta_n : n \ge 1\}$.

Using Assumption F.M1(i), a mean-value expansion gives

$$\text{(A.22)} \qquad \widehat{F}_n(z, \widehat{\theta}) = \widehat{F}_n(z, \theta_n) + \tfrac{1}{n}\textstyle\sum_{i=1}^n \frac{\partial}{\partial\theta'}F(y|X_i, \widehat{\theta}_n(z))(X_i \le x)(\widehat{\theta} - \theta_n)\ ,$$

where $\widehat{\theta}_n(z)$ lies on the line segment joining $\widehat{\theta}$ and θ_n.

Now, $\widehat{\theta} - \theta_0 = o_p(1)$ implies that there exists a sequence of constants $\{r_n^* : n \ge 1\}$ such that $r_n^* \to 0$ and $P(\|\widehat{\theta} - \theta_0\| > r_n^*) \to 0$. Given the definition of $\widehat{\theta}_n(z)$, this gives $P(\sup_{z\in\mathcal{Z}} \|\widehat{\theta}_n(z) - \theta_0\| > r_n^*) \to 0$. Let

$$\text{(A.23)} \qquad R_n = \sup_{z\in\mathcal{Z}} \left| \tfrac{1}{n}\textstyle\sum_{i=1}^n \frac{\partial}{\partial\theta}F(y|X_i, \widehat{\theta}_n(z))(X_i \le x) - \Delta_0(z)\right| \quad \text{and}$$
$$\varepsilon_n = \sup_{z\in\mathcal{Z}} \sup_{\theta:\|\theta-\theta_0\|\le r_n^*} \left| \tfrac{1}{n}\textstyle\sum_{i=1}^n \frac{\partial}{\partial\theta}F(y|X_i, \theta)(X_i \le x) - \Delta_0(z)\right|\ .$$

By construction, $P(R_n \le \varepsilon_n) \to 1$. And, by Assumption F.M1(ii), $\varepsilon_n \to 0$. In consequence, $R_n = o_p(1)$.

We now obtain the desired result

$$\text{(A.24)} \qquad \sqrt{n}\sup_{z\in\mathcal{Z}} |\widehat{F}_n(z, \widehat{\theta}) - \widehat{F}_n(z, \theta_n) - \Delta_0(z)' D_0 \overline{\psi}_n(\theta_n)|$$
$$\le \sqrt{n}\sup_{z\in\mathcal{Z}} \left| \tfrac{1}{n}\textstyle\sum_{i=1}^n \frac{\partial}{\partial\theta'}F(y|X_i, \widehat{\theta}_n(z))(X_i \le x)(\widehat{\theta} - \theta_n) - \Delta_0(z)' D_0\overline{\psi}_n(\theta_n)\right|$$
$$\le R_n\sqrt{n}\|\widehat{\theta} - \theta_n\| + \sup_{z\in\mathcal{Z}} |\Delta_0(z)'(\sqrt{n}(\widehat{\theta} - \theta_n) - \sqrt{n}D_0\overline{\psi}_n(\theta_n))| = o_p(1)\ ,$$

where the first inequality holds by (A.22), the second holds by the triangle inequality, and the equality holds by the results above, Assumption F.M1(iii), and Assumption F.E2(i). □

PROOF OF LEMMA A.5: Using Theorem 10.2 of Pollard (1990), Lemma A.5 follows from (a) the total boundedness of $\mathcal{Z}$ under the pseudometric ρ, (b) the stochastic equicontinuity (with respect to ρ) of $\nu_n(\cdot, \theta_n)$ on $\mathcal{Z}$, and (c) the convergence in distribution of each of the finite dimensional (fidi) distributions of $(\nu_n(\cdot, \theta_n), \sqrt{n}\,\overline{\psi}_n(\theta_n))$ to the corresponding fidi distributions of $(\nu(\cdot), \nu_0)$.

Conditions (a) and (b) above can be verified by using the functional CLT of Pollard (1990, Thm. 10.6, p. 53) applied to $\nu_n(\cdot, \theta_n)$, since total boundedness of $\mathcal{Z}$ is part of the conclusion of Pollard's Thm. 10.6 and since a functional CLT implies stochastic equicontinuity (see the converse part of Thm. 10.2 of Pollard (1990)). Pollard's functional CLT has five conditions (i)–(v). His "manageability" condition (i) holds because $\{g(z) : g(z) = (z \leq z_1) \text{ for some } z_1 \in R^{V+K}\}$ consists of indicator functions of sets in a Vapnik–Cervonenkis class of sets. His condition (ii) holds by Assumptions F.C2(i) and F.M2(i). His conditions (iii) and (iv) hold since one can take $F_{ni} = 2/\sqrt{n}\ \forall i, n$. His condition (v) requires that $\lim_{n\to\infty} \rho_n(z_1, z_2) = \rho(z_1, z_2)\ \forall z_1, z_2 \in \mathcal{Z}$ and, for all deterministic sequences $\{z_{1n} : n \geq 1\}$ and $\{z_{2n} : n \geq 1\}$, if $\rho(z_{1n}, z_{2n}) \to 0$ then $\rho_n(z_{1n}, z_{2n}) \to 0$, where

$$\text{(A.25)} \qquad \rho_n(z_1, z_2) = \left(\iint [(z \leq z_1) - (z \leq z_2)]^2 f(y|x, \theta_n) d\mu(y) d\widehat{G}_n(x)\right)^{1/2} .$$

Define $\rho_n^0(z_1, z_2)$ to equal $\rho_n(z_1, z_2)$ with θ_n replaced by θ_0. Note that

$$\text{(A.26)} \qquad \rho_n(z_1, z_2)^2 = \widehat{F}_n(z_1, \theta_n) + \widehat{F}_n(z_2, \theta_n) - 2\widehat{F}_n(z_3, \theta_n) ,$$

where $z_3 = \min\{z_1, z_2\}$ (element-by-element). By the argument used to establish (A.24), $\sup_{z\in\mathcal{Z}} |\widehat{F}_n(z, \theta_n) - \widehat{F}_n(z, \theta_0)| = o(1)$. In consequence,

$$\text{(A.27)} \qquad \sup_{z_1, z_2 \in \mathcal{Z}} |\rho_n(z_1, z_2)^2 - \rho_n^0(z_1, z_2)^2| = o(1) .$$

Assumption F.C1(iii) implies $\lim_{n\to\infty} \rho_n^0(z_1, z_2) = \rho(z_1, z_2)$. This and (A.27) establish the first part of Pollard's condition (v). To establish the second part, assume $\rho(z_{1n}, z_{2n}) \to 0$. Then,

$$\begin{aligned}\text{(A.28)} \quad \rho_n(z_{1n}, z_{2n}) &\leq |\rho_n(z_{1n}, z_{2n}) - \rho_n^0(z_{1n}, z_{2n})| + |\rho_n^0(z_{1n}, z_{2n}) - \rho(z_{1n}, z_{2n})| + \rho(z_{1n}, z_{2n}) \\ &\leq \sup_{z_1, z_2 \in \mathcal{Z}} |\rho_n(z_1, z_2) - \rho_n^0(z_1, z_2)| + \sup_{z_1, z_2 \in \mathcal{Z}} |\rho_n^0(z_1, z_2) - \rho(z_1, z_2)| + \rho(z_{1n}, z_{2n}) \\ &= o(1) ,\end{aligned}$$

where the equality uses (A.27) and Assumption F.C1(iii). Thus, Pollard's condition (v) holds, which completes the proof of the functional CLT for $\nu_n(\cdot, \theta_n)$.

We now verify condition (c) above. We need to show that $(\nu_n(z_1, \theta_n), ..., \nu_n(z_J, \theta_n), \sqrt{n}\overline{\psi}_n(\theta_n)')'$ converges in distribution to $(\nu(z_1), ..., \nu(z_J), \nu_0')'$ $\forall z_j \in \mathcal{Z}$, $\forall j \leq J$, $\forall J \geq 1$. We establish this result using the Cramer–Wold device and a CLT for triangular arrays of row-wise independent mean zero rv's, e.g., see Hall and Hyde (1980, Cor. 3.3.1, p. 58). It suffices to verify (i) the Lindeberg condition and (ii) the convergence of covariance matrices, i.e., $C(z_1, z_2, \theta_n, \widehat{G}_n) \to C(z_1, z_2, \theta_0, G)$ $\forall z_1, z_2 \in \mathcal{Z}$. Condition (i) holds using the boundedness of indicator functions and the Liapunov moment condition on $\psi(Z_i, \theta_n)$ specified by Assumption F.E2(ii). Condition (ii) holds by Assumptions F.C2(i) and F.M2(i). □

PROOF OF LEMMA A.6: Let $\gamma = P(Z \in A) > 0$ and $p_{ni} = P(Z_{ni} \in A^c)$, where A^c denotes the complement of A. Let r_n be the number of p_{ni}'s (for $i \leq n$) for which $p_{ni} \leq 1-\gamma$. By assumption (b), $\frac{1}{n}\sum_{i=1}^n p_{ni} \to 1-\gamma$, and hence, $r_n \to \infty$. In consequence, $\Pi_{i=1}^n p_{ni} \leq (1-\gamma)^{r_n} \to 0$. Using this result, we obtain

$$\text{(A.29)} \qquad \lim_{n\to\infty} P(Z_{ni} \in A \text{ for some } i \leq n) = 1 - \lim_{n\to\infty} \mathrm{P}(\cap_{i=1}^n (\mathrm{Z_{ni}} \subset \mathrm{A^c}))$$
$$= 1 - \lim_{n\to\infty} \Pi_{i=1}^n p_{ni} = 1 \ . \ \square$$

PROOF OF THEOREM A.3: Let $H_n(z) = \frac{1}{n}\sum_{i=1}^n H(y|X_i)(X_i \leq x)$. We have

$$\text{(A.30)} \qquad \sup_{z\in\mathcal{Z}} |\widehat{H}_n(z) - H_n(z)| = o_p(1)\,, \ \sup_{z\in\mathcal{Z}} |H_n(z) - H(z)| = o(1)\,,$$
$$\sup_{z\in\mathcal{Z}} |\widehat{F}_n(z,\widehat{\theta}) - \widehat{F}_n(z, \theta_1)| = o_p(1)\,, \text{ and } \sup_{z\in\mathcal{Z}} |\widehat{F}_n(z, \theta_1) - F(z, \theta_1)| = o(1)\,,$$

where the first result holds by an inid empirical process uniform weak LLN for a Vapnik–Cervonenkis class of sets (i.e., the Glivenko–Cantelli Theorem for inid rv's) since $H_n(z) = E\widehat{H}_n(z)$, the second and fourth results hold by Assumptions F.C3(i) and F.C3(iii), respectively, and the third result holds by a slight alteration of the proof of Lemma A.4 with θ_n replaced by θ_1 throughout and Assumption F.E2 replaced by Assumption F.E3.

Combining these results gives

$$\text{(A.31)}\ \frac{1}{\sqrt{n}} CK_n = \max_{j \le n} |\widehat{H}_n(Z_j) - \widehat{F}_n(Z_j, \widehat{\theta})|$$
$$= \max_{j \le n} |[\widehat{H}_n(Z_j) - H_n(Z_j)] + [H_n(Z_j) - H(Z_j)]$$
$$+ [H(Z_j) - F(Z_j, \theta_1)] + [F(Z_j, \theta_1) - \widehat{F}_n(Z_j, \theta_1)] + [\widehat{F}_n(Z_j, \theta_1) - \widehat{F}_n(Z_j, \widehat{\theta})]|$$
$$= \max_{j \le n} |H(Z_j) - F(Z_j, \theta_1)| + o_p(1) .$$

By Assumption H1, there exists a point $z \in \mathcal{Z}$ for which

$$\text{(A.32)} \qquad H(z) - F(z, \theta_1) = \int (H(y|x^*) - F(y|x^*, \theta_1))(x^* \le x) dG(x^*) \neq 0 .$$

This implies $P_H(H(Z) \neq F(Z, \theta_1)) > 0$, where $Z \sim H$. In turn, this implies that $\exists k \ge 1$ for which $P_H(|H(Z) - F(Z, \theta_1)| > 1/k) > 0$. To see why, suppose not. Then, we get a contradiction:

$$\text{(A.33)}\ 0 = \lim_{k \to \infty} P_H(|H(Z) - F(Z, \theta_1)| > 1/k) = \int \lim_{k \to \infty} (|H(z) - F(z, \theta_1)| > 1/k) dH(z)$$
$$= \int (|H(z) - F(z, \theta_1)| > 0) dH(z) = P_H(H(Z) \neq F(Z, \theta_1)) .$$

Now, let $A = \{z \in R^{V+K} : |H(z) - F(z, \theta_1)| > 1/k\}$. By the inequality above, $P_H(Z \in A) > 0$. And by Assumption F.C3(ii), $\frac{1}{n} \sum_{i=1}^n P(Z_i \in A) \to P_H(Z \in A)$. Hence, by Lemma A.6,

$$\text{(A.34)} \qquad 1 = \lim_{n \to \infty} P(Z_j \in A \text{ for some } \mathrm{j} \le \mathrm{n}) = \lim_{\mathrm{n} \to \infty} \mathrm{P}(\max_{\mathrm{j} \le \mathrm{n}} |\mathrm{H}(\mathrm{Z_j}) - \mathrm{F}(\mathrm{Z_j}, \theta_1)| > 1/\mathrm{k}) .$$

Combining (A.31) and (A.34) gives the desired result:

$$\text{(A.35)} \qquad 1 = \lim_{n \to \infty} P\left(\frac{1}{\sqrt{n}} CK_n > 1/k\right) \le \lim_{n \to \infty} P(CK_n > c_n) . \quad \square$$

Proof of Theorem A.4: First, we establish that the distribution of $\{Z_i : i \le n\}$ when $\{Y_i : i \le n\}$ are independent with conditional distribution $Q_n(\cdot|X_i)$ of Y_i given X_i and X_i are non-random is contiguous to the distribution of $\{Z_i : i \le n\}$ when $\{Y_i : i \le n\}$ are independent with conditional distribution $F(\cdot|X_i, \theta_0)$ of Y_i given X_i and X_i are non-random. It suffices to show that the log likelihood ratio $LR_n = \sum_{i=1}^n (\log q_n(Y_i|X_i) - \log f(Y_i|X_i, \theta_0))$ converges in distribution under $F(\cdot|\cdot, \theta_0)$ to a rv Z for which $E \exp(Z) = 1$. (For example, see Strasser (1985), Thms. 16.8

and 18.11.) A two-term Taylor expansion of $\log q_n(Y_i|X_i) = \log(f(Y_i|X_i,\, \theta_0) + d(Z_i)/\sqrt{n})$ about $\log f(Y_i|X_i,\, \theta_0)$ gives

$$\text{(A.36)}\qquad \left|LR_n - \textstyle\sum_{i=1}^n (1/f(Y_i|X_i,\theta_0))d(Z_i)/\sqrt{n} + \frac{1}{2}\sum_{i=1}^n (1/f(Y_i|X_i,\theta_0))^2 d^2(Z_i)/n\right|$$
$$\leq \tfrac{1}{6}\textstyle\sum_{i=1}^n \sup_{\lambda\in[0,\delta]} 2|1/(f(Y_i|X_i,\theta_0)+\lambda d(Z_i))|^3 \cdot |d(Z_i)|^3/n^{3/2} = o_p(1)\,,$$

where the equality uses Markov's inequality and Assumption F.CA(i).

Next, we have $E_F d(Z_i)/f(Y_i|X_i,\theta_0) = \int d(y,\, X_i)d\mu(y) = 0$ for all $i \geq 1$, where F abbreviates $F(\cdot|\cdot,\, \theta_0)$. In addition, the Lyapunov condition $\overline{\lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^n E_F(d(Z_i)/f(Y_i|X_i,\theta_0))^3 < \infty$ holds by Assumption F.CA(i) and $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \mathrm{Var}_F(d(Z_i)/f(Y_i|X_i,\theta_0)) \to E_{F_0}(d(Z)/f(Y|X,\theta_0))^2$ by Assumption F.CA(ii), where E_{F_0} denotes expectation when $Z \sim F(\cdot,\, \theta_0)$. In consequence, by the CLT for independent mean zero finite variance rv's,

$$\text{(A.37)}\qquad \tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n d(Z_i)/f(Y_i|X_i,\theta_0) \xrightarrow{d} N\left(0,\, E_{F_0}(d(Z)/f(Y|X,\theta_0))^2\right)$$

under $F(\cdot|\cdot,\, \theta_0)$. Furthermore, by the WLLN for inid rv's and Assumption F.CA(i),

$$\text{(A.38)}\qquad -\tfrac{1}{2n}\textstyle\sum_{i=1}^n (d(Z_i)/f(Y_i|X_i,\theta_0))^2 \xrightarrow{p} -\frac{1}{2}E_{F_0}(d(Z)/f(Y|X,\theta_0))^2\,.$$

Combining (A.36), (A.37), and (A.38) gives

$$\text{(A.39)}\qquad LR_T \xrightarrow{d} N\left(-\tfrac{1}{2}E_{F_0}(d(Z)/f(Y|X,\theta_0))^2,\ E_{F_0}(d(Z)/f(Y|X,\theta_0))^2\right)$$

under $F(\cdot|\cdot,\, \theta_0)$. Since $E\exp(Z) = 1$ when $Z \sim N\left(-\frac{1}{2}\sigma^2,\, \sigma^2\right)$, contiguity holds.

Now, to prove part (a) of Theorem A.4, we note that the result of Lemma A.4 holds under $\{Q_n(\cdot|\cdot) : n \geq 1\}$ by contiguity. In place of Lemma A.5, it suffices to show that

$$\text{(A.40)}\qquad \begin{pmatrix} \nu_n(\cdot,\, \theta_0) \\ \sqrt{n}\,\overline{\psi}_n(\theta_0) \end{pmatrix} \Rightarrow \begin{pmatrix} \nu(\cdot) + \sqrt{n_0}\int (Q(\cdot|x^*) - F(\cdot|x^*,\, \theta_0))(x^* \leq \cdot)dG(x^*) \\ \nu_0 + \sqrt{n_0}\int\int \psi(z^*,\, \theta_0)q(y^*|x^*)d\mu(y^*)dG(x^*) \end{pmatrix}$$

under $\{Q_n(\cdot|\cdot) : n \geq 1\}$. This follows by a modification of the proof of Lemma A.5. In particular, θ_n is replaced by θ_0 and the data are distributed under $\{Q_n(\cdot|\cdot) : n \geq 1\}$ rather than under $\{F(\cdot|\cdot,\, \theta_n) : n \geq 1\}$. Pollard's condition (ii) with $C(z_1,\, z_2,\, \theta_0,\, \widehat{G}_n)$ defined with $f(\cdot|\cdot,\theta_0)$ replaced by $q_n(\cdot|\cdot)$ is verified by Assumptions F.C1(iv) and F.CA(i) by straightforward calculations.

Pollard's condition (v) is verified with $\rho_n(z_1, z_2)$ replaced by $\rho_n^*(z_1, z_2)$, where $\rho_n^*(z_1, z_2)$ equals $\rho_n(z_1, z_2)$ (defined in (A.25)) with $f(y|x, \theta_n)$ replaced by $q_n(y|x)$. We have

$$\text{(A.41)} \qquad \sup_{z_1, z_2 \in \mathcal{Z}} |\rho_n^*(z_1, z_2)^2 - \rho_n^0(z_1, z_2)^2| \leq \int\int |q_n(y|x) - f(y|x, \theta_0)| d\mu(y) d\widehat{G}_n(x)$$
$$\leq \sqrt{\tfrac{n_0}{n}} \int\int |q(y|x) - f(y|x, \theta_0)| d\mu(y) d\widehat{G}_n(x) \leq 2\sqrt{\tfrac{n_0}{n}} = o(1) \ .$$

Using (A.41) in place of (A.27), the rest of the verification of condition (v) given in the proof of Lemma A.5 holds.

Verification of condition (c) of the proof of Lemma A.5 requires the alterations that (1) the limit of the means of $\nu_n(z, \theta_0)$ and $\sqrt{n}\,\overline{\psi}_n(\theta_0)$ under $\{Q_n(\cdot|\cdot) \ : \ n \geq 1\}$ are $\sqrt{n_0} \int (Q(y|x^*) - F(y|x^*, \theta_0))(x^* \leq x) dG(x^*)$ and $\sqrt{n_0} \int\int \psi(z, \theta_0) q(y|x) d\mu(y) dG(x)$, respectively, by Assumptions F.C1(iii), F.CA(iii), and F.CA(iv), rather than zero, (2) the Liapunov condition holds by Assumption F.CA(i), and (3) the covariances $C(z_1, z_2, \theta_0, \widehat{G}_n)$ defined with $f(\cdot|\cdot, \theta_0)$ replaced by $q_n(\cdot|\cdot)$ have the same limit as without this replacement by straightforward calculations. This completes the proof of (A.40).

Combining the results of Lemma A.4 and (A.40) and using the continuous mapping theorem (as in the proof of Theorem A.2(a)) establishes Theorem A.4(a).

Theorem A.4(b) holds given Theorem A.4(a) by the same argument as used to prove Theorem A.2(b) with (A.17) replaced by

$$\text{(A.42)} \qquad \left|\tfrac{1}{n} \textstyle\sum_{i=1}^n P_{Q_n}(Z_i \in B(z, 1/k)) - P_{\theta_0}(Z \in B(z, 1/k))\right|$$
$$\leq \left|\tfrac{1}{n} \textstyle\sum_{i=1}^n [P_{Q_n}(Z_i \in B(z, 1/k)) - P_{\theta_0}(Z_i \in B(z, 1/k))]\right|$$
$$+ \left|\tfrac{1}{n} \textstyle\sum_{i=1}^n P_{\theta_0}(Z_i \in B(z, 1/k)) - P_{\theta_0}(Z \in B(z, 1/k))\right| \to 0 \ ,$$

where P_{Q_n} denotes probability when Y_i given X_i has df $Q_n(\cdot|X_i)$. The second summand on the right-hand side of (A.42) is $o(1)$ by Assumption F.C1(v). The first summand is $o(1)$ because it is less than or equal to $\frac{1}{n} \sum_{i=1}^n \int |d(y, X_i)/\sqrt{n}| d\mu(y) \leq 2\sqrt{\frac{n_0}{n}} = o(1)$. □

FOOTNOTES

[1]I thank the co-editor, two referees, and David Pollard for helpful comments. I gratefully acknowledge research support from the National Science Foundation through grant number SBR–9410675.

I dedicate this paper to Anne, Brett, Alex, my parents, the rest of my family, and many friends and colleagues for their support over the past year and a half. A special thanks to Bill and Lorraine for their willingness to donate and to Jeff for doing so. (It looks like I won't be making any returns, but I am hanging onto your lifetime guarantee. A question, though, for whose lifetime is it guaranteed?) I thank Dr. Paul Genecin, Molly Meyer, Dr. Art Levy, and the oncology staffs at the Yale Health Plan and Yale–New Haven Hospital for their excellent care and continual support. I thank Dr. Joel Rappaport, Dr. Laura DeCastro, and all the staff at the bone marrow transplant unit and clinic at Yale–New Haven Hospital for their superb care and extraordinary dedication. I am indebted to you all.

[2]This follows because $\lim_{n\to\infty} P(T_n \in B|X) = P(T \in B)$ with $\{X_i : i \geq 1\}$ probability one implies $\lim_{n\to\infty} P(T_n \in B) = P(T \in B)$ by the bounded convergence theorem, where $P(\cdot|X)$ denotes conditional probability given $\{X_i : i \geq 1\}, T$ and $\{T_n : n \geq 1\}$ are random vectors, and B is some measurable set. Thus, $T_n \xrightarrow{d} T$ (or $T_n \xrightarrow{p} 0$) as $n \to \infty$ conditional on $\{X_i : i \geq 1\}$ with $\{X_i : i \geq 1\}$ probability one implies $T_n \xrightarrow{d} T$ ($T_n \xrightarrow{p} 0$) as $n \to \infty$ unconditionally.

[3]More generally, if $f(y|x,\theta)$ is not pointwise differentiable in θ, but the square-root density $p(y|x,\theta) = f^{1/2}(y|x,\theta)$ is $L^2(\mu)$ differentiable in θ with $L^2(\mu)$ derivative $\dot{p}(y|x,\theta)$, as in "regular" parametric models, then $\psi(z,\theta)$ is the score function $2\dot{p}(y|x,\theta)(p(y|x,\theta) > 0)/p(y|x,\theta)$ and D_0 is the inverse of the information matrix $4\int\int \dot{p}(y|x,\theta)\dot{p}(y|x,\theta)d\mu(y)dG(x)$.

[4]To see this, note that the Skorokhod representation theorem guarantees the existence of a probability space and rv's $\{\widetilde{\theta}_n : n \geq 1\}$ on it such that $\mathcal{L}(\widetilde{\theta}_n|X) = \mathcal{L}(\widehat{\theta}_n|X)\ \forall n \geq 1$, where $\widehat{\theta} = \widehat{\theta}_n$, and $\widetilde{\theta}_n \to \theta_0$ a.s. cond'l on X wp1. By (4.4), we have $c_{\alpha n}(\widetilde{\theta}_n) \to c_\alpha(\theta_0)$ a.s. cond'l on X wp1, which implies $c_{\alpha n}(\widetilde{\theta}_n) \xrightarrow{p} c_\alpha(\theta_0)$ cond'l on X wp1. Since $c_{\alpha n}(\widetilde{\theta}_n)$ and $c_{\alpha n}(\widehat{\theta})$ have the same distribution, this yields $c_{\alpha n}(\widehat{\theta}) \xrightarrow{p} c_\alpha(\theta_0)$ cond'l on X wp1.

REFERENCES

Andrews, D. W. K. (1988): "Chi-square Diagnostic Tests for Econometric Models: Theory," *Econometrica*, 56, 1419–1453.

Beran, R. and P. W. Miller (1989): "A Stochastic Minimum Distance Test for Multivariate Parametric Models," *Annals of Statistics*, 17, 125–140.

Bierens, H. J. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.

Billingsley, P. (1979): *Probability and Measure.* New York: Wiley.

Donsker, M. D. (1952): "Justification and Extension of Doob's Heuristic Approach to the Kolmogorov–Smirnov Theorems," *Annals of Mathematical Statistics*, 23, 277–281.

Doob, J. L. (1949): "Heuristic Approach to the Kolmogorov–Smirnov Theorems," *Annals of Mathematical Statistics*, 20, 393–403.

Durbin, J. (1973a): *Distribution Theory for Tests Based on the Sample Distribution Function.* Philadelphia: SIAM.

______ (1973b): "Weak Convergence of the Sample Distribution Function When Parameters Are Estimated," *Annals of Statistics*, 1, 279–290.

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

Ibragimov, I. A. and R. Z. Has'minskii (1981): *Statistical Estimation: Asymptotic Theory.* New York: Springer.

Kolmogorov, A. N. (1933): "Sulla Determinazione Empirica Di Una Legge Di Distribuzione," *Giornale Dell'Instituto Ital. Degli Attuari*, 4, 83–91.

Lifshits, M. A. (1982): "On the Absolute Continuity of Distributions of Functionals of Random Processes," *Theory of Probability and Its Applications*, 27, 600–607.

Maddala, G. S. (1983): *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press.

McCullagh, P. and J. A. Nelder (1983): *Generalized Linear Models.* London: Chapman and Hall.

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

______ (1990): *Empirical Processes: Theory and Applications.* CBMS Conference Series in Probability and Statistics, Vol. 2. Hayward, CA: Institute of Mathematical Statistics.

Smirnov, N. (1939): "On the Estimation of the Discrepancy between Empirical Curves of Distribution for Two Independent Samples," *Bulletin Mathèmatique de l'Université de Moscou*, 2, fasc. 2.

Stinchcombe, M.B. and H. White (1993): "Consistent Specification Testing with Unidentified Nuisance Parameters Using Duality and Banach Space Limit Theory," Discussion Paper 93–14, Department of Economics, University of California, San Diego.

Strasser, H. (1985): *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory.* New York: de Gruyter.

Zheng, J. X. (1993): "A Consistent Test of Conditional Parametric Distributions," unpublished manuscript, Department of Economics, University of Texas, Austin.

——— (1994): "A Specification Test of Conditional Parametric Distributions Using Kernel Estimations Methods," unpublished manuscript, Department of Economics, University of Texas, Austin.

TABLE 1

True Size of the Conditional Kolmogorov Test

	Sample Size	Nominal Significance Level 1%	5%	10%
(a) $\sigma_X^2 = 1/2$	25	.016 (.002)	.076 (.003)	.143 (.005)
	50	.013 (.002)	.059 (.003)	.119 (.005)
	100	.016 (.002)	.073 (.003)	.137 (.005)
	250	.006 (.002)	.048 (.005)	.101 (.007)
(b) $\sigma_X^2 = 1$	25	.013 (.002)	.073 (.003)	.149 (.005)
	50	.016 (.002)	.067 (.003)	.136 (.005)
	100	.011 (.002)	.052 (.003)	.111 (.005)
	250	.012 (.002)	.056 (.005)	.102 (.007)
(c) $\sigma_X^2 = 3$	25	.014 (.002)	.052 (.003)	.138 (.005)
	50	.009 (.002)	.063 (.003)	.146 (.005)
	100	.018 (.002)	.064 (.003)	.122 (.005)
	250	.014 (.002)	.051 (.005)	.114 (.007)