University of Connecticut

Department of Economics Working Paper Series

Testing Conditional Symmetry Without Smoothing

Tao Chen
University of Connecticut

Gautam Tripathi
University of Connecticut

Abstract

We test the assumption of conditional symmetry used to identify and estimate parameters in regression models with endogenous regressors without making any distributional assumptions. The specification test proposed here is computationally tractable, does not require nonparametric smoothing, and can detect n1/2-deviations from the null. Since the limiting distribution of the test statistic turns out to be a non-pivotal gaussian process, the critical values for implementing the test are obtained by simulation. In a Monte Carlo study we use the approach proposed here to test the assumption of conditional symmetry maintained in the seminal paper of Powell (1986b). Results from this finite sample experiment suggest that our test can work very well in moderately sized samples.

Journal of Economic Literature Classification: C12, C14

1. INTRODUCTION

Let $(Y, X)_{1+\dim(X) \times 1}$ be an observed random vector and μ a real valued function of X known up to a finite dimensional parameter such that for some $\theta_0 \in \Theta \subset \mathbb{R}^{\dim(\Theta)}$,

$$Y = \mu(X, \theta_0) + \varepsilon. \tag{1.1}$$

The unobserved "error" ε represents the aggregate effect of all factors affecting Y that are not part of the model and may therefore be correlated with some or all coordinates of X, i.e., the explanatory variables may be potentially endogenous. We assume that ε is continuously distributed with full support on $\mathbb{R}$.

To identify θ_0, the parameter of interest in (1.1), we assume the existence of a random vector $W_{\dim(W) \times 1}$ (the instrumental variables for X) such that the unknown conditional distribution of ε given W is symmetric about the origin, i.e., the null hypothesis is that

$$H_0 : \quad \text{Law}(\varepsilon|W) = \text{Law}(-\varepsilon|W). \tag{1.2}$$

W and X can have elements in common because the exogenous coordinates of X act as their own instruments. If all regressors are exogenous then, of course, $W = X$. For notational ease, the vector containing the distinct coordinates of X and W is written simply as (X, W).

The main objective of this paper is to propose a smoothing-free test for (1.2) against the alternative that it is false, i.e., the alternative hypothesis is that

$$H_1 : \quad \forall \theta \in \Theta, \ \text{Law}(\varepsilon(\theta)|W) \neq \text{Law}(-\varepsilon(\theta)|W), \tag{1.3}$$

where $\varepsilon(\theta) := Y - \mu(X, \theta)$ (of course, $\varepsilon(\theta_0) = \varepsilon$).

Symmetry is a powerful shape restriction because it unambiguously fixes the location under minimal assumptions, e.g., ε need not possess any moments. Moreover, it yields additional information about the underlying distribution in the form of an infinite number of moment conditions (namely, that the mean of every odd functions vanishes) that can be exploited to increase the efficiency of estimators. These fundamental properties are essentially the reason why symmetry is often imposed on statistical models to identify parameters and estimate them more efficiently; cf., e.g., Bickel (1982), Powell (1986b), Newey and Powell (1987), and Newey (1988). Symmetry considerations can also play important roles in economic modeling. For instance, since constraints in the nominal wage adjustment process can affect the smooth functioning of labor markets, much attention has been paid in empirical labor economics to determine whether the changes in nominal wages are symmetrically distributed in order to test the hypothesis that nominal wages are downwards rigid, cf., e.g., Card and Hyslop (1996), Christofides and Stengos (2001), and Stengos and Wu (2004). In empirical finance, the pricing of risky assets depends upon the skewness of the distribution of their returns, i.e., assets that make the portfolio returns more left (right) skewed command higher (lower) expected returns,

cf. Harvey and Siddique (2000) and the references therein. Additional examples of symmetry restrictions in empirical macroeconomics and finance can be found in Bai and Ng (2001).

Since symmetry restrictions are widely used, it is not surprising that nonparametric tests to determine whether the assumption of symmetry is supported by the data have been extensively studied. For instance, there is a large statistical literature on testing symmetry in univariate models; cf. Randles and Wolfe (1979, Section 3.5) and the references therein. For multivariate models, Fan and Gencay (1995), Ahmad and Li (1997), and Zheng (1998) use kernel smoothing, Bai and Ng (2001) employ the martingale transformation proposed by Khmaladze (1993), and Neumeyer, Dette, and Nagel (2005), Delgado and Escanciano (2007), and Neumeyer and Dette (2007) rely on empirical process theory. While the smoothing approach leads to tests that are consistent against fixed alternatives, these tests possess zero power against $n^{1/2}$-deviations from the null and moreover they depend upon smoothing parameters which are not always easy to choose. Khmaladze's transformation leads to distribution free tests which have power against $n^{1/2}$-alternatives, although it too requires nonparametric smoothing. The empirical process based tests require no smoothing but are not distribution free so that critical values for implementing the test cannot be tabulated in advance; instead, they are obtained by simulation.

Although our test is also based on comparing two empirical processes, it differs from the aforementioned papers in some important ways. First, unlike these papers, we allow the regressors in our model to be endogenous. This is a non-trivial extension because endogeneity of regressors does affect the limiting distribution of the test statistics (cf. Example 3.1). Second, the Kolmogorov-Smirnov (KS) type statistic we propose is computationally more tractable than the one proposed in these papers because it only requires a search over a finite number of points (cf. Section 2). By contrast, the KS statistics in Delgado and Escanciano and Neumeyer et al are implemented by searching over the uncountable set $\mathbb{R}^d$, where d depends upon the dimension of the variables involved. Third, unlike the earlier papers who only use the empirical distribution function to construct the test statistic, our technical treatment is very general and encompasses a large class of test functions allowing us to determine the asymptotic properties of a large class of test statistics in a unified manner. Finally, in the simulation experiment in Section 6 we study how the approach proposed here can be used to test the assumption of conditional symmetry maintained in Powell (1986b). To the best of our knowledge, this issue appears not to have been examined earlier in the literature (except perhaps for a brief mention in Powell (1986a, p. 155) suggesting a quantile matching approach to test for symmetry).

The remainder of the paper is organized as follows. Section 2 motivates the test statistic, and its large sample behavior under the null hypothesis is described in Sections 3. As the limiting distribution of the test statistic turns out to be non-pivotal, Section 4 shows how to simulate the critical values. In Section 5 we demonstrate that our test is consistent against

fixed alternatives and possesses non-trivial power against sequences of alternatives that are $n^{1/2}$-distant from the null. Small sample properties of the test are examined in Section 6. Finally, Section 7 concludes the paper. All proofs are in the appendices.

2. THE TEST STATISTIC

Our test for conditional symmetry is easy to motivate and is based on the almost obvious fact (whose proof, for the sake of completeness, is provided in the appendix) that

$$\text{Law}(\varepsilon|W) = \text{Law}(-\varepsilon|W) \iff (\varepsilon, W) \stackrel{\text{d}}{=} (-\varepsilon, W), \tag{2.1}$$

where "$\stackrel{\text{d}}{=}$" is shorthand for "equal in distribution". This suggests that a test for (1.2) can be obtained by comparing the empirical distributions of $Z := (\varepsilon, W)_{1+\dim(W)\times 1}$ and $Z^{\text{r}} := (-\varepsilon, W)$, where Z^{r} denotes Z with its first coordinate reflected about the origin. Since Z and Z^{r} are unobserved as they depend upon ε, we compare the empirical distributions of their feasible versions $\hat{Z} := (\hat{\varepsilon}, W)$ and $\hat{Z}^{\text{r}} := (-\hat{\varepsilon}, W)$ instead, where $\hat{\varepsilon} := Y - \mu(X, \hat{\theta})$ and $\hat{\theta}$ is an estimator of θ_0, e.g., the IV estimator, obtained using data $(Y_1, X_1, W_1), \ldots, (Y_n, X_n, W_n)$.

Given $z \in \mathbb{R}^{1+\dim(W)}$, let $(-\infty, z] := (-\infty, z^{(1)}] \times \ldots (-\infty, z^{(1+\dim(W))}]$ denote the closed lower orthant in $\mathbb{R}^{1+\dim(W)}$. Then one possible statistic for testing (1.2) is

$$\hat{\text{R}} := \sup_{z \in \mathbb{R}^{1+\dim(W)}} |n^{-1} \sum_{j=1}^{n} \mathbb{1}_{(-\infty, z]}(\hat{Z}_j) - n^{-1} \sum_{j=1}^{n} \mathbb{1}_{(-\infty, z]}(\hat{Z}_j^{\text{r}})|,$$

which resembles the usual KS statistic for testing whether two samples come from the same distribution (here $\mathbb{1}$ is the indicator function). The null hypothesis is rejected if the observed value of $\hat{\text{R}}$ is large enough. The computational cost of implementing $\hat{\text{R}}$ can be reduced by restricting search to the finite subset $\hat{\mathcal{Z}} := \{\hat{Z}_1, \ldots, \hat{Z}_n\}$ that in a certain sense becomes dense in $\text{supp}(Z)$, the support of Z, as the sample size increases. Therefore, motivated by Andrews (1997), we focus our attention on

$$\hat{\text{R}}_{\max} := \max_{z \in \hat{\mathcal{Z}}} |n^{-1} \sum_{j=1}^{n} \mathbb{1}_{(-\infty, z]}(\hat{Z}_j) - n^{-1} \sum_{j=1}^{n} \mathbb{1}_{(-\infty, z]}(\hat{Z}_j^{\text{r}})|.$$

It will be shown later that $\hat{\text{R}}$ and $\hat{\text{R}}_{\max}$ behave similarly in large samples. [Note that since $\hat{\mathcal{Z}}$ consists of estimated observations (unlike the statistic in Andrews's paper which is maximized over the "true" observations), the effect of estimating θ_0 has to be taken into account when showing that $\hat{\mathcal{Z}}$ is dense in $\text{supp}(Z)$ as $n \to \infty$; cf. Lemma F.3 for exact details.]

As far as checking the equality of two probability measures is concerned, it is worth keeping in mind that the lower orthants used in the construction of $\hat{\text{R}}$ and $\hat{\text{R}}_{\max}$ can be substituted by any measure determining class of sets. For instance, letting $\mathbb{S}^{\dim(W)}$ denote the unit

sphere in $\mathbb{R}^{1+\dim(W)}$, (1.2) could also be tested with

$$\hat{\mathrm{H}} := \sup_{(z,t)\in\mathbb{S}^{\dim(W)}\times\mathbb{R}} |n^{-1}\sum_{j=1}^{n}\mathbb{1}_{H(z,t)}(\hat{Z}_j) - n^{-1}\sum_{j=1}^{n}\mathbb{1}_{H(z,t)}(\hat{Z}_j^{\mathrm{r}})|,$$

which employs closed half-spaces $H(z,t) := \{s \in \mathbb{R}^{1+\dim(W)} : s'z \le t\}$ instead of orthants. Analogous statistics when closed/open lower orthants are replaced by closed/open upper orthants or when closed half-spaces are replaced by open half-spaces etc. follow mutatis mutandis.

It is useful to consider the statistics described above as special cases of a general statistic so that asymptotic properties of a large class of test statistics can be investigated in a unified manner. To accomplish this goal we introduce some compact notation. Henceforth, let $\mathbb{P}_U$ be the distribution of a random vector U and $\hat{\mathbb{P}}_U$ denote the empirical measure induced by n observations on U, i.e., $\hat{\mathbb{P}}_U := n^{-1}\sum_{j=1}^{n}\delta_{U_j}$ where δ_{U_j} is the Dirac (i.e., point) measure at U_j; e.g., $\hat{\mathbb{P}}_{\hat{Z}} := n^{-1}\sum_{j=1}^{n}\delta_{\hat{Z}_j}$ and $\hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}} := n^{-1}\sum_{j=1}^{n}\delta_{\hat{Z}_j^{\mathrm{r}}}$ are the empirical measures induced by $\hat{Z}_1,\ldots,\hat{Z}_n$ and $\hat{Z}_1^{\mathrm{r}},\ldots,\hat{Z}_n^{\mathrm{r}}$, respectively. Let $\mathcal{F}$ be a collection of "test" functions from $\mathbb{R}^{1+\dim(W)} \to \mathbb{R}$ and $\ell^{\infty}(\mathcal{F})$ denote the set of bounded functions from $\mathcal{F} \to \mathbb{R}$. We often use linear functional notation to write the integral of $f \in \mathcal{F}$ with respect to $\mathbb{P}_Z$ as $\mathbb{P}_Z f := \int f(z)\,\mathbb{P}_Z(dz) = \int f(u,w)\,\mathbb{P}_{\varepsilon,W}(du,dw)$. [If the region of integration is omitted it means that integration is over the support of Z.]

A general KS statistic for testing (1.2) can now be defined as

$$\widehat{\mathrm{KS}}_{\mathcal{F}} := \sup_{f\in\mathcal{F}}|\hat{\mathbb{P}}_{\hat{Z}}f - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}}f| = \sup_{f\in\mathcal{F}}|(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f|. \tag{2.2}$$

The statistics described earlier are encompassed by $\widehat{\mathrm{KS}}_{\mathcal{F}}$ for appropriately chosen $\mathcal{F}$. For instance, if $\mathcal{F}_1 := \{\mathbb{1}_{(-\infty,z]} : z \in \mathbb{R}^{1+\dim(W)}\}$ is the set of indicator functions of lower orthants in $\mathcal{Z}$, then $\widehat{\mathrm{KS}}_{\mathcal{F}_1} = \hat{\mathrm{R}}$. Similarly, if $\mathcal{F}_2 := \{\mathbb{1}_{H(z,t)} : (z,t) \in \mathbb{S}^{\dim(W)} \times \mathbb{R}\}$ is the collection of indicators of closed half-spaces in $\mathbb{R}^{1+\dim(W)}$, then $\widehat{\mathrm{KS}}_{\mathcal{F}_2} = \hat{\mathrm{H}}$. Although $\hat{\mathrm{R}}_{\max}$ can also be written as $\widehat{\mathrm{KS}}_{\hat{\mathcal{F}}_1}$ with $\hat{\mathcal{F}}_1 := \{\mathbb{1}_{(-\infty,z]} : z \in \hat{\mathcal{Z}}\}$, for technical reasons we prefer that $\mathcal{F}$ not be random. Hence, we will deal with $\hat{\mathrm{R}}_{\max}$ separately when showing that it has the same asymptotic properties as $\hat{\mathrm{R}}$.

The expression for $\widehat{\mathrm{KS}}_{\mathcal{F}}$ in (2.2) can be further simplified. In particular, instead of expressing $\widehat{\mathrm{KS}}_{\mathcal{F}}$ as a contrast between $\hat{\mathbb{P}}_{\hat{Z}}$ and $\hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}}$ as we do in (2.2), there is an equivalent representation in terms of a single measure. To see this, let $f^{\mathrm{r}}(Z) := f(Z^{\mathrm{r}})$. Then,

$$\widehat{\mathrm{KS}}_{\mathcal{F}} = \sup_{f\in\mathcal{F}}|\hat{\mathbb{P}}_{\hat{Z}}(f - f^{\mathrm{r}})|. \tag{2.3}$$

Besides reducing computational cost (only one empirical measure has to be calculated instead of two), this equivalence also simplifies some technical arguments (e.g., showing the asymptotic tightness of $\widehat{\mathrm{KS}}_{\mathcal{F}}$) because we only have to deal with one measure instead of two. Notice that

since $f - f^{\mathrm{r}}$ is antisymmetric in its first coordinate, $\widehat{\mathrm{KS}}_{\mathcal{F}}$ will be small (ideally zero) if the null hypothesis is true. Hence, (2.3) also has an appealing interpretation.

The following additional notation is used for the remainder of the paper: $A(\theta) := (\varepsilon(\theta), X, W)$ is the random vector containing $\varepsilon(\theta)$ and the distinct coordinates of X and W; hence, if all regressors are exogenous, i.e., $W = X$, then $A(\theta) = (\varepsilon(\theta), X)$. We write $A_0 := A(\theta_0)$ for notational convenience. Keeping in mind that $\varepsilon(\theta)$ is continuously distributed whereas X and W may have discrete components, $p_{A(\theta)}(u, x, w) \, du \, \kappa(dx, dw) := \mathbb{P}_{A(\theta)}(du, dx, dw)$ denotes the density of $\mathbb{P}_{A(\theta)}$, where the dominating measure κ is a mixture of the Lebesgue and counting measures. Given a random vector U, $L_2(\mathbb{P}_U)$ is the set of real-valued functions of U that are square-integrable with respect to $\mathbb{P}_U$. The $L_2(\mathbb{P}_U)$ inner product and norm are $\langle a_1, a_2 \rangle_{\mathbb{P}_U} := \int a_1(u) a_2(u) \, \mathbb{P}_U(du)$ and $\|a\|_{2, \mathbb{P}_U} := \langle a, a \rangle_{\mathbb{P}_U}^{1/2}$. The euclidean norm is $\| \cdot \|$ and $B(\theta, \epsilon)$ is the open ball of radius ϵ centered at θ; its $\| \cdot \|$-closure is $\overline{B(\theta, \epsilon)}$.

Throughout the paper we maintain the assumption that the observations (Y_j, X_j, W_j), $j = 1, \ldots, n$, are iid. Unless stated otherwise, all limits are taken as the sample size $n \to \infty$.

3. Large sample results under the null

In this section, we derive the asymptotic distribution of $\widehat{\mathrm{KS}}_{\mathcal{F}}$ when the null hypothesis is true. To do so, we make use of the notion of convergence in distribution of stochastic processes taking values in $\ell^\infty(\mathcal{F})$. Cf. Chapter 1.5 of van der Vaart and Wellner (1996), henceforth referred to as V&W, for details regarding convergence in distribution in $\ell^\infty(\mathcal{F})$.

If θ_0 is known and H_0 is true, then $(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f = (\hat{\mathbb{P}}_Z - \mathbb{P}_Z)(f - f^{\mathrm{r}})$ for $f \in \mathcal{F}$. Hence, under standard conditions on $\mathcal{F}$, it can be shown that the process $\{n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f : f \in \mathcal{F}\}$ converges in distribution in $\ell^\infty(\mathcal{F})$ to a mean zero gaussian process $\{\mathbb{G}_0 f : f \in \mathcal{F}\}$ with covariance function $\mathbb{E} \mathbb{G}_0 f_1 \mathbb{G}_0 f_2 = \mathbb{E}((f_1 - f_1^{\mathrm{r}})(Z)(f_2 - f_2^{\mathrm{r}})(Z))$, $f_1, f_2 \in \mathcal{F}$. Consequently, $n^{1/2} \widehat{\mathrm{KS}}_{\mathcal{F}}$ converges in distribution to the random variable $\sup_{f \in \mathcal{F}} |\mathbb{G}_0 f|$. Of course, in practice θ_0 is unknown and has to be estimated. As we now show, the estimation error from using $\hat{\theta}$ instead of θ_0 shifts the limiting distribution of $\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}}$ and, hence, $\widehat{\mathrm{KS}}_{\mathcal{F}}$.

For maximum generality, we try to impose as few restrictions as possible on our model. For instance, we do not require μ to be smooth in θ; instead, we only assume that $\mu(x, \cdot)$ is mean-square differentiable at θ_0, i.e.,

Assumption 3.1. *There exists a neighborhood of θ_0 such that for all θ in this neighborhood, $\mu(x, \theta) = \mu(x, \theta_0) + \dot{\mu}'(x, \theta_0)(\theta - \theta_0) + \rho(x, \theta, \theta_0)$ where $\dot{\mu} : \mathbb{R}^{\dim(X)} \times \mathbb{R}^{\dim(\Theta)} \to \mathbb{R}^{\dim(\Theta)}$ satisfies $\int \|\dot{\mu}(x, \theta_0)\|^2 \, \mathbb{P}_X(dx) < \infty$ and the remainder term $\rho : \mathbb{R}^{\dim(X)} \times \Theta \to \mathbb{R}$ is such that $\int \sup_{\theta \in \overline{B(\theta_0, \delta)}} \rho^2(x, \theta, \theta_0) \, \mathbb{P}_X(dx) = o(\delta^2)$ as $\delta \to 0$. [Note that ρ is identically zero if μ is linear in parameters.]*

Since mean-square differentiability is weaker than pointwise differentiability, our setup allows for kinks in the functional forms for μ.

Let ∂_1 denote partial differentiation with respect to the first argument. The next assumption stipulates that

Assumption 3.2. (i) $u \mapsto p_{\varepsilon|X=x,W=w}(u)$ *is differentiable a.e. on* $\mathbb{R}$ *and the conditional second moment of* $\partial_1 p_{\varepsilon|X=x,W=w}$, *denoted by* $v_{A_0}(x,w) := \int (\partial_1 \log p_{\varepsilon|X=x,W=w})^2(u)\, p_{\varepsilon|X=x,W=w}(u)\, du$, *is uniformly bounded, i.e.,* $\|v_{A_0}\|_\infty := \sup_{(x,w)\in \mathrm{supp}(X)\times\mathrm{supp}(W)} |v_{A_0}(x,w)| < \infty$. (ii) *The conditional distribution of* $\varepsilon|X,W$ *is Lipschitz, i.e., there exists a nonnegative* $\zeta \in L_2(\mathbb{P}_{X,W})$ *such that* $|\int_{(-\infty,t_1]} p_{\varepsilon|X=x,W=w}(u)\, du - \int_{(-\infty,t_2]} p_{\varepsilon|X=x,W=w}(u)\, du| \le \zeta(x,w)|t_1 - t_2|$ *for all* $t_1, t_2 \in \mathbb{R}$.

(i) implies (cf. Hájek, 1972, Lemma A.3) that the square-root density $p_{\varepsilon|X=x,W=w}^{1/2}$ is mean-square differentiable, i.e., given $\delta \in L_2(\mathbb{P}_{X,W})$,

$$\frac{p_{\varepsilon|X=x,W=w}^{1/2}(u + \delta(x,w)) - p_{\varepsilon|X=x,W=w}^{1/2}(u)}{p_{\varepsilon|X=x,W=w}^{1/2}(u)}$$

$$= \frac{1}{2}\frac{\partial_1 p_{\varepsilon|X=x,W=w}(u)}{p_{\varepsilon|X=x,W=w}(u)}\delta(x,w) + r(\delta(x,w), u, x, w), \quad (3.1)$$

where, for each $\epsilon > 0$, $\int r^2(\delta(x,w), u, x, w)\, \mathbb{P}_{\varepsilon|X=x,W=w}(du) \le \epsilon \delta^2(x,w)$. As a consequence, $\|r(\delta, \cdot, \cdot, \cdot)\|_{2,\mathbb{P}_{A_0}} = o(\|\delta\|_{2,\mathbb{P}_{X,W}})$ as $\|\delta\|_{2,\mathbb{P}_{X,W}} \to 0$. This fact is used to bound remainder terms for (B.3) in the proof of Lemma 3.1. (ii) implies that $\mathcal{F}_1$ and $\mathcal{F}_2$ satisfy Assumption 3.3(v).

The requirements on the class of test functions are as follows:

Assumption 3.3. (i) $\mathcal{F}$ *separates probability measures on* $\mathbb{R}^{1+\dim(W)}$, *i.e., if* $\mathbb{P}_1 \ne \mathbb{P}_2$ *are probability measures on* $\mathbb{R}^{1+\dim(W)}$ *then* $\mathbb{P}_1 f \ne \mathbb{P}_2 f$ *for some* $f \in \mathcal{F}$. (ii) $\mathcal{F}$ *has a bounded envelope, i.e.,* $M_{\mathcal{F}} := \sup_{(f,z)\in\mathcal{F}\times\mathbb{R}^{1+\dim(W)}} |f(z)| < \infty$. (iii) $\mathcal{F}$ *is* $\mathbb{P}_{A_0}$-*Donsker.* (iv) *Elements of* $\mathcal{F}$ *are stable with respect to perturbations in their first argument in the sense that* $f \in \mathcal{F} \implies f(\cdot + t, \cdot) \in \mathcal{F}$ *for all* $t \in \mathbb{R}$. (v) *There exists a continuous function* $q : [0,\infty) \to [0,\infty)$ *such that* $q(0) = 0$, $\sup_{f\in\mathcal{F}} \|f(\cdot - \Delta(X,\theta,\theta_0), \cdot) - f(\cdot, \cdot)\|_{2,\mathbb{P}_{A_0}} \le q(\|\theta - \theta_0\|)$, *and* $\sup_{f\in\mathcal{F}} \|f^{\mathrm{r}}(\cdot - \Delta(X,\theta,\theta_0), \cdot) - f^{\mathrm{r}}(\cdot, \cdot)\|_{2,\mathbb{P}_{A_0}} \le q(\|\theta - \theta_0\|)$, *where* $\Delta(X,\theta_a,\theta_b) := \mu(X,\theta_a) - \mu(X,\theta_b)$.

(i) is necessary because we employ elements of $\mathcal{F}$ to distinguish between the distributions of Z and Z^{r} (cf. the proof of Theorem 5.1). $\mathcal{F}_1$ and $\mathcal{F}_2$ satisfy (i) due to the fact that orthants and half-spaces separate probability measures. (ii), clearly satisfied by indicator functions, helps to uniformly (in $\mathcal{F}$) bound remainder terms in (B.3) in the proof of Lemma 3.1. In (iii), $\mathcal{F}$ is required to be Donsker with respect to the "bigger" measure $\mathbb{P}_{A_0}$ rather than $\mathbb{P}_Z$ in order to control the estimation uncertainty (from estimating θ_0) associated with the endogenous components of X. (iii), which implies that $\mathcal{F}^{\mathrm{r}} := \{f^{\mathrm{r}} : f \in \mathcal{F}\}$ is also $\mathbb{P}_{A_0}$-Donsker because $\mathcal{F}$ and $\mathcal{F}^{\mathrm{r}}$ are covered by the same number of $L_2(\mathbb{P}_{A_0})$-brackets, is used in the proof of Lemma 3.2

to show that $\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}}$ converges in distribution in $\ell^\infty(\mathcal{F})$. Following V&W (Example 2.5.4), $\mathcal{F}_1$ (hence $\mathcal{F}_1^{\mathrm{r}}$) is $\mathbb{P}_{A_0}$-Donsker because its $L_2(\mathbb{P}_{A_0})$-bracketing entropy is finite; the same argument shows that $\mathcal{F}_2$ (hence $\mathcal{F}_2^{\mathrm{r}}$) is $\mathbb{P}_{A_0}$-Donsker. (iv) implies that test functions with estimation error as an argument are still test functions. To see this, let $f \in \mathcal{F}_1$ so that $f = \mathbb{1}_{(-\infty,\tau]\times(-\infty,v]}$ for some $(\tau, v) \in \mathbb{R} \times \mathbb{R}^{\dim(W)}$. Then $f(Y - \mu(X, \hat{\theta}), W) = \mathbb{1}_{(-\infty,\tau]\times(-\infty,v]}(\varepsilon - \Delta(X, \hat{\theta}, \theta_0), W) = \mathbb{1}_{(-\infty,\tau+\Delta(X,\hat{\theta},\theta_0)]\times(-\infty,v]}(\varepsilon, W)$, implying that $f(Y - \mu(X, \hat{\theta}), W) \in \mathcal{F}$; the same argument works for $\mathcal{F}_2$ as well. (v) is similar to (2.7) of Khmaladze and Koul (2004) and (3) of van der Vaart and Wellner (2007). Under Assumption 3.2(ii), $\mathcal{F}_1$ and $\mathcal{F}_2$ satisfy (v) with $q(t) \propto t^{1/2}$ (cf. Lemma F.2). (iii)–(v) plus an asymptotic equicontinuity argument helps uniformly (in $\mathcal{F}$) bound (B.2) in the proof of Lemma 3.1.

We also assume that $\hat{\theta}$ is a consistent estimator of θ_0, i.e.,

Assumption 3.4. $\mathrm{plim}(\hat{\theta}) = \theta_0$.

We are now ready to describe how estimation of θ_0 affects the empirical distributions of $\hat{Z}$ and $\hat{Z}^{\mathrm{r}}$; this will help us derive the limiting distribution of $\widehat{\mathrm{KS}}_\mathcal{F}$. Henceforth, $\dot{\mu}_0(\cdot) := \dot{\mu}(\cdot, \theta_0)$ and the symbol o_{Pr° indicates asymptotic negligibility in outer probability (Pr°) to take care of measurability issues that may arise when taking the supremum over an uncountable set.

Lemma 3.1. *Let Assumptions* 3.1–3.4 *hold. Then,*

$$\sup_{f\in\mathcal{F}} |(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_Z)f - \langle f, (\partial_1 \log p_{A_0})\dot{\mu}_0'\rangle_{\mathbb{P}_{A_0}}(\hat{\theta} - \theta_0)| = o_{\mathrm{Pr}^\circ}(n^{-1/2}) + o_{\mathrm{Pr}}(\|\hat{\theta} - \theta_0\|)$$

$$\sup_{f\in\mathcal{F}} |(\hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}})f - \langle f^{\mathrm{r}}, (\partial_1 \log p_{A_0})\dot{\mu}_0'\rangle_{\mathbb{P}_{A_0}}(\hat{\theta} - \theta_0)| = o_{\mathrm{Pr}^\circ}(n^{-1/2}) + o_{\mathrm{Pr}}(\|\hat{\theta} - \theta_0\|).$$

The term $\langle \cdot, (\partial_1 \log p_{A_0})\dot{\mu}_0'\rangle_{\mathbb{P}_{A_0}}(\hat{\theta} - \theta_0)$, which vanishes if $\hat{\theta} = \theta_0$, captures the uncertainty arising from estimating θ_0. It is useful to note that Lemma 3.1 holds whether the null hypothesis (1.2) is true or not (provided θ_0 is suitably redefined, cf. Section 4). We will exploit this fact in showing the consistency of $\widehat{\mathrm{KS}}_\mathcal{F}$ (cf. Lemma F.1).

Next, we derive limiting distribution of $\widehat{\mathrm{KS}}_\mathcal{F}$. Begin by observing that since $n^{1/2}\widehat{\mathrm{KS}}_\mathcal{F} := \sup_{f\in\mathcal{F}} |n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f|$ is a functional of the stochastic process $\{n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f : f \in \mathcal{F}\}$, its distribution is determined by the distribution of $\{n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f : f \in \mathcal{F}\}$. The latter can be identified from the joint distribution of its marginals. Towards this end, fix $f \in \mathcal{F}$ and write $(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f = (\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_Z)f - (\hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}})f + (\hat{\mathbb{P}}_Z - \hat{\mathbb{P}}_{Z^{\mathrm{r}}})f$. Hence, since $\mathbb{P}_Z = \mathbb{P}_{Z^{\mathrm{r}}}$ by (1.2), under the hull hypothesis we have that

$$\begin{aligned}
(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f = {}& ((\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_Z)f - \langle f, (\partial_1 \log p_{A_0})\dot{\mu}_0'\rangle_{\mathbb{P}_{A_0}}(\hat{\theta} - \theta_0)) \\
& - ((\hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}})f - \langle f^{\mathrm{r}}, (\partial_1 \log p_{A_0})\dot{\mu}_0'\rangle_{\mathbb{P}_{A_0}}(\hat{\theta} - \theta_0)) \\
& + (\hat{\mathbb{P}}_Z - \mathbb{P}_Z)f - (\hat{\mathbb{P}}_{Z^{\mathrm{r}}} - \mathbb{P}_{Z^{\mathrm{r}}})f + \langle f - f^{\mathrm{r}}, (\partial_1 \log p_{A_0})\dot{\mu}_0'\rangle_{\mathbb{P}_{A_0}}(\hat{\theta} - \theta_0).
\end{aligned}$$

Then, by Lemma 3.1 and the fact that $(\hat{\mathbb{P}}_Z - \mathbb{P}_Z)f - (\hat{\mathbb{P}}_{Z^r} - \mathbb{P}_{Z^r})f = (\hat{\mathbb{P}}_Z - \mathbb{P}_Z)(f - f^r)$,

$$n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^r})f = \hat{\mathbb{X}}_0(f) + o_{\mathrm{Pr}^\circ}(1) + o_{\mathrm{Pr}}(\|n^{1/2}(\hat{\theta} - \theta_0)\|) \qquad \text{unif. in } f \in \mathcal{F}, \qquad (3.2)$$

where $\hat{\mathbb{X}}_0(f) := n^{1/2}(\hat{\mathbb{P}}_Z - \mathbb{P}_Z)(f - f^r) + \langle f - f^r, (\partial_1 \log p_{A_0})\dot{\mu}_0'\rangle_{\mathbb{P}_{A_0}} n^{1/2}(\hat{\theta} - \theta_0)$. Notice that since $(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^r})f = (\hat{\mathbb{P}}_{\hat{Z}} - \mathbb{P}_Z)(f - f^r)$ under the null, we can rewrite (3.2) as

$$n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}} - \mathbb{P}_Z)(f - f^r) = n^{1/2}(\hat{\mathbb{P}}_Z - \mathbb{P}_Z)(f - f^r) + \langle f - f^r, (\partial_1 \log p_{A_0})\dot{\mu}_0'\rangle_{\mathbb{P}_{A_0}} n^{1/2}(\hat{\theta} - \theta_0)$$
$$+ o_{\mathrm{Pr}^\circ}(1) + o_{\mathrm{Pr}}(\|n^{1/2}(\hat{\theta} - \theta_0)\|) \qquad \text{unif. in } f \in \mathcal{F}.$$

This leads to a nice interpretation of (3.2), namely, that (3.2) represents a linearization of the process $\{n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}} - \mathbb{P}_Z)g : g \in \mathcal{A}_{\mathcal{F}}\}$ about θ_0, where $\mathcal{A}_{\mathcal{F}} := \{f - f^r : f \in \mathcal{F}\}$ is the set of functions generated by $\mathcal{F}$ that are antisymmetric in their first coordinate.

To identify the marginal distribution of $\hat{\mathbb{X}}_0$, we need some additional information about $\hat{\theta}$. In particular, we assume that

Assumption 3.5. *$\hat{\theta}$ is asymptotically linear with influence function φ, i.e., $n^{1/2}(\hat{\theta} - \theta_0) = n^{-1/2}\sum_{j=1}^n \varphi(Y_j, X_j, W_j, \theta_0) + o_{\mathrm{Pr}}(1)$, where $\mathbb{E}\varphi(Y, X, W, \theta_0) = 0$ and $\mathbb{E}\|\varphi(Y, X, W, \theta_0)\|^2 < \infty$.*

Thus, by the CLT for iid random vectors, $(n^{1/2}(\hat{\mathbb{P}}_Z - \mathbb{P}_Z)(f - f^r), n^{1/2}(\hat{\theta} - \theta_0))$ is asymptotically multivariate normal. In particular, $n^{1/2}(\hat{\theta} - \theta_0)$ converges in distribution in $\mathbb{R}^{\dim(\Theta)}$ to $N_{\varphi_0} \overset{\mathrm{d}}{=} \mathrm{N}(0_{\dim(\Theta)\times 1}, \mathbb{E}\varphi_0\varphi_0')$, where $\varphi_0 := \varphi(Y, X, W, \theta_0)$. Hence, for all $f \in \mathcal{F}$,

$$\hat{\mathbb{X}}_0(f) \overset{\mathrm{d}}{\to} \mathbb{X}_0(f) \quad in \ \mathbb{R}, \qquad (3.3)$$

where $\mathbb{X}_0(f) := \mathbb{G}_0 f + \langle f - f^r, (\partial_1 \log p_{A_0})\dot{\mu}_0'\rangle_{\mathbb{P}_{A_0}} N_{\varphi_0}$ and $\mathbb{G}_0$ is the gaussian process defined earlier. Since the remainder terms in (3.2) are asymptotically negligible in probability uniformly in $f \in \mathcal{F}$, the process $\{n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^r})f : f \in \mathcal{F}\}$ will converge in distribution in $\ell^\infty(\mathcal{F})$ to the limiting process $\{\mathbb{X}_0(f) : f \in \mathcal{F}\}$ provided $\{\hat{\mathbb{X}}_0(f) : f \in \mathcal{F}\}$ is asymptotically tight. As the next result shows, this is indeed the case.

Lemma 3.2. *Let Assumptions* 3.1–3.5 *hold and* (1.2) *be true. Then,*

$$\{n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^r})f : f \in \mathcal{F}\} \overset{\mathrm{d}}{\to} \{\mathbb{X}_0(f) : f \in \mathcal{F}\} \quad in \ \ell^\infty(\mathcal{F}).$$

Since $x \mapsto \sup_{f \in \mathcal{F}} |x(f)|$ (being a norm) is continuous on $\ell^\infty(\mathcal{F})$ and the limiting process in Lemma 3.2 takes values in $\ell^\infty(\mathcal{F})$, an application of the continuous mapping theorem (V&W, Theorem 1.3.6) yields the limiting distribution of $n^{1/2}\widehat{\mathrm{KS}}_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^r})f|$.

Theorem 3.1. *Let Assumptions* 3.1–3.5 *hold and* (1.2) *be true. Then, the random variable $n^{1/2}\widehat{\mathrm{KS}}_{\mathcal{F}}$ converges in distribution to the random variable $\sup_{f \in \mathcal{F}} |\mathbb{X}_0(f)|$.*

This leads immediately to the asymptotic distribution of $n^{1/2}\hat{\mathrm{R}} = n^{1/2}\widehat{\mathrm{KS}}_{\mathcal{F}_1}$.

Corollary 3.1. *Let the conditions of Theorem* 3.1 *hold. Then,* $n^{1/2}\hat{R}$ *converges in distribution to the random variable*

$$R_0 := \sup_{z \in \mathbb{R}^{1+\dim(W)}} |\mathbb{G}_0 \mathbb{1}_{(-\infty,z]} + \langle \mathbb{1}_{(-\infty,z]} - \mathbb{1}^{\mathrm{r}}_{(-\infty,z]}, (\partial_1 \log p_{A_0})\dot{\mu}'_0 \rangle_{\mathbb{P}_{A_0}} N_{\varphi_0}|.$$

Example 3.1 (No endogenous regressors). The limiting distribution of the test statistics simplifies if there are no endogenous regressors. To see this, suppose that all regressors are exogenous, i.e., $W = X$, so that $A_0 = Z$. Since $2f = (f + f^{\mathrm{r}}) + (f - f^{\mathrm{r}})$ and (1.2) implies that $\partial_1 p_Z$ is antisymmetric in its first coordinate, it follows from Corollary 3.1 that $n^{1/2}\hat{R}$ converges in distribution to the random variable $\sup_{z \in \mathbb{R}^{1+\dim(X)}} |\mathbb{G}_0 \mathbb{1}_{(-\infty,z]} + 2\langle \mathbb{1}_{(-\infty,z]}, (\partial_1 \log p_Z)\dot{\mu}'_0 \rangle_{\mathbb{P}_Z} N_{\varphi_0}|.$ □

We also expect $n^{1/2}\hat{R}_{\max}$ to converge in distribution to R_0 under the null hypothesis because $\hat{\mathcal{F}}_1$ is dense in $\mathcal{F}_1$ with probability approaching one (w.p.a.1) (cf. Lemma F.3). The next result confirms this intuition (cf. Andrews, 1997, p. 1105).

Lemma 3.3. *Let the conditions of Theorem* 3.1 *hold. Then,* $n^{1/2}\hat{R}_{\max}$ *and* $n^{1/2}\hat{R}$ *both converge in distribution to the same random variable, i.e.,* R_0.

Now that we know how $\hat{R}_{\max}$ behaves in large samples, a size-α test, $\alpha \in (0,1)$, for (1.2) based on $\hat{R}_{\max}$ can be formalized as follows: Reject H_0 if $n^{1/2}\hat{R}_{\max} \geq c_\alpha$, where c_α is the $1 - \alpha$ quantile of R_0. However, c_α cannot be obtained from a table because the distribution of R_0 depends upon θ_0 (via N_{φ_0}) and $\mathbb{P}_{A_0}$; i.e., in other words, $n^{1/2}\hat{R}_{\max}$ is not an asymptotically pivotal statistic. Instead, quantiles of R_0 can be simulated as described in Section 4.

4. The test using simulated critical values

Simulated critical values for specification tests have been used earlier; cf., e.g., Su and Wei (1991), Hansen (1996), Neumeyer, Dette, and Nagel (2005), Delgado, Domínguez, and Lavergne (2006), and Delgado and Escanciano (2007). Intuitively, the basic idea behind simulating critical values for $\hat{R}_{\max}$ is to introduce additional randomness in the data and construct an artificial random variable (say $\hat{R}^*_{\max}$) that has the same distribution as $\hat{R}_{\max}$ under the null hypothesis. Repeated draws of $\hat{R}^*_{\max}$ then yield a random sample from the distribution of $\hat{R}_{\max}$ under the null, which can be used to estimate the quantiles of $\hat{R}_{\max}$ to any desired level of accuracy since the number of draws is in control of the researcher.

The test using the simulated critical values is implemented as follows:

(i) Use $\hat{\theta}$ and the residuals $\hat{\varepsilon}_j := Y_j - \mu(X_j, \hat{\theta})$, $j = 1, \ldots, n$, to calculate $n^{1/2}\hat{R}_{\max}$.

(ii) Independent of the observed data $\mathcal{D}_n := \{(Y_j, X_j, W_j) : 1 \leq j \leq n\}$, use a random number generator to generate $R_1, \ldots, R_n \overset{\text{iid}}{\sim}$ Rademacher and define $Y^*_j := \mu(X_j, \hat{\theta}) + R_j\hat{\varepsilon}_j$, $j = 1, \ldots, n$. [Recall that a random variable R is said to have the Rademacher or symmetric Bernoulli distribution if $\mathbb{P}_R := (\delta_{-1} + \delta_1)/2$, i.e., R takes values -1 and 1 with equal

probability.] Since $\text{Law}(R\hat{\varepsilon}|W_1,\ldots,W_n) = \text{Law}(-R\hat{\varepsilon}|W_1,\ldots,W_n)$ by construction, the simulated sample $\{(R_j\hat{\varepsilon}_j, W_j) : 1 \le j \le n\}$ satisfies the null hypothesis.

(iii) Re-estimate θ_0 using (Y_j^*, X_j, W_j), $j = 1,\ldots,n$ to make the estimation error in the simulated sample resemble the estimation error in the data. Denote the resulting estimator by $\hat{\theta}^*$ and let $\hat{\varepsilon}_j^* := Y_j^* - \mu(X_j, \hat{\theta}^*)$, $j = 1,\ldots,n$, be the corresponding residuals.

(iv) Calculate $n^{1/2}\hat{R}_{\max}^* := n^{1/2}\max_{z\in\hat{\mathcal{Z}}^*}|n^{-1}\sum_{j=1}^n \mathbb{1}_{(-\infty,z]}(\hat{Z}_j^*) - n^{-1}\sum_{j=1}^n \mathbb{1}_{(-\infty,z]}(\hat{Z}_j^{*\mathrm{r}})|$, where $\hat{\mathcal{Z}}^* := \{\hat{Z}_1^*,\ldots,\hat{Z}_n^*\}$, $\hat{Z}^* := (\hat{\varepsilon}^*, W)$, and $\hat{Z}^{*\mathrm{r}} := (-\hat{\varepsilon}^*, W)$. As shown subsequently, $n^{1/2}\hat{R}_{\max}^*$ has the same limiting distribution as $n^{1/2}\hat{R}_{\max}$ when the null hypothesis is true.

(v) Repeat (ii)–(iv) B times, to get B random draws from the distribution of $n^{1/2}\hat{R}_{\max}$ under the null hypothesis. Calculate the $1 - \alpha$ sample quantile $(c_{\alpha,B}^*)$ for these draws.

(vi) The decision rule "Reject H_0 if the value of $n^{1/2}\hat{R}_{\max}$ observed in (i) exceeds $c_{\alpha,B}^*$" then leads to a size-α test for (1.2). Alternatively, the p-value can be obtained by calculating the fraction of draws in (v) that exceed the observed value of $n^{1/2}\hat{R}_{\max}$.

We now show that (vi) is justified asymptotically. To summarize our approach, we first prove that $n^{1/2}\widehat{\text{KS}}_{\mathcal{F}}^* := n^{1/2}\sup_{f\in\mathcal{F}}|(\hat{\mathbb{P}}_{\hat{Z}^*} - \hat{\mathbb{P}}_{\hat{Z}^{*\mathrm{r}}})f|$ has a well defined limiting distribution irrespective of whether the null hypothesis is true or not. From this it will follow that $n^{1/2}\hat{R}^* := n^{1/2}\sup_{z\in\mathbb{R}^{1+\dim(W)}}|n^{-1}\sum_{j=1}^n \mathbb{1}_{(-\infty,z]}(\hat{Z}_j^*) - n^{-1}\sum_{j=1}^n \mathbb{1}_{(-\infty,z]}(\hat{Z}_j^{*\mathrm{r}})|$ has the same limiting distribution as $n^{1/2}\hat{R}$ under the null and is bounded in probability otherwise. The proof ends by showing that $n^{1/2}\hat{R}_{\max}^*$ has the same limiting distribution as $n^{1/2}\hat{R}^*$ whether the null is true or not, implying in particular that $n^{1/2}\hat{R}_{\max}^*$ (hence its quantiles) is bounded in probability even when the null is false. Therefore, the test using critical values from the simulated distribution of $n^{1/2}\hat{R}_{\max}^*$ is consistent, i.e., it rejects a false null w.p.a.1, because $n^{1/2}\hat{R}_{\max} \to \infty$ w.p.a.1 when the null hypothesis is false (cf. Section 5).

We begin by assuming that $\hat{\theta}$ and $\hat{\theta}^*$ have a well defined limit even when the null hypothesis is false, i.e., irrespective of whether (1.2) is true or not,

Assumption 4.1. $\text{plim}((\hat{\theta}^*, \hat{\theta}) - (\theta_1, \theta_1)) = 0$ *for some* $\theta_1 \in \Theta$.

θ_1 is called the "pseudo-true value" and it exists under very general conditions (of course, if the null is true then $\theta_1 = \theta_0$). Let $Z(\theta) := (\varepsilon(\theta), W)$, $Z^{\mathrm{r}}(\theta) := (-\varepsilon(\theta), W)$, $A_1 := A(\theta_1)$, and let $\mathscr{R}$ denote the linear operator that "Rademacherizes" the first component of its argument, e.g., $\mathscr{R}(\hat{Z}) := (R\hat{\varepsilon}, W)$ and $\mathscr{R}(A_1) := (R\varepsilon(\theta_1), X, W)$. Given $f \in \mathcal{F}$, we can write (cf. the remark at the end of this section)

$$(\hat{\mathbb{P}}_{\hat{Z}^*} - \hat{\mathbb{P}}_{\hat{Z}^{*\mathrm{r}}})f = (\hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))} - \hat{\mathbb{P}}_{\mathscr{R}(Z^{\mathrm{r}}(\theta_1))})f + (\hat{\mathbb{P}}_{\hat{Z}^*} - \hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))})f - (\hat{\mathbb{P}}_{\hat{Z}^{*\mathrm{r}}} - \hat{\mathbb{P}}_{\mathscr{R}(Z^{\mathrm{r}}(\theta_1))})f$$
$$= (\hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))} - \mathbb{P}_{\mathscr{R}(Z(\theta_1))})(f - f^{\mathrm{r}})$$
$$+ (\hat{\mathbb{P}}_{\hat{Z}^*} - \hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))})f - (\hat{\mathbb{P}}_{\hat{Z}^{*\mathrm{r}}} - \hat{\mathbb{P}}_{\mathscr{R}(Z^{\mathrm{r}}(\theta_1))})f. \tag{4.1}$$

To get some intuition behind why it makes sense to center the last two terms about $\hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))}$ and $\hat{\mathbb{P}}_{\mathscr{R}(Z^{\mathrm{r}}(\theta_1))}$, respectively, notice that $\hat{\varepsilon}^* = R\hat{\varepsilon} - \Delta(X, \hat{\theta}^*, \hat{\theta}) = R\varepsilon(\theta_1) - \Delta(X, \hat{\theta}^*, \hat{\theta}) - R\Delta(X, \hat{\theta}, \theta_1)$. Hence, $\hat{\mathbb{P}}_{\hat{Z}^*} - \hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))}$ captures the estimation error from re-estimating θ_0 using the simulated sample; a similar interpretation holds for $\hat{\mathbb{P}}_{\hat{Z}^{*\mathrm{r}}} - \hat{\mathbb{P}}_{\mathscr{R}(Z^{\mathrm{r}}(\theta_1))}$.

To study the properties of $\hat{\mathbb{P}}_{\hat{Z}^*}$ and $\hat{\mathbb{P}}_{\hat{Z}^{*\mathrm{r}}}$ without assuming the null to be true, we strengthen Assumptions 3.1–3.3 as follows.

Assumption 4.2. (i) *There exists a neighborhood of θ_1 such that for each θ in this neighborhood, $\mu(x, \hat{\theta}^*) = \mu(x, \theta) + \dot{\mu}'(x, \theta)(\hat{\theta}^* - \theta) + \rho(x, \hat{\theta}^*, \theta)$ with $\| \sup_{\tilde{\theta} \in \overline{B(\theta, \delta)}} \rho(\cdot, \tilde{\theta}, \theta) \|_{2, \mathbb{P}_X} = o(\delta)$ as $\delta \to 0$, $\|\dot{\mu}(\cdot, \theta_1)\| \in L_2(\mathbb{P}_X)$, and $\|\dot{\mu}(\cdot, \hat{\theta}) - \dot{\mu}(\cdot, \theta_1)\|_{2, \mathbb{P}_X} = o_{\mathrm{Pr}}(1)$. (ii) $p_{\varepsilon(\theta_1)|X, W}$ satisfies Assumption 3.2. (iii) Assumption 3.3 holds with $\mathbb{P}_{A_0}$ replaced by $\mathbb{P}_{\mathscr{R}(A_1)}$ and $\Delta(X, \theta, \theta_0)$ replaced by $\tilde{\Delta}(-1, X, \hat{\theta}^*, \hat{\theta}, \theta_1)$ and $\tilde{\Delta}(1, X, \hat{\theta}^*, \hat{\theta}, \theta_1)$, where $\tilde{\Delta}(R, X, \theta_a, \theta_b, \theta_1) := \Delta(X, \theta_a, \theta_b) + R\Delta(X, \theta_b, \theta_1)$.*

Since $\operatorname{plim}(\hat{\theta}) = \theta_1$ by Assumption 4.1, (i) implies that, w.p.a.1, $\mu(x, \hat{\theta}^*) = \mu(x, \hat{\theta}) + \dot{\mu}'(x, \hat{\theta})(\hat{\theta}^* - \hat{\theta}) + \rho(x, \hat{\theta}^*, \hat{\theta})$ and $\| \sup_{\tilde{\theta} \in \overline{B(\hat{\theta}, \|\hat{\theta}^* - \hat{\theta}\|)}} \rho(\cdot, \tilde{\theta}, \hat{\theta}) \|_{2, \mathbb{P}_X} = o(\|\hat{\theta}^* - \hat{\theta}\|)$. Under (iii), $\|f^{\hat{\theta}^*, \hat{\theta}} - f^{\theta_1, \theta_1}\|_{2, \mathbb{P}_{\mathscr{R}(A_1)}} \leq q(\|(\hat{\theta}^*, \hat{\theta}) - (\theta_1, \theta_1)\|)$. These facts help prove the first result in this section, which shows how the re-estimation step affects the empirical distributions of $\hat{Z}^*$ and $\hat{Z}^{*\mathrm{r}}$ (cf. Lemma 3.1 for a comparison). Henceforth, let $\dot{\mu}_1(\cdot) := \dot{\mu}(\cdot, \theta_1)$ and, for $f \in \mathcal{F}$,

$$f^{\theta_a, \theta_b}(R\varepsilon(\theta_1), X, W) := f(R\varepsilon(\theta_1) - \tilde{\Delta}(R, X, \theta_a, \theta_b, \theta_1), W)$$

$$f^{\mathrm{r}\theta_a, \theta_b}(R\varepsilon(\theta_1), X, W) := f^{\mathrm{r}}(R\varepsilon(\theta_1) - \tilde{\Delta}(R, X, \theta_a, \theta_b, \theta_1), W).$$

Lemma 4.1. *Let Assumptions 4.1 and 4.2 hold. Then, whether or not the null is true,*

$$\sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{\hat{Z}^*} - \hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))})f - \mathbb{P}_{\mathscr{R}(A_1)}(f^{\hat{\theta}, \hat{\theta}} - f^{\theta_1, \theta_1})$$

$$- 0.5\langle f - f^{\mathrm{r}}, (\partial_1 \log p_{A_1})\dot{\mu}_1' \rangle_{\mathbb{P}_{A_1}}(\hat{\theta}^* - \hat{\theta})| = o_{\mathrm{Pr}^\circ}(n^{-1/2}) + o_{\mathrm{Pr}}(\|\hat{\theta}^* - \hat{\theta}\|)$$

and

$$\sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{\hat{Z}^{*\mathrm{r}}} - \hat{\mathbb{P}}_{\mathscr{R}(Z^{\mathrm{r}}(\theta_1))})f - \mathbb{P}_{\mathscr{R}(A_1)}(f^{\mathrm{r}\hat{\theta}, \hat{\theta}} - f^{\mathrm{r}\theta_1, \theta_1})$$

$$- 0.5\langle f^{\mathrm{r}} - f, (\partial_1 \log p_{A_1})\dot{\mu}_1' \rangle_{\mathbb{P}_{A_1}}(\hat{\theta}^* - \hat{\theta})| = o_{\mathrm{Pr}^\circ}(n^{-1/2}) + o_{\mathrm{Pr}}(\|\hat{\theta}^* - \hat{\theta}\|).$$

As shown in the appendix,

$$\mathbb{P}_{\mathscr{R}(A_1)}(f^{\hat{\theta}, \hat{\theta}} - f^{\theta_1, \theta_1}) = \mathbb{P}_{\mathscr{R}(A_1)}(f^{\mathrm{r}\hat{\theta}, \hat{\theta}} - f^{\mathrm{r}\theta_1, \theta_1}). \tag{4.2}$$

Therefore, by (4.1) and Lemma 4.1 and irrespective of whether the null is true or not,

$$n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}^*} - \hat{\mathbb{P}}_{\hat{Z}^{*\mathrm{r}}})f = \hat{\mathbb{X}}_1^*(f) + o_{\mathrm{Pr}^\circ}(1) + o_{\mathrm{Pr}}(\|n^{1/2}(\hat{\theta}^* - \hat{\theta})\|) \qquad \text{unif. in } f \in \mathcal{F},$$

where

$$\hat{\mathbb{X}}_1^*(f) := n^{1/2}(\hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))} - \mathbb{P}_{\mathscr{R}(Z(\theta_1))})(f - f^{\mathrm{r}}) + \langle f - f^{\mathrm{r}}, (\partial_1 \log p_{A_1})\dot{\mu}_1' \rangle_{\mathbb{P}_{A_1}} n^{1/2}(\hat{\theta}^* - \hat{\theta}). \quad (4.3)$$

The empirical process $\{n^{1/2}(\hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))} - \mathbb{P}_{\mathscr{R}(Z(\theta_1))})(f - f^{\mathrm{r}}) : f \in \mathcal{F}\}$ converges in distribution in $\ell^\infty(\mathcal{F})$ to $\mathbb{G}_1$, a mean zero gaussian process with covariance function $\mathbb{E}\mathbb{G}_1 f_1 \mathbb{G}_1 f_2 = \mathbb{E}((f_1 - f_1^{\mathrm{r}})(Z(\theta_1))(f_2 - f_2^{\mathrm{r}})(Z(\theta_1)))$, $f_1, f_2 \in \mathcal{F}$; cf. the proof of Lemma 4.2. $\hat{\mathbb{X}}_1^*$ is thus a simulated version of $\hat{\mathbb{X}}$ that remains well defined even if the null hypothesis is false. This also illustrates the importance of re-estimating θ_0 using the simulated sample: If θ_0 was not re-estimated, i.e., set $\hat{\theta}^* = \hat{\theta}$ in (4.3), then $\hat{\mathbb{X}}_1^*$ would not mimic $\hat{\mathbb{X}}$ under the null and there would be no reason to believe why $n^{1/2}\widehat{\mathrm{KS}}_{\mathcal{F}}^*$ would possess the same limiting distribution as $n^{1/2}\widehat{\mathrm{KS}}_{\mathcal{F}}$ when the null hypothesis is true.

Let Pr^* and $\mathbb{E}^*$ stand for probability and expectation conditional on $\mathcal{D}_n$, i.e., integrals with respect to the Rademacher distribution (because given $\mathcal{D}_n$, the only source of randomness in $(Y^*, X, W, \hat{\theta})$ is R). Stochastic order symbols under Pr^* are written as o_{Pr^*} and O_{Pr^*}.

To derive the distribution of $\{\hat{\mathbb{X}}_1^*(f) : f \in \mathcal{F}\}$, we assume that

Assumption 4.3. (i) *Conditional on $\mathcal{D}_n$, $\hat{\theta}^*$ is asymptotically linear with influence function φ^*, i.e., $n^{1/2}(\hat{\theta}^* - \hat{\theta}) = n^{-1/2}\sum_{j=1}^n \varphi^*(Y_j^*, X_j, W_j, \hat{\theta}) + o_{\mathrm{Pr}^*}(1)$, where $\mathbb{E}^*\varphi^*(Y^*, X, W, \hat{\theta}) = 0$ and $\mathbb{E}^*\|\varphi^*(Y^*, X, W, \hat{\theta})\|^2 < \infty$. (ii) Lindeberg's condition is satisfied, i.e., for all $\epsilon > 0$ and $\lambda \in \mathbb{R}^{\dim(\Theta)}$, $\mathbb{E}^*[|\lambda'\varphi^*(Y^*, X, W, \hat{\theta})|^2 \mathbb{1}(|\lambda'\varphi^*(Y^*, X, W, \hat{\theta})| > \epsilon n^{1/2}\sigma_\lambda^*)] = o(\sigma_\lambda^{*2})$, where $\sigma_\lambda^{*2} := \mathbb{E}^*[n^{-1/2}\sum_{j=1}^n \lambda'\varphi^*(Y_j^*, X_j, W_j, \hat{\theta})]^2$. (iii) $\sigma_\lambda^{*2} = \mathbb{E}[\lambda'\varphi_1]^2 + o_{\mathrm{Pr}}(1)$, where $\varphi_1 := \varphi(Y, X, W, \theta_1)$ and φ is the influence function defined earlier in Assumption 3.5.*

This assumption is straightforward to verify if μ is linear in parameters, cf. Neumeyer, Dette, and Nagel (2005, p. 705). Conditional on $\mathcal{D}_n$, $n^{1/2}\lambda'(\hat{\theta}^* - \hat{\theta})/\sigma_\lambda^*$ converges in distribution to a standard gaussian random variable by (i), (ii), and Lindeberg's CLT. Hence, by (iii), $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ converges in distribution, conditional on $\mathcal{D}_n$, to N_{φ_1}, a mean zero gaussian random vector with variance $\mathbb{E}[\varphi_1\varphi_1']$. A dominated convergence argument (Andrews, 1997, p. 1101, footnote 2) then implies that $n^{1/2}(\hat{\theta}^* - \hat{\theta})$ also converges in distribution unconditionally to N_{φ_1}. This leads to the following result about the limiting distribution of $\hat{\mathbb{P}}_{\hat{Z}^*} - \hat{\mathbb{P}}_{\hat{Z}^{*\mathrm{r}}}$.

Lemma 4.2. *Let Assumptions 4.1–4.3 hold. Then, whether the null hypothesis is true or not, $\{n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}^*} - \hat{\mathbb{P}}_{\hat{Z}^{*\mathrm{r}}})f : f \in \mathcal{F}\} \xrightarrow{\mathrm{d}} \{\mathbb{X}_1(f) : f \in \mathcal{F}\}$ in $\ell^\infty(\mathcal{F})$ conditionally on $\mathcal{D}_n$, hence unconditionally, where $\mathbb{X}_1(f) := \mathbb{G}_1 f + \langle f - f^{\mathrm{r}}, (\partial_1 \log p_{A_1})\dot{\mu}_1' \rangle_{\mathbb{P}_{A_1}} N_{\varphi_1}$.*

It follows by the continuous mapping theorem that

Corollary 4.1. *Under the assumptions maintained in Lemma 4.2, $n^{1/2}\widehat{\mathrm{KS}}_{\mathcal{F}}^*$ converges in distribution to $\sup_{f \in \mathcal{F}} |\mathbb{X}_1(f)|$ whether the null hypothesis is true or not. Consequently, letting*

$\mathcal{F} = \mathcal{F}_1$, *the same holds for* $n^{1/2}\hat{R}^*$, *i.e., it converges in distribution to the random variable*

$$R_1 := \sup_{z \in \mathbb{R}^{1+\dim(W)}} |\mathbb{G}_1 \mathbb{1}_{(-\infty,z]} + \langle \mathbb{1}_{(-\infty,z]} - \mathbb{1}^{r}_{(-\infty,z]}, (\partial_1 \log p_{A_1})\dot{\mu}'_1 \rangle_{\mathbb{P}_{A_1}} N_{\varphi_1}|.$$

Therefore, $n^{1/2}\hat{R}^*$ has the same limiting distribution as $n^{1/2}\hat{R}$ under the null (because then $\theta_1 = \theta_0$) and is bounded in probability otherwise. Finally, we have that

Lemma 4.3. *Under Assumptions 4.1–4.3,* $n^{1/2}\hat{R}^*_{\max}$ *and* $n^{1/2}\hat{R}^*$ *have the same limiting distribution whether or not the null hypothesis is true.*

Since $n^{1/2}\hat{R}_{\max}$ and $n^{1/2}\hat{R}$ have the same asymptotic distribution under the null (cf. Lemma 3.3), it follows by Corollary 4.1 and Lemma 4.3 that $n^{1/2}\hat{R}^*_{\max}$ has the same limiting distribution as $n^{1/2}\hat{R}_{\max}$ when the null hypothesis is true and is bounded in probability otherwise. Therefore, under the null, the simulated critical value $c^*_{\alpha,B}$ converges in probability (conditional on $\mathcal{D}_n$) to c_α, the $1 - \alpha$ quantile of R_0, provided $B \to \infty$ as $n \to \infty$ (cf. Andrews, 1997, p. 1108). This completes our argument justifying the use of simulated critical values.

Remark. By Rademacherization, $(R\varepsilon(\theta_1), W) \overset{d}{=} (-R\varepsilon(\theta_1), W)$. Thus, for $f \in \mathcal{F}$,

$$\mathbb{P}_{\mathscr{R}(Z(\theta_1))} f = \mathbb{E} f(\mathscr{R}(Z(\theta_1))) = \mathbb{E} f(R\varepsilon(\theta_1), W)$$
$$= \mathbb{E} f(-R\varepsilon(\theta_1), W) = \mathbb{E} f^{r}(R\varepsilon(\theta_1), W) = \mathbb{P}_{\mathscr{R}(Z(\theta_1))} f^{r}.$$

Hence, $\mathbb{P}_{\mathscr{R}(Z(\theta_1))}(f - f^{r}) = 0$. Moreover,

$$f(\mathscr{R}(Z^{r}(\theta_1))) = f(-R\varepsilon(\theta_1), W) = f^{r}(R\varepsilon(\theta_1), W) = f^{r}(\mathscr{R}(Z(\theta_1)))$$

$$\implies \hat{\mathbb{P}}_{\mathscr{R}(Z^{r}(\theta_1))} f = n^{-1} \sum_{j=1}^{n} f(\mathscr{R}(Z^{r}_j)) = n^{-1} \sum_{j=1}^{n} f^{r}(\mathscr{R}(Z_j(\theta_1))) = \hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))} f^{r}.$$

Therefore, $(\hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))} - \hat{\mathbb{P}}_{\mathscr{R}(Z^{r}(\theta_1))}) f = (\hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))} - \mathbb{P}_{\mathscr{R}(Z(\theta_1))})(f - f^{r})$.

5. Large sample results under the alternative

We begin by showing that our test is consistent, i.e., it rejects any deviation from conditional symmetry with w.p.a.1.

Theorem 5.1. *Let* (1.3) *be true and Assumptions 3.1–3.4 hold with* (θ_0, ε) *replaced by* $(\theta_1, \varepsilon(\theta_1))$. *Then,* $\lim_{n\to\infty} n^{1/2}\widehat{KS}_{\mathcal{F}} = \infty$ *w.p.a.1.*

Hence, $n^{1/2}\widehat{KS}_{\mathcal{F}}$ will reject (1.2) w.p.a.1 as $n \to \infty$ because the simulated critical values are bounded in probability (cf. Section 4). Letting $\mathcal{F} = \mathcal{F}_1$ in Theorem 5.1, we have that $n^{1/2}\hat{R}$ is consistent. With a little additional effort, we can also show that

Theorem 5.2. $\lim_{n\to\infty} n^{1/2}\hat{R}_{\max} = \infty$ *w.p.a.1 under the conditions of Theorem 5.1.*

Hence, $n^{1/2}\hat{R}_{\max}$ is consistent against fixed alternatives as well. Next, we derive the power of $n^{1/2}\widehat{KS}_{\mathcal{F}}$ against a sequence of alternatives that lie in a $n^{1/2}$-neighborhood of the null. To create the local alternatives of interest, begin by assuming that the null hypothesis is true, i.e., $\mathbb{P}_Z = \mathbb{P}_{Z^r}$. Next, let θ_n be a sequence in Θ such that $\lim_{n\to\infty} \theta_n = \theta_0$ and assume that $A_n := (\varepsilon(\theta_n), X, W)$ is drawn from the perturbed measure $\mathbb{P}_{A_n} := \mathbb{P}_{A_0}(1 + n^{-1/2}h)$, where $h : \mathbb{R}^{1+\dim(W)} \to \mathbb{R}$ is such that $\|h\|_\infty := \sup_{\mathbb{R}\times\mathrm{supp}(W)} |h(u,w)| < \infty$ and $\int h\, d\mathbb{P}_Z = 0$ (these conditions ensure that $\mathbb{P}_{A_n}$ is a probability measures for $n \geq \|h\|_\infty^2$). Reflecting the first coordinate about the origin, this leads to a sequence of distributions for $Z(\theta_n) := (\varepsilon(\theta_n), W)$ and $Z^r(\theta_n) := (-\varepsilon(\theta_n), W)$ given by

$$H_{1n} : \qquad \mathbb{P}_{Z(\theta_n)} := \mathbb{P}_Z(1 + n^{-1/2}h) \quad \& \quad \mathbb{P}_{Z^r(\theta_n)} := \mathbb{P}_Z(1 + n^{-1/2}h^r). \qquad (5.1)$$

Henceforth, let $h^r \neq h$. Then, for $f \in \mathcal{F}$,

$$(5.1) \implies (\mathbb{P}_{Z(\theta_n)} - \mathbb{P}_{Z^r(\theta_n)})f = n^{-1/2}\mathbb{P}_Z(h - h^r)f =: n^{-1/2}\Gamma_h(f)$$

and $\sup_{f\in\mathcal{F}} |\Gamma_h(f)| > 0$ because $\mathcal{F}$ is measure determining and h is not symmetric in its first coordinate. Thus, (5.1) defines a sequence of local alternatives for (2.1).

Since observations under H_{1n} are independently but not identically distributed because the underlying measures $\mathbb{P}_{Z(\theta_n)}$ and $\mathbb{P}_{Z^r(\theta_n)}$ depend upon n, the assumptions introduced in Section 3 have to be strengthened in order to derive the distribution of $n^{1/2}\widehat{KS}_{\mathcal{F}}$ under H_{1n}. We begin with Assumption 3.1.

Assumption 5.1. *There exists a neighborhood of θ_0 such that for each θ_n in this neighborhood, $\mu(x, \cdot)$ is mean-square differentiable at θ_n, i.e., $\mu(x, \theta) = \mu(x, \theta_n) + \dot\mu'(x, \theta_n)(\theta - \theta_n) + \rho(x, \theta, \theta_n)$ with $\|\sup_{\theta\in\overline{B(\theta_n,\delta)}} \rho(\cdot, \theta, \theta_n)\|_{2,\mathbb{P}_X} = o(\delta)$ as $\delta \to 0$, $\|\dot\mu(\cdot, \theta_0)\| \in L_2(\mathbb{P}_X)$, and $\|\dot\mu(\cdot, \theta_n) - \dot\mu(\cdot, \theta_0)\|_{2,\mathbb{P}_X} = o(1)$.*

Next, we strengthen Assumption 3.2.

Assumption 5.2. (i) *There exists a neighborhood of θ_0 such that for each θ_n in this neighborhood, $p_{\varepsilon(\theta_n)|X=x,W=w}$ is a.e. differentiable on $\mathbb{R}$ and the conditional second moment of the derivative is uniformly bounded on $\mathbb{N} \times \mathrm{supp}(X) \times \mathrm{supp}(W)$.* (ii) *For each θ_n in the aforementioned neighborhood, the conditional distribution of $\varepsilon(\theta_n)$ given (X, W) is Lipschitz and the nonnegative Lipschitz constants (ζ_n) are such that $\sup_n \zeta_n \in L_2(\mathbb{P}_{X,W})$.*

Assumption 3.3 is modified as follows:

Assumption 5.3. (i), (ii), *and* (iv) *are same as* (i), (ii), *and* (iv) *of Assumption* 3.3. (iii) *The bracketing integral $\int_0^\infty \sup_{n\in\mathbb{N}} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(\mathbb{P}_{A_n}))}\, d\epsilon < \infty$.* (v) *Same as Assumption* 3.3(v) *but with $(\mathbb{P}_{A_0}, \theta_0)$ replaced by $(\mathbb{P}_{A_n}, \theta_n)$.*

$\mathcal{F}_1$ and $\mathcal{F}_2$ satisfy (iii) because indicators of orthants and half-spaces are VC, hence, universally Donsker (V&W, Example 2.5.4 and Problem 2.6.14). The argument in Lemma F.2 shows that $\mathcal{F}_1$ and $\mathcal{F}_2$ also satisfy (v) under Assumption 5.2(ii). By V&W (Theorem 2.8.4), (ii) and (iii) imply that $\mathcal{F}$ is Donsker and pre-gaussian uniformly in $(\mathbb{P}_{A_n})$. Since $\mathcal{F}$ and $-\mathcal{F}^{\mathrm{r}}$ are covered by the same number of (pointwise) $L_2(\mathbb{P}_{A_n})$-brackets,

$$N_{[]}(\epsilon, \mathcal{F} - \mathcal{F}^{\mathrm{r}}, L_2(\mathbb{P}_{A_n})) \leq N_{[]}(\epsilon, \mathcal{F}, L_2(\mathbb{P}_{A_n})) \times N_{[]}(\epsilon, -\mathcal{F}^{\mathrm{r}}, L_2(\mathbb{P}_{A_n})) = N_{[]}^2(\epsilon, \mathcal{F}, L_2(\mathbb{P}_{A_n})).$$

Hence, by (iii),

$$\int_0^\infty \sup_{n\in\mathbb{N}} \sqrt{\log N_{[]}(\epsilon, \mathcal{F} - \mathcal{F}^{\mathrm{r}}, L_2(\mathbb{P}_{A_n}))}\, d\epsilon \leq \int_0^\infty \sup_{n\in\mathbb{N}} \sqrt{2\log N_{[]}(\epsilon, \mathcal{F}, L_2(\mathbb{P}_{A_n}))}\, d\epsilon < \infty.$$

Therefore, by V&W (Theorem 2.8.4), $\mathcal{F} - \mathcal{F}^{\mathrm{r}}$ is Donsker and pre-gaussian uniformly in $(\mathbb{P}_{A_n})$. This fact is used in the proof of Lemma 5.2.

Finally, Assumption 3.4 becomes

Assumption 5.4. $\mathrm{plim}(\hat{\theta} - \theta_n) = 0.$

Under these conditions it is straightforward to show that Lemma 3.1 remains valid with θ_0 replaced by θ_n (the proof of the following result is virtually identical to the proof of Lemmas 3.1 and 4.1 and is therefore omitted), i.e.,

Lemma 5.1. *Let Assumptions* 5.1–5.4 *hold. Then, under* H_{1n},

$$\sup_{f\in\mathcal{F}} |(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{Z(\theta_n)})f - \langle f, (\partial_1 \log p_{A_n})\dot{\mu}_0'\rangle_{\mathbb{P}_{A_n}}(\hat{\theta} - \theta_n)| = o_{\mathrm{Pr}^\circ}(n^{-1/2}) + o_{\mathrm{Pr}}(\|\hat{\theta} - \theta_n\|)$$

$$\sup_{f\in\mathcal{F}} |(\hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}(\theta_n)})f - \langle f^{\mathrm{r}}, (\partial_1 \log p_{A_n})\dot{\mu}_0'\rangle_{\mathbb{P}_{A_n}}(\hat{\theta} - \theta_n)| = o_{\mathrm{Pr}^\circ}(n^{-1/2}) + o_{\mathrm{Pr}}(\|\hat{\theta} - \theta_n\|).$$

As in Section 3, we use Lemma 5.1 to derive the distribution of $n^{1/2}\widehat{\mathrm{KS}}_{\mathcal{F}}$ under H_{1n}. Begin by observing that

$$\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}} = (\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{Z(\theta_n)}) - (\hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}(\theta_n)}) + (\hat{\mathbb{P}}_{Z(\theta_n)} - \mathbb{P}_{Z(\theta_n)}) - (\hat{\mathbb{P}}_{Z^{\mathrm{r}}(\theta_n)} - \mathbb{P}_{Z^{\mathrm{r}}(\theta_n)})$$
$$+ (\mathbb{P}_{Z(\theta_n)} - \mathbb{P}_{Z^{\mathrm{r}}(\theta_n)}).$$

Hence, by (5.1) and Lemma 5.1,

$$n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f = \hat{\mathbb{X}}_{\theta_n}(f) + \Gamma_h(f) + o_{\mathrm{Pr}^\circ}(1) + o_{\mathrm{Pr}}(\|n^{1/2}(\hat{\theta} - \theta_n)\|) \quad \text{unif. in } f \in \mathcal{F},$$

where $\hat{\mathbb{X}}_{\theta_n}(f) := n^{1/2}(\hat{\mathbb{P}}_{Z(\theta_n)} - \mathbb{P}_{Z(\theta_n)})(f - f^{\mathrm{r}}) + \langle f - f^{\mathrm{r}}, (\partial_1 \log p_{A_n})\dot{\mu}'\rangle_{\mathbb{P}_{A_n}} n^{1/2}(\hat{\theta} - \theta_n)$. To identify the marginal distribution of $\hat{\mathbb{X}}_{\theta_n}$ assume that (cf. Andrews, 1997, Assumption E2(i)):

Assumption 5.5. (i) *Assumption* 3.5 *holds with* θ_0 *replaced by* θ_n. (ii) *The Lindeberg condition* $\mathbb{E}[|\lambda'\varphi_n|^2 \mathbb{1}(|\lambda'\varphi_n| > \epsilon n^{1/2}\sigma_{\lambda,n})] = o(\sigma_{\lambda,n}^2)$ *holds for all* $\epsilon > 0$ *and* $\lambda \in \mathbb{R}^{\dim(\Theta)}$, *where* $\varphi_n := \varphi(Y, X, W, \theta_n)$ *and* $\sigma_{\lambda,n}^2 := \mathbb{E}[\lambda'\varphi_n]^2 \to \mathbb{E}[\lambda'\varphi_0]^2$.

By Assumption 5.5, $n^{1/2}(\hat{\theta} - \theta_n)$ converges in distribution to N_{φ_0} by Lindeberg's CLT for inid random variables. Let $\partial_1 h$ exist and $\|\partial_1 h\|_\infty < \infty$. Since $\log p_{A_n} = \log p_{A_0} + \log(1 + n^{-1/2}h)$ (the second term is well defined for $n \geq \|h\|_\infty^2$),

$$\langle f - f^{\mathrm{r}}, (\partial_1 \log p_{A_n})\dot{\mu}_0\rangle_{\mathbb{P}_{A_n}} = \langle f - f^{\mathrm{r}}, (\partial_1 \log p_{A_0})\dot{\mu}_0\rangle_{\mathbb{P}_{A_n}} + n^{-1/2}\langle f - f^{\mathrm{r}}, (\partial_1 h)\dot{\mu}_0\rangle_{\mathbb{P}_{A_0}}$$

so that, by Cauchy-Schwarz,

$$\|\langle f - f^{\mathrm{r}}, (\partial_1 \log p_{A_n})\dot{\mu}_0\rangle_{\mathbb{P}_{A_n}} - \langle f - f^{\mathrm{r}}, (\partial_1 \log p_{A_0})\dot{\mu}_0\rangle_{\mathbb{P}_{A_0}}\|$$
$$\leq n^{-1/2}\langle f - f^{\mathrm{r}}, (h + \partial_1 h)(\partial_1 \log p_{A_0})\dot{\mu}_0\rangle_{\mathbb{P}_{A_0}}$$
$$\leq n^{-1/2}\|f - f^{\mathrm{r}}\|_{2,\mathbb{P}_{A_0}} \|h + \partial_1 h\|_\infty \|v_{A_0}\|_\infty \|(\dot{\mu}_0'\dot{\mu}_0)^{1/2}\|_{2,\mathbb{P}_X}.$$

Hence, for each $f \in \mathcal{F}$, $\hat{\mathbb{X}}_{\theta_n}(f)$ converges in distribution to $\mathbb{X}_0(f)$ by Lindeberg's CLT. Therefore, $n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})\mathcal{F}$ converges in distribution in $\ell^\infty(\mathcal{F})$ because $(\hat{\mathbb{X}}_{\theta_n} + \Gamma_h)\mathcal{F}$ is asymptotically tight.

Lemma 5.2. *Let Assumptions* 5.1–5.5 *hold. Then, under* H_{1n},

$$\{n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f : f \in \mathcal{F}\} \xrightarrow{\mathrm{d}} \{\mathbb{X}_0(f) + \Gamma_h(f) : f \in \mathcal{F}\} \quad in \ \ell^\infty(\mathcal{F}).$$

Consequently, by the continuous mapping theorem we have the limiting distribution of $n^{1/2}\widehat{\mathrm{KS}}_{\mathcal{F}}$ under H_{1n}.

Theorem 5.3. *Let Assumptions* 5.1–5.5 *hold. Then, under* H_{1n}, *the random variable* $n^{1/2}\widehat{\mathrm{KS}}_{\mathcal{F}}$ *converges in distribution to the random variable* $\sup_{f \in \mathcal{F}} |\mathbb{X}(f) + \Gamma_h(f)|$.

Letting $\mathcal{F} = \mathcal{F}_1$ in Theorem 5.3, we immediately get that

Corollary 5.1. *If Assumptions* 5.1–5.5 *hold then, under* H_{1n}, *the random variable* $n^{1/2}\hat{\mathrm{R}}$ *converges in distribution to the random variable*

$$\sup_{z \in \mathbb{R}^{1+\dim(W)}} |\mathbb{G}_0 \mathbb{1}_{(-\infty,z]} + \langle \mathbb{1}_{(-\infty,z]} - \mathbb{1}_{(-\infty,z]}^{\mathrm{r}}, (\partial_1 \log p_{A_0})\dot{\mu}_0'\rangle_{\mathbb{P}_{A_0}} N_{\varphi_0} + \Gamma_h(\mathbb{1}_{(-\infty,z]})|.$$

Given that we know that $n^{1/2}\hat{\mathrm{R}}_{\max}$ and $n^{1/2}\hat{\mathrm{R}}$ behave similarly under the null, it is also not surprising that

Lemma 5.3. *Under the conditions of Theorem* 5.3, $n^{1/2}\hat{\mathrm{R}}_{\max}$ *and* $n^{1/2}\hat{\mathrm{R}}$ *have the same limiting distribution under* H_{1n}.

Finally, as in Andrews (1997, p. 1114), it can be shown that $n^{1/2}\widehat{\mathrm{KS}}_{\mathcal{F}}$ is asymptotically locally unbiased (the same argument works for $n^{1/2}\hat{\mathrm{R}}_{\max}$ as well). Indeed, since $(\mathcal{F}, \|\cdot\|_{2,\mathbb{P}_{A_0}})$ is totally bounded (cf. the proof of Lemma 3.2), there exists a sequence of increasing finite sets $(\mathcal{F}_j)$ whose limit $\cup_{j=1}^\infty \mathcal{F}_j$ is dense (in the $\|\cdot\|_{2,\mathbb{P}_{A_0}}$ norm) in $\mathcal{F}$. Hence,

$$\sup_{f \in \mathcal{F}} |(\mathbb{X}_0 + \Gamma_h)f| = \sup_{f \in \overline{\cup_{j=1}^\infty \mathcal{F}_j}} |(\mathbb{X}_0 + \Gamma_h)f| = \sup_{f \in \cup_{j=1}^\infty \mathcal{F}_j} |(\mathbb{X}_0 + \Gamma_h)f|,$$

where the second equality follows because $\mathbb{X}_0 + \Gamma_h$ is $\|\cdot\|_{2,\mathbb{P}_{A_0}}$-continuous on $\mathcal{F}$ w.p.1. [Almost all sample paths of $\mathbb{X}_0(\mathcal{F})$ are uniformly $\|\cdot\|_{2,\mathbb{P}_{A_0}}$-continuous by V&W (Addendum 1.5.8), and uniform $\|\cdot\|_{2,\mathbb{P}_{A_0}}$-continuity of Γ_h follows because $|\Gamma_h(f)| \leq 2\|h\|_\infty \|f\|_{2,\mathbb{P}_{A_0}}$ for all $f \in \mathcal{F}$, i.e., Γ_h is a bounded linear functional on $\mathcal{F}$.] Therefore, if $B \to \infty$ as $n \to \infty$,

$$
\begin{aligned}
\lim_{n\to\infty} \mathrm{Pr}_{H_{1n}}(n^{1/2}\widehat{\mathrm{KS}}_\mathcal{F} > c^*_{\alpha,B}) &= \mathrm{Pr}(\sup_{f\in\mathcal{F}} |(\mathbb{X}_0 + \Gamma_h)f| > c_\alpha) \\
&= \mathrm{Pr}(\sup_{f\in\cup_{j=1}^\infty \mathcal{F}_j} |(\mathbb{X}_0 + \Gamma_h)f| > c_\alpha) \\
&= \lim_{j\to\infty} \mathrm{Pr}(\sup_{f\in\mathcal{F}_j} |(\mathbb{X}_0 + \Gamma_h)f| > c_\alpha) \quad \text{(continuity of prob. meas.)} \\
&\geq \lim_{j\to\infty} \mathrm{Pr}(\sup_{f\in\mathcal{F}_j} |\mathbb{X}_0(f)| > c_\alpha) \quad \text{(Anderson's lemma)} \\
&= \mathrm{Pr}(\sup_{f\in\cup_{j=1}^\infty \mathcal{F}_j} |\mathbb{X}_0(f)| > c_\alpha) \quad \text{(continuity of prob. meas.)} \\
&= \mathrm{Pr}(\sup_{f\in\mathcal{F}} |\mathbb{X}_0(f)| > c_\alpha) = \alpha.
\end{aligned}
$$

Hence, our test has non-trivial power against the local alternatives.

6. Monte Carlo study

In a seminal paper, Powell (1986b) showed how to identify and estimate censored or truncated regression models under the assumption that the error term in the latent regression model is symmetrically distributed conditional on the regressors (no additional distributional assumptions are needed). Applications of Powell's symmetric trimming approach that we are aware of, e.g., Levitt (1996), Chay and Powell (2001), Jacoby, Murgai, and Rehman (2004), and Kopczuk (2007), to name a few, simply assume the aforementioned conditional symmetry condition. However, since the latter is critical for Powell's approach to go through, it is important to have a test to determine whether the conditional symmetry assumption maintained in Powell's paper is supported by the data. In this section we demonstrate how our test can be applied for this purpose (we use the simulation designs in Powell's paper).

We begin by setting up Powell's model in our notation. Consider the linear regression model $Y^0 = X'\theta_0 + U^0$, where the response variable Y^0 is latent and U^0 is continuously distributed on $\mathbb{R}$. [Powell uses a "*" superscript to denote latent variables but we avoid this because in our paper a "*" superscript indicates simulated random variables, cf. Section 4.] Instead of observing Y^0, the researcher only observes its left-censored version

$$
Y := \begin{cases} Y^0 & \text{if } Y^0 > 0 \\ 0 & \text{otherwise} \end{cases} \iff U = \begin{cases} U^0 & \text{if } U^0 > -X'\theta_0 \\ -X'\theta_0 & \text{otherwise,} \end{cases}
$$

where $U := Y - X'\theta_0$ is the residual from the censored regression. We focus on left-censoring because we replicate Powell's simulation design; extension to right-censoring or truncation follows mutatis mutandis.

In his paper, Powell showed that if the structural error term U^0 in the latent model is symmetrically distributed about the origin conditional on the regressors, all of which were assumed to be exogenous, i.e.,

$$U^0|X \overset{\mathrm{d}}{=} -U^0|X, \tag{6.1}$$

then θ_0 is identified and the symmetrically censored least-squares (SCLS) estimator $\hat{\theta}$, defined via equation 2.10 of his paper, converges almost surely to θ_0.

Let V denote the symmetrically recensored version of U in Powell's approach, i.e.,

$$V := \begin{cases} U & \text{if } U < X'\theta_0 \\ X'\theta_0 & \text{otherwise} \end{cases} = \begin{cases} U^0 & \text{if } -X'\theta_0 < U^0 < X'\theta_0 \\ X'\theta_0 & \text{if } U^0 \geq X'\theta_0 \\ -X'\theta_0 & \text{if } U^0 \leq -X'\theta_0. \end{cases} \tag{6.2}$$

Since the conditional distribution of $V|X$ has support $[-X'\theta_0, X'\theta_0]$, it is necessary to assume that $X'\theta_0 > 0$ (at least with positive probability) to ensure that the support is not empty, cf. Powell (1986b, p. 1440). We use $\hat{\mathrm{R}}_{\max}$ to test (6.1) with the understanding that now $\hat{\varepsilon}$ in the definition of $\hat{\mathrm{R}}_{\max}$ is replaced by an estimator of V, namely,

$$\hat{V} := \begin{cases} \hat{U} & \text{if } \hat{U} < X'\hat{\theta} \\ X'\hat{\theta} & \text{if } \hat{U} \geq X'\hat{\theta}, \end{cases}$$

where $\hat{\theta}$ is Powell's SCLS estimator, $\hat{U} := Y - X'\hat{\theta}$, and all regressors are exogenous, i.e. $W = X$; thus, here

$$\hat{\mathrm{R}}_{\max} := \max_{(v,x) \in \{(\hat{V}_1, X_1), \ldots, (\hat{V}_n, X_n)\}} |n^{-1} \sum_{j=1}^{n} \mathbb{1}_{(-\infty, v] \times (-\infty, x]}(\hat{V}_j, X_j) - n^{-1} \sum_{j=1}^{n} \mathbb{1}_{(-\infty, v] \times (-\infty, x]}(-\hat{V}_j, X_j)|.$$

The motivation behind using $\hat{\mathrm{R}}_{\max}$ comes from the fact that, given X, the recensored residual V is symmetrically distributed about the origin if and only if U^0 is symmetrically distributed on $[-X'\theta_0, X'\theta_0]$, i.e.,

$$V|X \overset{\mathrm{d}}{=} -V|X \text{ on } [-X'\theta_0, X'\theta_0] \iff U^0|X \overset{\mathrm{d}}{=} -U^0|X \text{ on } [-X'\theta_0, X'\theta_0].$$

Indeed, the conditional density of $V|X = x$ (with respect to a dominating measure that is a mixture of the Lebesgue measure on $\mathbb{R}$ and the counting measure on $\{-x'\theta_0, x'\theta_0\}$) is given by

$$\mathrm{pdf}_{V|X=x}(t) := \mathrm{pdf}_{U^0|X=x}(t)\mathbb{1}_{(-x'\theta_0, x'\theta_0)}(t) + \mathbb{P}_{U^0|X=x}(U^0 \leq -x'\theta_0)\mathbb{1}_{\{-x'\theta_0\}}(t)$$
$$+ \mathbb{P}_{U^0|X=x}(U^0 \geq x'\theta_0)\mathbb{1}_{\{x'\theta_0\}}(t), \qquad t \in \mathbb{R}. \tag{6.3}$$

Hence, $\text{pdf}_{V|X=x}$ is symmetric on its support (which is $[-x'\theta_0, x'\theta_0]$) if and only if $\text{pdf}_{U^0|X=x}(t) = \text{pdf}_{U^0|X=x}(-t)$ for $t \in (-x'\theta_0, x'\theta_0)$ and $\mathbb{P}_{U^0|X=x}(U^0 \leq -x'\theta_0) = \mathbb{P}_{U^0|X=x}(U^0 \geq x'\theta_0)$. This makes it clear that the symmetrically censored residuals in Powell's approach can only be used to test whether (6.1) holds on the support of $\text{Law}(V|X)$.

The main technical difficulty in applying our test to verify Powell's conditional symmetry assumption comes from the fact that in Section 1 the error term ε was assumed to be continuously distributed with full support on $\mathbb{R}$ whereas the symmetrically recensored residual relevant for Powell's approach, defined in (6.2), has bounded support conditional on the regressors. However, as shown in Appendix E, our technical arguments still go through under some additional regularity conditions.

Before presenting the finite sample results we recall Powell's experimental setup; additional details can be found in Section 4 of his paper. Powell's latent model is the simple linear regression $Y^0 = \theta_0 + \theta_1 X + U^0$ with $X \stackrel{\text{d}}{=} \text{Unif}(-1.7, 1.7)$ and $U^0 \stackrel{\text{d}}{=} N(0, \sigma^2)$. The simulation results reported in Tables 1 and 2 are for testing that $V|X \stackrel{\text{d}}{=} -V|X$. Since Powell imposes independence between U^0 and X in his simulation design, by (6.3) this is equivalent to testing that $U^0 \stackrel{\text{d}}{=} -U^0$ on the support of $\text{Law}(V|X)$.

The empirical size of $n^{1/2}\hat{\text{R}}_{\max}$ is reported in Table 1 for three different sample sizes, namely, $n = 50, 100, 200$, and three specifications: $(\theta_0, \theta_1, \sigma^2) \in \{(0, 1, 2), (0, 1, 1), (1, 1, 1)\}$. These specifications are designs 3, 1 and 4 in Powell's paper and correspond to the cases when censoring is high (50%) and the scale of U^0 is high ($\sigma^2 = 2$), censoring is high and scale is low ($\sigma^2 = 1$), and censoring is low (25%) and scale is low. To keep the running time manageable, the results in Tables 1 and 2 are based on 300 simulations and 500 replications per simulation for simulating the critical values as described in Section 4 (i.e., we create a simulated sample (Y^*, X) with $Y^* := X'\hat{\theta} + R\hat{U}$ and use it to re-estimate θ^* by symmetric censoring to get $\hat{\theta}^*$, $\hat{U}^*$ and thus $\hat{V}^*$; the latter in turn is used to construct $\hat{\text{R}}^*_{\max}$ for simulating the critical values). Code for the simulation was written in R and is available from the authors.

The results in Table 1 suggest that our test is working pretty well. For each sample size we fail to reject the hypothesis that the true size is statistically different (at 5% level of significance) from its nominal value as the corresponding t-statistics are all bounded by 2. This finding holds whether censoring and scale are high or low suggesting that the size of $n^{1/2}\hat{\text{R}}_{\max}$ also appears to be fairly robust to the degree of censoring and the scale of U^0.

We also carried out a small simulation to examine the power of $n^{1/2}\hat{\text{R}}_{\max}$ in finite samples. The latent response in Powell's design was generated with $U^0 \stackrel{\text{d}}{=} \text{LogNormal}(\mu, \sigma^2)$ (an asymmetric distribution) and the parameters $\theta_0, \theta_1, \mu, \sigma^2$ were chosen to produce three levels of censoring, namely, $\approx 3\%$ (low), $\approx 25\%$ (medium), and $\approx 50\%$ (high), and three levels of skewness of V, namely, ≈ -0.2 (low), ≈ -0.5 (medium), and ≈ 0.9 (high). Since increasing the skewness of V should make the asymmetry of U^0 easier to detect, the least favorable model

TABLE 1. Empirical size of $n^{1/2}\hat{R}_{max}$.

			Nominal size (α)	
Design	n	0.01	0.05	0.10
	50	.01 (.006)	.037 (.013)	.077 (.017)
censoring $= 50\%$, scale $= 2$	100	.02 (.006)	.06 (.013)	.113 (.017)
	200	.013 (.006)	.05 (.013)	.097 (.017)
	50	.003 (.006)	.047 (.013)	.08 (.017)
censoring $= 50\%$, scale $= 1$	100	.013 (.006)	.057 (.013)	.12 (.017)
	200	.013 (.006)	.04 (.013)	.087 (.017)
	50	.003 (.006)	.057 (.013)	.087 (.017)
censoring $= 25\%$, scale $= 1$	100	.01 (.006)	.033 (.013)	.087 (.017)
	200	.013 (.006)	.05 (.013)	.087 (.017)

The last three columns report the fraction of simulations for which $n^{1/2}\hat{R}_{max} > c^*_{\alpha,B}$.
Monte Carlo standard errors in parenthesis.

in Table 2 is Design 3 (corresponding to high censoring and low skewness) whereas the most favorable one is Design 1 (low censoring and medium skewness). The results in Table 2 suggest that censoring and skewness of V both seem to affect the power of the test. Rejection rates are high for the low censoring level even when $n = 50$. Keeping the skewness of V fixed, an increase in the censoring reduces power although the rejection rates are close to ideal for $n = 200$. Overall, these results suggest that $n^{1/2}\hat{R}_{max}$ appears to have good power for moderate sample sizes even when the censoring is high and V is not very skewed.

7. Conclusion

We have shown how to test for conditional symmetry in the presence of endogenous regressors without making any distributional assumptions and without doing nonparametric smoothing. The Kolmogorov-Smirnov type statistic we propose is easy to implement, does not require optimization over an uncountable set, and can detect $n^{1/2}$-deviations from the null. Results from a simulation experiment suggest that our test can work very well in moderately sized samples even under nonstandard conditions.

TABLE 2. Empirical power of $n^{1/2}\hat{R}_{\max}$.

Design			n	0.01	0.05	0.10
					Nominal size (α)	
1	$\theta_0 = 1, \theta_1 = 1$	censoring $= 3\%$	50	.513	.763	.847
			100	.937	.993	1
	$\mu = 0, \sigma^2 = 1$	skewness $= 0.56$	200	1	1	1
2	$\theta_0 = 2, \theta_1 = 4.5$	censoring $= 27\%$	50	.453	.707	.8
			100	.917	.983	.997
	$\mu = 0, \sigma^2 = 1$	skewness $= -0.26$	200	1	1	1
3	$\theta_0 = -0.5, \theta_1 = 1.25$	censoring $= 48\%$	50	.11	.307	.483
			100	.457	.707	.81
	$\mu = -1, \sigma^2 = 1$	skewness $= -0.25$	200	.857	.94	.98
4	$\theta_0 = -0.5, \theta_1 = 3$	censoring $= 49\%$	50	.27	.527	.687
			100	.74	.913	.957
	$\mu = -1, \sigma^2 = 1$	skewness $= -0.53$	200	.99	1	1
5	$\theta_0 = 0, \theta_1 = 1$	censoring $= 50\%$	50	.413	.63	.757
			100	.883	.96	.98
	$\mu = -5, \sigma^2 = 1$	skewness $= 0.94$	200	1	1	1

The last three columns report the fraction of simulations for which $n^{1/2}\hat{R}_{\max} > c^*_{\alpha,B}$.

APPENDIX A. PROOFS FOR SECTION 2

Proof of (2.1). Let $p_{\varepsilon,W}$ denote the density of (ε, W) with respect to an appropriate dominating measure. Then,

$$(\varepsilon, W) \stackrel{\mathrm{d}}{=} (-\varepsilon, W) \iff p_{\varepsilon,W}(u,w) = p_{-\varepsilon,W}(u,w) \qquad \forall(u,w) \in \mathbb{R} \times \mathrm{supp}(W)$$
$$\iff p_{\varepsilon|W=w}(u)p_W(w) = p_{-\varepsilon|W=w}(u)p_W(w) \quad \forall(u,w) \in \mathbb{R} \times \mathrm{supp}(W)$$
$$\iff p_{\varepsilon|W=w}(u) = p_{-\varepsilon|W=w}(u) \qquad \forall(u,w) \in \mathbb{R} \times \mathrm{supp}(W)$$
$$\iff \varepsilon|W \stackrel{\mathrm{d}}{=} -\varepsilon|W. \qquad \square$$

APPENDIX B. PROOFS FOR SECTION 3

Proof of Lemma 3.1. Let $m_{A_0} := (\partial_1 \log p_{A_0})\dot{\mu}_0$. We only show that

$$\sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_Z)f - \langle f, m'_{A_0} \rangle_{\mathbb{P}_{A_0}}(\hat{\theta} - \theta_0)| = o_{\mathrm{Pr}^\circ}(n^{-1/2}) + o_{\mathrm{Pr}}(\|\hat{\theta} - \theta_0\|). \qquad (B.1)$$

Since $(\hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}})f = (\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_Z)f^{\mathrm{r}}$, the corresponding result for $\hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}}$ follows by replacing f with f^{r} in (B.1). So let $f \in \mathcal{F}$ and recall that $\Delta(X, \theta, \theta_0) := \mu(X, \theta) - \mu(X, \theta_0)$. Since

$$f(Y - \mu(X, \theta), W) = f(\varepsilon - \Delta(X, \theta, \theta_0), W) =: f^\theta(\varepsilon, X, W),$$

we have

$$(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_Z)f = n^{-1}\sum_{j=1}^{n}[f(\varepsilon_j - \Delta(X_j, \hat{\theta}, \theta_0), W_j) - f(\varepsilon_j - \Delta(X_j, \theta_0, \theta_0), W_j)]$$

$$= n^{-1}\sum_{j=1}^{n}[f^{\hat{\theta}}(\varepsilon_j, X_j, W_j) - f^{\theta_0}(\varepsilon_j, X_j, W_j)]$$

$$= \hat{\mathbb{P}}_{A_0}(f^{\hat{\theta}} - f^{\theta_0}).$$

Hence, we can write $(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_Z)f = (\hat{\mathbb{P}}_{A_0} - \mathbb{P}_{A_0})(f^{\hat{\theta}} - f^{\theta_0}) + \mathbb{P}_{A_0}(f^{\hat{\theta}} - f^{\theta_0})$. Consequently, to prove (B.1) it suffices to show that

$$\sup_{f \in \mathcal{F}}|n^{1/2}(\hat{\mathbb{P}}_{A_0} - \mathbb{P}_{A_0})(f^{\hat{\theta}} - f^{\theta_0})| = o_{\mathrm{Pr}^\circ}(1) \tag{B.2}$$

$$\sup_{f \in \mathcal{F}}|\mathbb{P}_{A_0}(f^{\hat{\theta}} - f^{\theta_0}) - \langle f, m'_{A_0}\rangle_{\mathbb{P}_{A_0}}(\hat{\theta} - \theta_0)| = o_{\mathrm{Pr}}(\|\hat{\theta} - \theta_0\|). \tag{B.3}$$

We will use equicontinuity of the empirical process $n^{1/2}(\hat{\mathbb{P}}_{A_0} - \mathbb{P}_{A_0})$ to demonstrate (B.2) and mean-square differentiability of $p_{A_0}^{1/2}$ to show that (B.3) holds. [Cf. the proof of Proposition 2.2 in Khmaladze and Koul (2004) for a similar approach.]

We begin with (B.2). Given $f \in \mathcal{F}$, we know that $f^\theta, f^{\theta_0} \in \mathcal{F}$ by Assumption 3.3(iv) and $\|f^\theta - f^{\theta_0}\|_{2,\mathbb{P}_{A_0}} \le q(\|\theta - \theta_0\|)$ by Assumption 3.3(v), where q is continuous and passes through the origin, i.e., $q(0) = 0$. Hence, as

$$|(\hat{\mathbb{P}}_{A_0} - \mathbb{P}_{A_0})(f^\theta - f^{\theta_0})| \le \sup_{f^\theta, f^{\theta_0} \in \mathcal{F}: \|f^\theta - f^{\theta_0}\|_{2,\mathbb{P}_{A_0}} \le q(\|\theta - \theta_0\|)}|(\hat{\mathbb{P}}_{A_0} - \mathbb{P}_{A_0})(f^\theta - f^{\theta_0})|$$

$$\le \sup_{g, h \in \mathcal{F}: \|g - h\|_{2,\mathbb{P}_{A_0}} \le q(\|\theta - \theta_0\|)}|(\hat{\mathbb{P}}_{A_0} - \mathbb{P}_{A_0})(g - h)|$$

and the right hand side does not depend upon f,

$$\sup_{f \in \mathcal{F}}|(\hat{\mathbb{P}}_{A_0} - \mathbb{P}_{A_0})(f^\theta - f^{\theta_0})| \le \sup_{g, h \in \mathcal{F}: \|g - h\|_{2,\mathbb{P}_{A_0}} \le q(\|\theta - \theta_0\|)}|(\hat{\mathbb{P}}_{A_0} - \mathbb{P}_{A_0})(g - h)|. \tag{B.4}$$

Since $\mathcal{F}$ is $\mathbb{P}_{A_0}$-Donsker by Assumption 3.3(iii), the empirical process $\{n^{1/2}(\hat{\mathbb{P}}_{A_0} - \mathbb{P}_{A_0})f : f \in \mathcal{F}\}$ is asymptotically equicontinuous (V&W, Section 2.1.2), i.e., $\forall \epsilon > 0$,

$$\lim_{\theta \to \theta_0}\limsup_{n \to \infty}\mathrm{Pr}^\circ(\sup_{g, h \in \mathcal{F}: \|g - h\|_{2,\mathbb{P}_{A_0}} \le q(\|\theta - \theta_0\|)}|n^{1/2}(\hat{\mathbb{P}}_{A_0} - \mathbb{P}_{A_0})(g - h)| > \epsilon) = 0. \tag{B.5}$$

Therefore, (B.2) follows by (B.4) and (B.5) because $\hat{\theta}$ is a consistent estimator of θ_0 (Assumption 3.4).

Next, we show (B.3). Begin by observing that, for $\theta \in \Theta$,

$$
\begin{aligned}
\mathbb{P}_{A_0} f^{\theta} &= \int f(u - \Delta(x, \theta, \theta_0), w) \, \mathbb{P}_{A_0}(du, dx, dw) \\
&= \int f(u - \Delta(x, \theta, \theta_0), w) p_{A_0}(u, x, w) \, du \, \kappa(dx, dw) \\
&= \int f(t, w) p_{A_0}(t + \Delta(x, \theta, \theta_0), x, w) \, dt \, \kappa(dx, dw)
\end{aligned}
$$

by the translation invariance of Lebesgue measure. Hence,

$$
\mathbb{P}_{A_0}(f^{\hat{\theta}} - f^{\theta_0}) = \int f(t, w)[p_{A_0}(t + \Delta(x, \hat{\theta}, \theta_0), x, w) / p_{A_0}(t, x, w) - 1] \, \mathbb{P}_{A_0}(dt, dx, dw). \quad \text{(B.6)}
$$

Since

$$
\frac{p_{A_0}^{1/2}(u + \Delta(x, \theta, \theta_0), x, w) - p_{A_0}^{1/2}(u, x, w)}{p_{A_0}^{1/2}(u, x, w)} \stackrel{(3.1)}{=} \frac{1}{2} \frac{\partial_1 p_{A_0}(u, x, w)}{p_{A_0}(u, x, w)} \Delta(x, \theta, \theta_0)
$$
$$
+ r(\Delta(x, \theta, \theta_0), u, x, w),
$$

it follows that

$$
\begin{aligned}
\frac{p_{A_0}(t + \Delta(x, \hat{\theta}, \theta_0), x, w)}{p_{A_0}(t, x, w)} - 1 &= (\partial_1 \log p_{A_0}(t, x, w)) \Delta(x, \hat{\theta}, \theta_0) \\
&\quad + 0.25(\partial_1 \log p_{A_0}(t, x, w))^2 \Delta^2(x, \hat{\theta}, \theta_0) \\
&\quad + r^2(\Delta(x, \hat{\theta}, \theta_0), t, x, w) + 2r(\Delta(x, \hat{\theta}, \theta_0), t, x, w) \\
&\quad + r(\Delta(x, \hat{\theta}, \theta_0), t, x, w)(\partial_1 \log p_{A_0}(t, x, w)) \Delta(x, \hat{\theta}, \theta_0). \quad \text{(B.7)}
\end{aligned}
$$

We use the decomposition in (B.7) to handle (B.6). By Assumption 3.1,

$$
\begin{aligned}
\int f(t, w)(\partial_1 \log p_{A_0}(t, x, w)) & \Delta(x, \hat{\theta}, \theta_0) \, \mathbb{P}_{A_0}(dt, dx, dw) \\
&= \int f(t, w)(\partial_1 \log p_{A_0}(t, x, w)) \dot{\mu}_0'(x)(\hat{\theta} - \theta_0) \, \mathbb{P}_{A_0}(dt, dx, dw) \\
&\quad + \int f(t, w)(\partial_1 \log p_{A_0}(t, x, w)) \rho(x, \hat{\theta}) \, \mathbb{P}_{A_0}(dt, dx, dw).
\end{aligned}
$$

By Assumptions 3.2 and 3.3(ii), $\|v_{A_0}\|_\infty \vee M_{\mathcal{F}} < \infty$. Hence, by Jensen and Assumption 3.1,

$$
\begin{aligned}
&|\int f(t,w)(\partial_1 \log p_{A_0}(t,x,w))\rho(x,\hat{\theta})\,\mathbb{P}_{A_0}(dt,dx,dw)|^2 \\
&\qquad \le M_{\mathcal{F}}^2 \int (\partial_1 \log p_{A_0}(t,x,w))^2 \left(\sup_{\theta \in \overline{B(\theta_0, \|\hat{\theta}-\theta_0\|)}} |\rho(x,\theta,\theta_0)| \right)^2 \mathbb{P}_{A_0}(dt,dx,dw) \\
&\qquad = M_{\mathcal{F}}^2 \int v_{A_0}(x,w) \sup_{\theta \in \overline{B(\theta_0, \|\hat{\theta}-\theta_0\|)}} \rho^2(x,\theta,\theta_0)\, \mathbb{P}_{X,W}(dx,dw) \\
&\qquad \le M_{\mathcal{F}}^2 \|v_{A_0}\|_\infty \int \sup_{\theta \in \overline{B(\theta_0, \|\hat{\theta}-\theta_0\|)}} \rho^2(x,\theta,\theta_0)\, \mathbb{P}_X(dx) \\
&\qquad = o(\|\hat{\theta}-\theta_0\|^2).
\end{aligned}
\tag{B.8}
$$

Therefore,

$$
\sup_{f \in \mathcal{F}} | \int f(t,w)(\partial_1 \log p_{A_0}(t,x,w))\Delta(x,\hat{\theta},\theta_0)\, \mathbb{P}_{A_0}(dt,dx,dw)
$$
$$
- \langle f, m'_{A_0}\rangle_{\mathbb{P}_{A_0}}(\hat{\theta}-\theta_0)| = o(\|\hat{\theta}-\theta_0\|). \tag{B.9}
$$

Next, by Assumption 3.1,

$$
\begin{aligned}
&|\int f(t,w)(\partial_1 \log p_{A_0}(t,x,w))^2 \Delta^2(x,\hat{\theta},\theta_0)\, \mathbb{P}_{A_0}(dt,dx,dw)| \\
&\le 2M_{\mathcal{F}} \int (\partial_1 \log p_{A_0}(t,x,w))^2 (\|\dot{\mu}_0(x)\|^2 \|\hat{\theta}-\theta_0\|^2 + (\sup_{\theta \in \overline{B(\theta_0, \|\hat{\theta}-\theta_0\|)}} |\rho(x,\theta,\theta_0)|)^2)\, \mathbb{P}_{A_0}(dt,dx,dw) \\
&= 2M_{\mathcal{F}} \int v_{A_0}(x,w)(\|\dot{\mu}_0(x)\|^2 \|\hat{\theta}-\theta_0\|^2 + \sup_{\theta \in \overline{B(\theta_0, \|\hat{\theta}-\theta_0\|)}} \rho^2(x,\theta,\theta_0))\, \mathbb{P}_{X,W}(dx,dw) \\
&\le 2M_{\mathcal{F}}\|v_{A_0}\|_\infty \int (\|\dot{\mu}_0(x)\|^2 \|\hat{\theta}-\theta_0\|^2 + \sup_{\theta \in \overline{B(\theta_0, \|\hat{\theta}-\theta_0\|)}} \rho^2(x,\theta,\theta_0))\, \mathbb{P}_X(dx) \\
&\le 2M_{\mathcal{F}}\|v_{A_0}\|_\infty (\int \|\dot{\mu}_0(x)\|^2\, \mathbb{P}_X(dx) + o(1))\|\hat{\theta}-\theta_0\|^2.
\end{aligned}
$$

Therefore,

$$
\sup_{f \in \mathcal{F}} | \int f(t,w)(\partial_1 \log p_{A_0}(t,x,w))^2 \Delta^2(x,\hat{\theta},\theta_0)\, \mathbb{P}_{A_0}(dt,dx,dw)| = O(\|\hat{\theta}-\theta_0\|^2)
$$
$$
= o_{\text{Pr}}(\|\hat{\theta}-\theta_0\|) \tag{B.10}
$$

because $\text{plim}(\hat{\theta}) = \theta_0$. Next, let $\epsilon > 0$. Then, by Assumption 3.1 and (3.1),

$$|\int f(t,w)r^2(\Delta(x,\hat{\theta},\theta_0),t,x,w)\,\mathbb{P}_{A_0}(dt,dx,dw)|$$

$$\leq M_{\mathcal{F}}\int r^2(\Delta(x,\hat{\theta},\theta_0),t,x,w)\,\mathbb{P}_{A_0}(dt,dx,dw)$$

$$\leq M_{\mathcal{F}}\epsilon\int \Delta^2(x,\hat{\theta},\theta_0)\,\mathbb{P}_X(dx)$$

$$\leq 2M_{\mathcal{F}}\epsilon\int (\|\dot{\mu}_0(x)\|^2\,\|\hat{\theta}-\theta_0\|^2 + \sup_{\theta\in B(\theta_0,\|\hat{\theta}-\theta_0\|)}\rho^2(x,\theta,\theta_0))\,\mathbb{P}_X(dx)$$

$$\leq 2M_{\mathcal{F}}\epsilon(\int \|\dot{\mu}_0(x)\|^2\,\mathbb{P}_X(dx) + o(1))\|\hat{\theta}-\theta_0\|^2.$$

Therefore, since ϵ was arbitrary,

$$\sup_{f\in\mathcal{F}}|\int f(t,w)r^2(\Delta(x,\hat{\theta},\theta_0),t,x,w)\,\mathbb{P}_{A_0}(dt,dx,dw)| = o(\|\hat{\theta}-\theta_0\|^2). \qquad (\text{B}.11)$$

A similar argument using Jensen's inequality reveals that

$$\sup_{f\in\mathcal{F}}|\int f(t,w)r(\Delta(x,\hat{\theta},\theta_0),t,x,w)\,\mathbb{P}_{A_0}(dt,dx,dw)| = o(\|\hat{\theta}-\theta_0\|). \qquad (\text{B}.12)$$

Finally, since

$$|\int f(t,w)r(\Delta(x,\hat{\theta},\theta_0),t,x,w)\partial_1\log p_{A_0}(t,x,w)\Delta(x,\hat{\theta},\theta_0)\,\mathbb{P}_{A_0}(dt,dx,dw)|$$

$$\leq (\int |f(t,w)|r^2(\Delta(x,\hat{\theta},\theta_0),t,x,w)\mathbb{P}_{A_0}(dt,dx,dw))^{1/2}$$

$$\times (\int |f(t,w)|(\partial_1\log p_{A_0}(t,x,w))^2\Delta^2(x,\hat{\theta},\theta_0)\,\mathbb{P}_{A_0}(dt,dx,dw))^{1/2}$$

by Cauchy-Schwarz, the arguments leading to (B.11) and (B.10) show that

$$\sup_{f\in\mathcal{F}}|\int f(t,w)r(\Delta(x,\hat{\theta},\theta_0),t,x,w)(\partial_1\log p_{A_0}(t,x,w))\Delta(x,\hat{\theta},\theta_0)\,\mathbb{P}_{A_0}(dt,dx,dw)|$$

$$= o(\|\hat{\theta}-\theta_0\|^2). \quad (\text{B}.13)$$

Therefore, (B.3) follows by (B.6)–(B.13). $\qquad\square$

Proof of Lemma 3.2. Recall that $\hat{\mathbb{X}}_0(f) := n^{1/2}(\hat{\mathbb{P}}_Z-\mathbb{P}_Z)(f-f^{\mathrm{r}})+\langle f-f^{\mathrm{r}}, m'_{A_0}\rangle_{\mathbb{P}_{A_0}}n^{1/2}(\hat{\theta}-\theta_0)$, $f\in\mathcal{F}$. Assumption 3.3(ii) implies that the sample paths of $\hat{\mathbb{X}}_0$ are bounded functions on $\mathcal{F}$. Hence, by V&W (Theorem 1.5.4), $\{\hat{\mathbb{X}}_0(f) : f\in\mathcal{F}\}$ converges in distribution in $\ell^\infty(\mathcal{F})$ to a tight limit process $\{\mathbb{X}_0(f) : f\in\mathcal{F}\}\subset\ell^\infty(\mathcal{F})$ if $\{\hat{\mathbb{X}}_0(f) : f\in\mathcal{F}\}$ is asymptotically tight and its marginals $\hat{\mathbb{X}}_0(f_1),\ldots,\hat{\mathbb{X}}_0(f_k)$ converge in distribution in $\mathbb{R}^k$ to the marginals $\mathbb{X}_0(f_1),\ldots,\mathbb{X}_0(f_k)$ for every finite subset $f_1,\ldots,f_k$ of $\mathcal{F}$. From (3.3) we know that the limiting process, if it exists,

is given by $\mathbb{X}_0(f) := \mathbb{G}_0 f + \langle f - f^{\mathrm{r}}, m'_{A_0} \rangle_{\mathbb{P}_{A_0}} N_{\varphi_0}$. To verify its existence, it only remains to show that $\{\hat{\mathbb{X}}_0(f) : f \in \mathcal{F}\}$ is asymptotically tight. We proceed as follows. First, for each $f \in \mathcal{F}$, $\hat{\mathbb{X}}_0(f)$ is asymptotically tight in $\mathbb{R}$ by (3.3). Next, since $\mathcal{F}$ is $\mathbb{P}_{A_0}$-Donsker (Assumption 3.3(iii)), $(\mathcal{F}, \|\cdot\|_{2,\mathbb{P}_{A_0}})$ is totally bounded by Assumption 3.3(ii) and V&W (Problem 2.1.1). Hence, by V&W (Theorem 1.5.7), $\{\hat{\mathbb{X}}_0(f) : f \in \mathcal{F}\}$ is asymptotically tight if it is asymptotically uniformly $\|\cdot\|_{2,\mathbb{P}_{A_0}}$-equicontinuous in probability, i.e., for all $\epsilon > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \Pr^{\circ}\left(\sup_{f,g \in \mathcal{F}_{\delta, \mathbb{P}_{A_0}}} |\hat{\mathbb{X}}_0(f - g)| > \epsilon \right) = 0, \tag{B.14}$$

where $\mathcal{F}_{\delta, \mathbb{P}_{A_0}} := \{f, g \in \mathcal{F} : \|f - g\|_{2, \mathbb{P}_{A_0}} \leq \delta\}$. Thus, we are done if we can show (B.14). Since $(f - g)^{\mathrm{r}} = f^{\mathrm{r}} - g^{\mathrm{r}}$, by definition of $\hat{\mathbb{X}}_0$ and the triangle inequality

$$\Pr^{\circ}\left(\sup_{f,g \in \mathcal{F}_{\delta, \mathbb{P}_{A_0}}} |\hat{\mathbb{X}}_0(f - g)| > 2\epsilon \right) \leq a_{\delta,n} + b_{\delta,n},$$

where $a_{\delta,n} := \Pr^{\circ}(\sup_{f,g \in \mathcal{F}_{\delta, \mathbb{P}_{A_0}}} |n^{1/2}(\hat{\mathbb{P}}_Z - \mathbb{P}_Z)((f - f^{\mathrm{r}}) - (g - g^{\mathrm{r}}))| > \epsilon)$ and

$$b_{\delta,n} := \Pr^{\circ}\left(\sup_{f,g \in \mathcal{F}_{\delta, \mathbb{P}_{A_0}}} |\langle (f - f^{\mathrm{r}}) - (g - g^{\mathrm{r}}), m'_{A_0} \rangle_{\mathbb{P}_{A_0}} n^{1/2}(\hat{\theta} - \theta_0)| > \epsilon \right).$$

Now, $\mathcal{F}^{\mathrm{r}}$ is $\mathbb{P}_{A_0}$-Donsker because this is equivalent to $\mathcal{F}$ being $\mathbb{P}_{A_0}$-Donsker (Assumption 3.3(iii)). Consequently, $\mathcal{F} - \mathcal{F}^{\mathrm{r}} := \{f - g^{\mathrm{r}} : f, g \in \mathcal{F}\}$ is $\mathbb{P}_{A_0}$-Donsker by V&W (Theorem 2.10.2) because the difference map is Lipschitz. Hence, since the empirical process $n^{1/2}(\hat{\mathbb{P}}_Z - \mathbb{P}_Z)(\mathcal{F} - \mathcal{F}^{\mathrm{r}})$ is asymptotically tight under $\|\cdot\|_{2,\mathbb{P}_{A_0}}$ (V&W (2.1.8)),

$$\lim_{\delta \to 0} \limsup_{n \to \infty} a_{\delta,n} \leq \lim_{\delta \to 0} \limsup_{n \to \infty} \Pr^{\circ}\left(\sup_{f,g \in (\mathcal{F} - \mathcal{F}^{\mathrm{r}})_{\delta, \mathbb{P}_{A_0}}} |n^{1/2}(\hat{\mathbb{P}}_Z - \mathbb{P}_Z)(f - g)| > \epsilon \right) = 0.$$

Next, by repeated applications of Cauchy-Schwarz,

$$|\langle (f - f^{\mathrm{r}}) - (g - g^{\mathrm{r}}), m'_{A_0} \rangle_{\mathbb{P}_{A_0}}(\hat{\theta} - \theta_0)| \leq \|\langle (f - g) - (f^{\mathrm{r}} - g^{\mathrm{r}}), m_{A_0} \rangle_{\mathbb{P}_{A_0}}\| \|\hat{\theta} - \theta_0\|$$

$$\leq 2\|f - g\|_{2, \mathbb{P}_{A_0}} \|v_{A_0}\|_{\infty} \|(\dot{\mu}'_0 \dot{\mu}_0)^{1/2}\|_{2, \mathbb{P}_X} \|\hat{\theta} - \theta_0\|$$

because $\|f^{\mathrm{r}} - g^{\mathrm{r}}\|_{2, \mathbb{P}_{A_0}} = \|f - g\|_{2, \mathbb{P}_{A_0}}$ for all $f, g \in \mathcal{F}$ under the null hypothesis. Indeed,

$$\|f^{\mathrm{r}} - g^{\mathrm{r}}\|^2_{2, \mathbb{P}_{A_0}} = \int (f - g)^2(-u, w) \, \mathbb{P}_{\varepsilon, X, W}(du, dx, dw)$$

$$= \int (f - g)^2(-u, w) \, \mathbb{P}_{\varepsilon, W}(du, dw) \qquad \text{(marginal integration)}$$

$$= \int (f - g)^2(u, w) \, \mathbb{P}_{\varepsilon, W}(du, dw) \qquad \text{(change of variables and (2.1))}$$

$$= \int (f - g)^2(u, w) \, \mathbb{P}_{\varepsilon, X, W}(du, dx, dw) \qquad \text{(marginal integration)}$$

$$= \|f - g\|^2_{2, \mathbb{P}_{A_0}}.$$

Hence,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} b_{\delta,n} \leq \lim_{\delta \to 0} \limsup_{n \to \infty} \Pr^\circ(\epsilon < 2\delta \|v_{A_0}\|_\infty \|(\dot{\mu}_0' \dot{\mu}_0)^{1/2}\|_{2,\mathbb{P}_X} \|n^{1/2}(\hat{\theta} - \theta_0)\|) = 0$$

because $\hat{\theta}$ is a $n^{1/2}$-consistent estimator of θ_0 (Assumption 3.5). Therefore, (B.14) holds. $\square$

Proof of Lemma 3.3. We follow the approach taken by Andrews (1997) in proving his Theorem A.2 while accounting for the fact that (unlike Andrews) our statistic is optimized over the estimated set of observations $\hat{\mathcal{Z}}$. We prove the stated result by showing that the cdf of the nonnegative random variable $n^{1/2}\hat{R}_{\max}$ converges pointwise to the cdf of R_0. In particular, we show that for all $c \in [0, \infty)$,

$$\limsup_{n \to \infty} \Pr(n^{1/2}\hat{R}_{\max} > c) \leq \Pr(R_0 > c) \leq \liminf_{n \to \infty} \Pr(n^{1/2}\hat{R}_{\max} > c). \tag{B.15}$$

The first inequality is easy: Since $\hat{R}_{\max} \leq \hat{R}$, it follows by Corollary 3.1 that

$$\limsup_{n \to \infty} \Pr(n^{1/2}\hat{R}_{\max} > c) \leq \limsup_{n \to \infty} \Pr(n^{1/2}\hat{R} > c) = \Pr(R_0 > c). \tag{B.16}$$

The second inequality requires a bit more effort. Let $\hat{\mathbb{Y}} := n^{1/2}(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^r})$ and for $g \in \mathcal{F}_1$ define $B_{\mathbb{P}_{A_0}}(g, \epsilon) := \{h \in \mathcal{F}_1 : \|g - h\|_{2,\mathbb{P}_{A_0}} < \epsilon\}$. By Lemma F.3, given $f \in \mathcal{F}_1$ and $r > 0$ there exists $\hat{f} \in \hat{\mathcal{F}}_1 \cap B_{\mathbb{P}_{A_0}}(f, r)$ w.p.a.1. Consequently, w.p.a.1,

$$\inf_{g \in B_{\mathbb{P}_{A_0}}(f,r)} |\hat{\mathbb{Y}}(g)| \leq |\hat{\mathbb{Y}}(\hat{f})| \leq \max_{h \in \hat{\mathcal{F}}_1} |\hat{\mathbb{Y}}(h)| = n^{1/2}\hat{R}_{\max}.$$

In fact, since the upper bound does not depend upon f, it follows that given $T \in \mathbb{N}$ and $f_1, \ldots, f_T \in \mathcal{F}_1$, the event

$$E_n := \{\max_{t \leq T} \inf_{g \in B_{\mathbb{P}_{A_0}}(f_t,r)} |\hat{\mathbb{Y}}(g)| \leq n^{1/2}\hat{R}_{\max}\}$$

holds w.p.a.1. Now,

$$\Pr(\max_{t \leq T} \inf_{g \in B_{\mathbb{P}_{A_0}}(f_t,r)} |\hat{\mathbb{Y}}(g)| > c, E_n) \leq \Pr(n^{1/2}\hat{R}_{\max} > c, E_n)$$

or equivalently

$$\Pr(\max_{t \leq T} \inf_{g \in B_{\mathbb{P}_{A_0}}(f_t,r)} |\hat{\mathbb{Y}}(g)| > c) - \Pr(\max_{t \leq T} \inf_{g \in B_{\mathbb{P}_{A_0}}(f_t,r)} |\hat{\mathbb{Y}}(g)| > c, E_n^{\complement})$$

$$\leq \Pr(n^{1/2}\hat{R}_{\max} > c) - \Pr(n^{1/2}\hat{R}_{\max} > c, E_n^{\complement}).$$

Hence, since $\lim_{n\to\infty} \Pr(E_n^{\complement}) = 0$,

$$\liminf_{n\to\infty} \Pr(n^{1/2}\hat{\mathrm{R}}_{\max} > c) \geq \liminf_{n\to\infty} \Pr(\max_{t\leq T} \inf_{g\in B_{\mathbb{P}_{A_0}}(f_t,r)} |\hat{\mathbb{Y}}(g)| > c)$$

$$= \Pr(\max_{t\leq T} \inf_{g\in B_{\mathbb{P}_{A_0}}(f_t,r)} |\mathbb{X}_0(g)| > c) \tag{B.17}$$

$$\geq \sup_{T\in\mathbb{N}} \sup_{r>0} \Pr(\max_{t\leq T} \inf_{g\in B_{\mathbb{P}_{A_0}}(f_t,r)} |\mathbb{X}_0(g)| > c), \tag{B.18}$$

where (B.17) follows by Lemma 3.2 applied to $\mathcal{F}_1$ plus the continuous mapping theorem and (B.18) is the tightest lower bound. From the proof of Lemma 3.2 we know that $(\mathcal{F}_1, \|\cdot\|_{2,\mathbb{P}_{A_0}})$ is totally bounded, i.e., given $r > 0$, there exists $T_r < \infty$ such that $\mathcal{F}_1 \subset \cup_{t=1}^{T_r} B_{\mathbb{P}_{A_0}}(f_t,r)$ with $f_1,\ldots,f_{T_r} \in \mathcal{F}_1$. Moreover, from V&W (Addendum 1.5.8) it follows that almost all sample paths of $\{\mathbb{X}_0(f) : f \in \mathcal{F}_1\}$ are uniformly $\|\cdot\|_{2,\mathbb{P}_{A_0}}$-continuous so that for all $\eta > 0$ there exists $r_\eta > 0$ such that the probability of the event

$$U_\eta := \{ \sup_{g,h\in\mathcal{F}_1: \|g-h\|_{2,\mathbb{P}_{A_0}}<2r_\eta} |\mathbb{X}_0(g) - \mathbb{X}_0(h)| < \eta \}$$

approaches one as $\eta \to 0$. Therefore, since being a convex functional of a gaussian process the random variable $\mathrm{R}_0 := \sup_{f\in\mathcal{F}_1} |\mathbb{X}_0(f)|$ is continuously distributed on $(0,\infty)$, cf. Davydov, Lifshits, and Smorodina (1998, Theorem 11.1),

$$\Pr(\mathrm{R}_0 > c) = \lim_{\eta\to 0} \Pr(\sup_{f\in\mathcal{F}_1} |\mathbb{X}_0(f)| > c + \eta)$$

$$\leq \lim_{\eta\to 0} \Pr(\max_{t\leq T_{r_\eta}} \sup_{f\in B_{\mathbb{P}_{A_0}}(f_t,r_\eta)} |\mathbb{X}_0(f)| > c + \eta)$$

$$= \lim_{\eta\to 0} \Pr(\max_{t\leq T_{r_\eta}} \sup_{f\in B_{\mathbb{P}_{A_0}}(f_t,r_\eta)} |\mathbb{X}_0(f)| > c + \eta, U_\eta)$$

$$\leq \lim_{\eta\to 0} \Pr(\sup_{f\in B_{\mathbb{P}_{A_0}}(f_t,r_\eta)} |\mathbb{X}_0(f)| > c + \eta \text{ for some } t \leq T_{r_\eta}, U_\eta)$$

$$\leq \lim_{\eta\to 0} \Pr(\inf_{f\in B_{\mathbb{P}_{A_0}}(f_t,r_\eta)} |\mathbb{X}_0(f)| > c \text{ for some } t \leq T_{r_\eta})$$

$$\leq \lim_{\eta\to 0} \Pr(\max_{t\leq T_{r_\eta}} \inf_{f\in B_{\mathbb{P}_{A_0}}(f_t,r_\eta)} |\mathbb{X}_0(f)| > c)$$

$$\leq \sup_{T\in\mathbb{N}} \sup_{r>0} \Pr(\max_{t\leq T} \inf_{f\in B_{\mathbb{P}_{A_0}}(f_t,r)} |\mathbb{X}_0(f)| > c). \tag{B.19}$$

Hence, (B.15) follows by (B.16), (B.18), and (B.19). $\qquad\square$

APPENDIX C. PROOFS FOR SECTION 4

Proof of Lemma 4.1. For $f \in \mathcal{F}$,

$$
\begin{aligned}
(\hat{\mathbb{P}}_{\hat{Z}^*} - \hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))})f &= n^{-1} \sum_{j=1}^{n} [f(\hat{\varepsilon}_j^*, W_j) - f(R_j \varepsilon_j(\theta_1), W_j)] \\
&= n^{-1} \sum_{j=1}^{n} [f(R_j \varepsilon_j(\theta_1) - \tilde{\Delta}(R_j, X_j, \hat{\theta}^*, \hat{\theta}, \theta_1), W_j) \\
&\qquad\qquad - f(R_j \varepsilon_j(\theta_1) - \tilde{\Delta}(R_j, X_j, \theta_1, \theta_1, \theta_1), W_j)] \\
&= \hat{\mathbb{P}}_{\mathscr{R}(A_1)}(f^{\hat{\theta}^*, \hat{\theta}} - f^{\theta_1, \theta_1}) \\
&= (\hat{\mathbb{P}}_{\mathscr{R}(A_1)} - \mathbb{P}_{\mathscr{R}(A_1)})(f^{\hat{\theta}^*, \hat{\theta}} - f^{\theta_1, \theta_1}) + \mathbb{P}_{\mathscr{R}(A_1)}(f^{\hat{\theta}^*, \hat{\theta}} - f^{\theta_1, \theta_1}).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
(\hat{\mathbb{P}}_{\hat{Z}^*} - \hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))})&f - \mathbb{P}_{\mathscr{R}(A_1)}(f^{\hat{\theta}, \hat{\theta}} - f^{\theta_1, \theta_1}) \\
&= (\hat{\mathbb{P}}_{\mathscr{R}(A_1)} - \mathbb{P}_{\mathscr{R}(A_1)})(f^{\hat{\theta}^*, \hat{\theta}} - f^{\theta_1, \theta_1}) + \mathbb{P}_{\mathscr{R}(A_1)}(f^{\hat{\theta}^*, \hat{\theta}} - f^{\hat{\theta}, \hat{\theta}}).
\end{aligned}
$$

To prove the first result (we only show the first result as the proof of the second one is similar), it thus suffices to show that

$$
\sup_{f \in \mathcal{F}} |n^{1/2}(\hat{\mathbb{P}}_{\mathscr{R}(A_1)} - \mathbb{P}_{\mathscr{R}(A_1)})(f^{\hat{\theta}^*, \hat{\theta}} - f^{\theta_1, \theta_1})| = o_{\mathrm{Pr}^\circ}(1) \tag{C.1}
$$

$$
\sup_{f \in \mathcal{F}} |\mathbb{P}_{\mathscr{R}(A_1)}(f^{\hat{\theta}^*, \hat{\theta}} - f^{\hat{\theta}, \hat{\theta}}) - 0.5\langle f - f^{\mathrm{r}}, (\partial_1 \log p_{A_1})\hat{\mu}_1' \rangle_{\mathbb{P}_{A_1}}(\hat{\theta}^* - \hat{\theta})| = o_{\mathrm{Pr}}(\|\hat{\theta}^* - \hat{\theta}\|). \tag{C.2}
$$

As in the proof of Lemma 3.1, we use equicontinuity of the $n^{1/2}(\hat{\mathbb{P}}_{\mathscr{R}(A_1)} - \mathbb{P}_{\mathscr{R}(A_1)})$ process to demonstrate (C.1) and mean-square differentiability of $p_{A_1}^{1/2}$ to show (C.2). We begin with the first claim. Given $f \in \mathcal{F}$, we know that $f^{\hat{\theta}^*, \hat{\theta}}, f^{\theta_1, \theta_1} \in \mathcal{F}$ and

$$
\|f^{\hat{\theta}^*, \hat{\theta}} - f^{\theta_1, \theta_1}\|_{2, \mathbb{P}_{\mathscr{R}(A_1)}} \leq q(\|(\hat{\theta}^*, \hat{\theta}) - (\theta_1, \theta_1)\|) =: \hat{q}
$$

by Assumption 4.2(iii). Hence, as $|(\hat{\mathbb{P}}_{\mathscr{R}(A_1)} - \mathbb{P}_{\mathscr{R}(A_1)})(f^{\hat{\theta}^*, \hat{\theta}} - f^{\theta_1, \theta_1})|$ is bounded from above by $\sup_{g, h \in \mathcal{F} : \|g-h\|_{2, \mathbb{P}_{\mathscr{R}(A_1)}} \leq \hat{q}} |(\hat{\mathbb{P}}_{\mathscr{R}(A_1)} - \mathbb{P}_{\mathscr{R}(A_1)})(g - h)|$ and the bound does not depend upon f,

$$
\sup_{f \in \mathcal{F}} |n^{1/2}(\hat{\mathbb{P}}_{\mathscr{R}(A_1)} - \mathbb{P}_{\mathscr{R}(A_1)})(f^{\hat{\theta}^*, \hat{\theta}} - f^{\theta_1, \theta_1})| \leq \sup_{g, h \in \mathcal{F} : \|g-h\|_{2, \mathbb{P}_{\mathscr{R}(A_1)}} \leq \hat{q}} |n^{1/2}(\hat{\mathbb{P}}_{\mathscr{R}(A_1)} - \mathbb{P}_{\mathscr{R}(A_1)})(g - h)|.
$$

Then (C.1) follows because the empirical process $\{n^{1/2}(\hat{\mathbb{P}}_{\mathscr{R}(A_1)} - \mathbb{P}_{\mathscr{R}(A_1)})f : f \in \mathcal{F}\}$ is asymptotically equicontinuous (due to the fact that $\mathcal{F}$ is $\mathbb{P}_{\mathscr{R}(A_1)}$-Donsker by Assumption 4.2(iii)) and $\mathrm{plim}(\hat{q}) = 0$ by Assumption 4.1 and the properties of q.

Next, we show (C.2). Begin by observing that

$$\mathbb{P}_{\mathscr{R}(A_1)}f^{\theta_a,\theta_b} = \int f(su - \tilde{\Delta}(s,x,\theta_a,\theta_b,\theta_1),w)\,\mathbb{P}_R(ds)\,\mathbb{P}_{A_1}(du,dx,dw)$$

$$= 0.5\int f(u - \tilde{\Delta}(1,x,\theta_a,\theta_b,\theta_1),w)p_{A_1}(u,x,w)\,du\,\kappa(dx,dw)$$

$$+ 0.5\int f(-u - \tilde{\Delta}(-1,x,\theta_a,\theta_b,\theta_1),w)p_{A_1}(u,x,w)\,du\,\kappa(dx,dw)$$

$$= 0.5\int f(t,w)p_{A_1}(t + \tilde{\Delta}(1,x,\theta_a,\theta_b,\theta_1),x,w)\,dt\,\kappa(dx,dw)$$

$$+ 0.5\int f(t,w)p_{A_1}(-t - \tilde{\Delta}(-1,x,\theta_a,\theta_b,\theta_1),x,w)\,dt\,\kappa(dx,dw)$$

$$= \int f(t,w)p_{A_1}(st + s\tilde{\Delta}(s,x,\theta_a,\theta_b,\theta_1),x,w)\,\mathbb{P}_R(ds)\,dt\,\kappa(dx,dw), \qquad \text{(C.3)}$$

where the third equality follows by the translation invariance of the Lebesgue measure. Hence,

$$\mathbb{P}_{\mathscr{R}(A_1)}(f^{\hat{\theta}^*,\hat{\theta}} - f^{\hat{\theta},\hat{\theta}}) = \int f(t,w)(p_{A_1}(st + s\tilde{\Delta}(s,x,\hat{\theta}^*,\hat{\theta},\theta_1),x,w)$$

$$- p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1),x,w))\,\mathbb{P}_R(ds)\,dt\,\kappa(dx,dw). \quad \text{(C.4)}$$

Under Assumption 4.2(ii), $p_{A_1}^{1/2}$ is mean-square differentiable. Hence, expanding as in (B.7) while keeping $st + \Delta(x,\hat{\theta},\theta_1)$ fixed,

$$p_{A_1}(st + s\tilde{\Delta}(s,x,\hat{\theta}^*,\hat{\theta},\theta_1),x,w) - p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1),x,w)$$

$$= p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1) + s\Delta(x,\hat{\theta}^*,\hat{\theta}),x,w) - p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1),x,w)$$

$$= (\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1),x,w))s\Delta(x,\hat{\theta}^*,\hat{\theta})$$

$$+ 0.25(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1),x,w))^2 p_{A_1}^{-1}(st + \Delta(x,\hat{\theta},\theta_1),x,w)(s\Delta(x,\hat{\theta}^*,\hat{\theta}))^2$$

$$+ r^2(s\Delta(x,\hat{\theta}^*,\hat{\theta}),st + \Delta(x,\hat{\theta},\theta_1),x,w)p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1),x,w)$$

$$+ 2r(s\Delta(x,\hat{\theta}^*,\hat{\theta}),st + \Delta(x,\hat{\theta},\theta_1),x,w)p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1),x,w)$$

$$+ r(s\Delta(x,\hat{\theta}^*,\hat{\theta}),st + \Delta(x,\hat{\theta},\theta_1),x,w)(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1),x,w))s\Delta(x,\hat{\theta}^*,\hat{\theta}).$$

We use this decomposition to handle (C.4). First, by Assumption 4.2(i),

$$\int f(t,w)(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1),x,w))s\Delta(x,\hat{\theta}^*,\hat{\theta})\,\mathbb{P}_R(ds)\,dt\,\kappa(dx,dw)$$

$$= \int f(t,w)(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1),x,w))s\dot{\mu}'(x,\hat{\theta})(\hat{\theta}^* - \hat{\theta})\,\mathbb{P}_R(ds)\,dt\,\kappa(dx,dw)$$

$$+ \int f(t,w)(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1),x,w))s\rho(x,\hat{\theta}^*,\hat{\theta})\,\mathbb{P}_R(ds)\,dt\,\kappa(dx,dw).$$

By a change of variable,

$$\int f(t,w)(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1), x, w)) s\rho(x,\hat{\theta}^*,\hat{\theta}) \, \mathbb{P}_R(ds) \, dt \, \kappa(dx,dw)$$

$$= 0.5 \int f(t,w)(\partial_1 p_{A_1}(t + \Delta(x,\hat{\theta},\theta_1), x, w)) \rho(x,\hat{\theta}^*,\hat{\theta}) \, dt \, \kappa(dx,dw)$$

$$- 0.5 \int f(t,w)(\partial_1 p_{A_1}(-t + \Delta(x,\hat{\theta},\theta_1), x, w)) \rho(x,\hat{\theta}^*,\hat{\theta}) \, dt \, \kappa(dx,dw)$$

$$= 0.5 \int f(u - \Delta(x,\hat{\theta},\theta_1), w)(\partial_1 p_{A_1}(u, x, w)) \rho(x,\hat{\theta}^*,\hat{\theta}) \, du \, \kappa(dx,dw)$$

$$- 0.5 \int f(-(u - \Delta(x,\hat{\theta},\theta_1)), w)(\partial_1 p_{A_1}(u, x, w)) \rho(x,\hat{\theta}^*,\hat{\theta}) \, du \, \kappa(dx,dw)$$

$$= 0.5 \int (f - f^{\mathrm{r}})(u - \Delta(x,\hat{\theta},\theta_1), w)(\partial_1 \log p_{A_1}(u, x, w)) \rho(x,\hat{\theta}^*,\hat{\theta}) \, \mathbb{P}_{A_1}(du, dx, dw). \quad \text{(C.5)}$$

Hence, following the argument leading to (B.8),

$$| \int f(t,w)(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1), x, w)) s\rho(x,\hat{\theta}^*,\hat{\theta}) \, \mathbb{P}_R(ds) \, dt \, \kappa(dx,dw) |^2$$

$$\leq M_{\mathcal{F}}^2 \|v_{A_1}\|_\infty \int \sup_{\theta \in \overline{B(\hat{\theta}, \|\hat{\theta}^* - \hat{\theta}\|)}} \rho^2(x,\theta,\hat{\theta}) \, \mathbb{P}_X(dx)$$

$$= o_{\mathrm{Pr}}(\|\hat{\theta}^* - \hat{\theta}\|^2) \quad \text{(C.6)}$$

by Assumption 4.2(i). Arguing similarly,

$$| \int f(t,w)(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1), x, w)) s(\dot{\mu}(x,\hat{\theta}) - \dot{\mu}'(x,\theta_1))'(\hat{\theta}^* - \hat{\theta})) \, \mathbb{P}_R(ds) \, dt \, \kappa(dx,dw) |$$

$$\leq M_{\mathcal{F}} \|v_{A_1}\|_\infty \|\dot{\mu}(\cdot,\hat{\theta}) - \dot{\mu}(\cdot,\theta_1)\|_{2,\mathbb{P}_X} \|\hat{\theta}^* - \hat{\theta}\|$$

$$= o_{\mathrm{Pr}}(\|\hat{\theta}^* - \hat{\theta}\|) \quad \text{(C.7)}$$

by Assumption 4.2(i). Therefore, by (C.6) and (C.7),

$$| \int f(t,w)(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1), x, w)) s\Delta(x,\hat{\theta}^*,\hat{\theta}) \, \mathbb{P}_R(ds) \, dt \, \kappa(dx,dw)$$

$$- \int f(t,w)(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1), x, w)) s\dot{\mu}'(x,\theta_1)(\hat{\theta}^* - \hat{\theta}) \, \mathbb{P}_R(ds) \, dt \, \kappa(dx,dw) |$$

$$= o_{\mathrm{Pr}}(\|\hat{\theta}^* - \hat{\theta}\|). \quad \text{(C.8)}$$

Now, the argument leading to (C.5) shows that

$$\int f(t,w)(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1), x, w)) s\dot{\mu}'(x,\theta_1)(\hat{\theta}^* - \hat{\theta}) \, \mathbb{P}_R(ds) \, dt \, \kappa(dx,dw)$$

$$= 0.5 \int (f - f^{\mathrm{r}})(u - \Delta(x,\hat{\theta},\theta_1), w)(\partial_1 \log p_{A_1}(u, x, w)) \dot{\mu}'(x,\theta_1)(\hat{\theta}^* - \hat{\theta}) \, \mathbb{P}_{A_1}(du, dx, dw).$$

Thus, by Assumption 4.2(iii) and Cauchy-Schwarz,

$$
\begin{aligned}
\Big| \int f(t,w)(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1)&, x, w) \\
- \partial_1 p_{A_1}(st, x, w))s\dot{\mu}'(x,\theta_1)(\hat{\theta}^* - \hat{\theta})\, \mathbb{P}_R(ds)\, dt\, \kappa(dx, dw) \Big| & \\
\leq q(\|\hat{\theta} - \theta_1\|)\, \|v_{A_1}\|_\infty\, \|(\dot{\mu}_1'\dot{\mu}_1)^{1/2}\|_{2,\mathbb{P}_X}\, \|\hat{\theta}^* - \hat{\theta}\| &= o_{\Pr}(\|\hat{\theta}^* - \hat{\theta}\|) \qquad \text{(C.9)}
\end{aligned}
$$

because $\mathrm{plim}(\hat{\theta}) = \theta_1$, q is continuous, and $q(0) = 0$. Therefore, by (C.8) and (C.9),

$$
\begin{aligned}
\Big| \int f(t,w)(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1), x, w))&s\Delta(x,\hat{\theta}^*,\hat{\theta})\, \mathbb{P}_R(ds)\, dt\, \kappa(dx, dw) \\
- \int f(t,w)(\partial_1 p_{A_1}(st, x, w))s\dot{\mu}'(x,\theta_1)(\hat{\theta}^* - \hat{\theta})\, \mathbb{P}_R(ds)\, dt\, \kappa(dx, dw) \Big| &= o_{\Pr}(\|\hat{\theta}^* - \hat{\theta}\|).
\end{aligned}
$$

But as in (C.5),

$$
\int f(t,w)(\partial_1 p_{A_1}(st, x, w))s\dot{\mu}(x,\theta_1)\, \mathbb{P}_R(ds)\, dt\, \kappa(dx, dw) = 0.5\langle f - f^{\mathrm{r}}, (\partial_1 \log p_{A_1})\dot{\mu}_1\rangle_{\mathbb{P}_{A_1}}.
$$

Therefore,

$$
\begin{aligned}
\sup_{f \in \mathcal{F}} \Big| \int f(t,w)(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1), x, w))&s\Delta(x,\hat{\theta}^*,\hat{\theta})\, \mathbb{P}_R(ds)\, dt\, \kappa(dx, dw) \\
- 0.5\langle f - f^{\mathrm{r}}, (\partial_1 \log p_{A_1}(u, x, w)\dot{\mu}_1\rangle_{\mathbb{P}_{A_1}} \Big| &= o_{\Pr}(\|\hat{\theta}^* - \hat{\theta}\|). \qquad \text{(C.10)}
\end{aligned}
$$

Next, since $R^2 = 1$, the argument leading to (C.5) reveals that

$$
\begin{aligned}
\int f(t,w)(\partial_1 p_{A_1}(st + \Delta(x,\hat{\theta},\theta_1), x, w))^2 & p_{A_1}^{-1}(st + \Delta(x,\hat{\theta},\theta_1), x, w) \\
\times (s\Delta(x,\hat{\theta}^*,\hat{\theta}))^2\, \mathbb{P}_R(ds)\, dt\, \kappa(dx, dw) & \\
= 0.5 \int (f + f^{\mathrm{r}})(u - \Delta(x,\hat{\theta},\theta_1), w)&(\partial_1 \log p_{A_1}(u, x, w))^2 \\
\times \Delta^2(x,\hat{\theta}^*,\hat{\theta})\, \mathbb{P}_{A_1}(du, dx, dw). &
\end{aligned}
$$

Hence, by Assumption 4.2(ii),

$$\sup_{f\in\mathcal{F}}|\int f(t,w)(\partial_1 p_{A_1}(st+\Delta(x,\hat{\theta},\theta_1),x,w))^2 p_{A_1}^{-1}(st+\Delta(x,\hat{\theta},\theta_1),x,w)$$

$$\times (s\Delta(x,\hat{\theta}^*,\hat{\theta}))^2\,\mathbb{P}_R(ds)\,dt\,\kappa(dx,dw)|$$

$$\leq M_{\mathcal{F}}\,\|v_{A_1}\|_\infty \int (\|\dot{\mu}(x,\hat{\theta})\|^2\,\|\hat{\theta}^*-\hat{\theta}\|^2 + \sup_{\theta\in\overline{B(\hat{\theta},\|\hat{\theta}^*-\hat{\theta}\|)}}\rho^2(x,\theta,\hat{\theta}))\,\mathbb{P}_X(dx)$$

$$\leq M_{\mathcal{F}}\,\|v_{A_1}\|_\infty((\|\dot{\mu}(\cdot,\hat{\theta})-\dot{\mu}(\cdot,\theta_1)\|_{2,\mathbb{P}_X}^2 + \|\dot{\mu}(\cdot,\theta_1)\|_{2,\mathbb{P}_X}^2)\,\|\hat{\theta}^*-\hat{\theta}\|^2$$

$$+\int \sup_{\theta\in\overline{B(\hat{\theta},\|\hat{\theta}^*-\hat{\theta}\|)}}\rho^2(x,\theta,\hat{\theta})\,\mathbb{P}_{A_1}(dx))$$

$$= O_{\mathrm{Pr}}(\|\hat{\theta}^*-\hat{\theta}\|^2). \tag{C.11}$$

Next, since

$$\int f(t,w)r^2(s\Delta(x,\hat{\theta}^*,\hat{\theta}),st+\Delta(x,\hat{\theta},\theta_1),x,w)$$

$$\times p_{A_1}(st+\Delta(x,\hat{\theta},\theta_1),x,w)\,\mathbb{P}_R(ds)\,dt\,\kappa(dx,dw)$$

$$= 0.5\int (f(u-\Delta(x,\hat{\theta},\theta_1),w)r^2(\Delta(x,\hat{\theta}^*,\hat{\theta}),u,x,w)$$

$$+ f^{\mathrm{r}}(u-\Delta(x,\hat{\theta},\theta_1),w)r^2(-\Delta(x,\hat{\theta}^*,\hat{\theta}),u,x,w))\,\mathbb{P}_{A_1}(du,dx,dw),$$

following the argument leading to (C.11) we obtain that

$$\sup_{f\in\mathcal{F}}|\int f(t,w)r^2(s\Delta(x,\hat{\theta}^*,\hat{\theta}),st+\Delta(x,\hat{\theta},\theta_1),x,w)$$

$$\times p_{A_1}(st+\Delta(x,\hat{\theta},\theta_1),x,w)\,\mathbb{P}_R(ds)\,dt\,\kappa(dx,dw)|$$

$$\leq M_{\mathcal{F}}\,o(\|\Delta(\cdot,\hat{\theta}^*,\hat{\theta})\|_{2,\mathbb{P}_X}^2) = o(\|\hat{\theta}^*-\hat{\theta}\|^2). \tag{C.12}$$

In a similar manner, it can be shown that

$$\sup_{f\in\mathcal{F}}|\int f(t,w)r(s\Delta(x,\hat{\theta}^*,\hat{\theta}),st+\Delta(x,\hat{\theta},\theta_1),x,w)$$

$$\times p_{A_1}(st+\Delta(x,\hat{\theta},\theta_1),x,w)\,\mathbb{P}_R(ds)\,dt\,\kappa(dx,dw)| = o(\|\hat{\theta}^*-\hat{\theta}\|^2) \tag{C.13}$$

and

$$\sup_{f\in\mathcal{F}}|\int f(t,w)r(s\Delta(x,\hat{\theta}^*,\hat{\theta}),st+\Delta(x,\hat{\theta},\theta_1),x,w)(\partial_1 p_{A_1}(st+\Delta(x,\hat{\theta},\theta_1),x,w))$$

$$\times s\Delta(x,\hat{\theta}^*,\hat{\theta})\,\mathbb{P}_R(ds)\,dt\,\kappa(dx,dw)| = o(\|\hat{\theta}^*-\hat{\theta}\|^2). \tag{C.14}$$

The desired result follows by (C.10)–(C.14). □

Proof of (4.2). By (C.3) and the fact that $R^2 = 1$,

$$\mathbb{P}_{\mathscr{R}(A_1)} f^{\hat{\theta}, \hat{\theta}} = \int f(t, w) p_{A_1}(st + \Delta(x, \hat{\theta}, \theta_1), x, w) \, \mathbb{P}_R(ds) \, dt \, \kappa(dx, dw)$$

$$\mathbb{P}_{\mathscr{R}(A_1)} f^{\theta_1, \theta_1} = \int f(t, w) p_{A_1}(st, x, w) \, \mathbb{P}_R(ds) \, dt \, \kappa(dx, dw).$$

Similarly, by a change of variable and the fact that $\mathrm{supp}(R) = \{-1, 1\}$,

$$\mathbb{P}_{\mathscr{R}(A_1)} f^{\mathrm{r}\hat{\theta}, \hat{\theta}} = \int f^{\mathrm{r}}(t, w) p_{A_1}(st + \Delta(x, \hat{\theta}, \theta_1), x, w) \, \mathbb{P}_R(ds) \, dt \, \kappa(dx, dw)$$

$$= \int f(-t, w) p_{A_1}(st + \Delta(x, \hat{\theta}, \theta_1), x, w) \, \mathbb{P}_R(ds) \, dt \, \kappa(dx, dw)$$

$$= \int f(u, w) p_{A_1}(-su + \Delta(x, \hat{\theta}, \theta_1), x, w) \, \mathbb{P}_R(ds) \, du \, \kappa(dx, dw)$$

$$= \int f(u, w) p_{A_1}(ru + \Delta(x, \hat{\theta}, \theta_1), x, w) \, \mathbb{P}_R(dr) \, du \, \kappa(dx, dw)$$

$$= \mathbb{P}_{\mathscr{R}(A_1)} f^{\hat{\theta}, \hat{\theta}}.$$

The desired result follows because $\mathbb{P}_{\mathscr{R}(A_1)} f^{\mathrm{r}\theta_1, \theta_1} = \mathbb{P}_{\mathscr{R}(A_1)} f^{\theta_1, \theta_1}$ by setting Δ to be identically zero in the preceding display. $\qquad\square$

Proof of Lemma 4.2. Since $\mathcal{F}$ is $\mathbb{P}_{\mathscr{R}(A_1)}$-Donsker by Assumption 4.2(iii), tightness of $\hat{\mathbb{X}}_1^*$ can be established as in the proof of Lemma 3.2. It only remains to prove that the empirical process $\{n^{1/2}(\hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))} - \mathbb{P}_{\mathscr{R}(Z(\theta_1))})(f - f^{\mathrm{r}}) : f \in \mathcal{F}\}$ has the same limiting marginal distribution and covariance function as $\mathbb{G}_1$. So let $f \in \mathcal{F}$. Then, since $(f - f^{\mathrm{r}})(Z^{\mathrm{r}}(\theta_1)) = -(f - f^{\mathrm{r}})(Z(\theta_1))$ and $\mathbb{1}_{\{1\}}(R) - \mathbb{1}_{\{-1\}}(R) = R$,

$$\hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))}(f - f^{\mathrm{r}}) = n^{-1} \sum_{j=1}^{n} (f - f^{\mathrm{r}})(\mathscr{R}(Z_j(\theta_1)))$$

$$= n^{-1} \sum_{j=1}^{n} (f - f^{\mathrm{r}})(Z_j(\theta_1)) \mathbb{1}_{\{1\}}(R_j) + (f - f^{\mathrm{r}})(Z_j^{\mathrm{r}}(\theta_1)) \mathbb{1}_{\{-1\}}(R_j)$$

$$= n^{-1} \sum_{j=1}^{n} (\mathbb{1}_{\{1\}}(R_j) - \mathbb{1}_{\{-1\}}(R_j))(f - f^{\mathrm{r}})(Z_j(\theta_1))$$

$$= n^{-1} \sum_{j=1}^{n} R_j (f - f^{\mathrm{r}})(Z_j(\theta_1)).$$

Hence, as $\mathbb{P}_{\mathscr{R}(Z(\theta_1))}(f - f^{\mathrm{r}}) = 0$ (cf. the remark at the end of Section 4), the CLT for iid random variables shows that the empirical process $\{n^{1/2}(\hat{\mathbb{P}}_{\mathscr{R}(Z(\theta_1))} - \mathbb{P}_{\mathscr{R}(Z(\theta_1))})(f - f^{\mathrm{r}}) : f \in \mathcal{F}\}$ has the same limiting marginal distribution and covariance function as $\mathbb{G}_1$. $\qquad\square$

Proof of Lemma 4.3. From Corollary 4.1 we know that $n^{1/2}\hat{R}^*$ converges in distribution to R_1 whether the null is true or not. Therefore, it suffices to prove the same for $n^{1/2}\hat{R}^*_{\max}$. In particular, we show that for all $c \in [0, \infty)$,

$$\limsup_{n\to\infty} \Pr(n^{1/2}\hat{R}^*_{\max} > c) \le \Pr(R_1 > c) \le \liminf_{n\to\infty} \Pr(n^{1/2}\hat{R}^*_{\max} > c).$$

This follows exactly as in the proof of Lemma 3.3 provided $\hat{\mathcal{F}}^*_1 := \{\mathbb{1}_{(-\infty, z]} : z \in \hat{\mathcal{Z}}^*\}$ is dense in $\mathcal{F}_1$ in the sense that for each $f \in \mathcal{F}_1$ and $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr(\exists \hat{f} \in \hat{\mathcal{F}}^*_1 : \|\hat{f} - f\|_{2, \mathbb{P}_{\mathcal{R}(A_1)}} \vee \|\hat{f}^{\mathrm{r}} - f^{\mathrm{r}}\|_{2, \mathbb{P}_{\mathcal{R}(A_1)}} < \epsilon) = 1.$$

But the denseness result above follows as in the proof of Lemma F.3 upon replacing ε by $\varepsilon^* := R\varepsilon(\theta_1)$, Z by $Z^* := (\varepsilon^*, W)$, $\hat{Z}$ by $\hat{Z}^* := (\hat{\varepsilon}^*, W)$ with $\hat{\varepsilon}^*$ as defined in Section 4, $\Delta(X, \hat{\theta}, \theta_0)$ by $\tilde{\Delta}(R, X, \hat{\theta}^*, \hat{\theta}, \theta_1)$ (because $\hat{\varepsilon}^* - \varepsilon^* = -\tilde{\Delta}(R, X, \hat{\theta}^*, \hat{\theta}, \theta_1)$), and using the fact that $\mathrm{plim}(\hat{\theta}^*, \hat{\theta}) = (\theta_1, \theta_1)$ by Assumption 4.1. $\qquad\square$

APPENDIX D. PROOFS FOR SECTION 5

Proof of Theorem 5.1. Under the alternative, $\mathbb{P}_{Z(\theta)} \ne \mathbb{P}_{Z^{\mathrm{r}}(\theta)}$ for each $\theta \in \Theta$; in particular, $\mathbb{P}_{Z(\theta_1)} \ne \mathbb{P}_{Z^{\mathrm{r}}(\theta_1)}$. Now write

$$\sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f| = \sup_{f \in \mathcal{F}} |(\mathbb{P}_{Z(\theta_1)} - \mathbb{P}_{Z^{\mathrm{r}}(\theta_1)})f| + \hat{\Delta}_1(\mathcal{F}) + \hat{\Delta}_2(\mathcal{F}),$$

where $\hat{\Delta}_1(\mathcal{F}) := \sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f| - \sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{Z(\theta_1)} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}(\theta_1)})f|$ and

$$\hat{\Delta}_2(\mathcal{F}) := \sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{Z(\theta_1)} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}(\theta_1)})f| - \sup_{f \in \mathcal{F}} |(\mathbb{P}_{Z(\theta_1)} - \mathbb{P}_{Z^{\mathrm{r}}(\theta_1)})f|.$$

By the reverse triangle inequality and Lemma F.1,

$$\begin{aligned} |\hat{\Delta}_1(\mathcal{F})| &\le \sup_{f \in \mathcal{F}} ||(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f| - |(\hat{\mathbb{P}}_{Z(\theta_1)} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}(\theta_1)})f|| \\ &\le \sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f - (\hat{\mathbb{P}}_{Z(\theta_1)} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}(\theta_1)})f| \\ &\le \sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{Z(\theta_1)})f| + \sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}(\theta_1)})f| \\ &= o_{\Pr^\circ}(1). \end{aligned} \tag{D.1}$$

Next, since $\mathcal{F}$ is $\mathbb{P}_{A(\theta_1)}$-Donsker $\implies$ $\mathcal{F}$ and $\mathcal{F}^{\mathrm{r}}$ are $\mathbb{P}_{A(\theta_1)}$-Glivenko-Cantelli $\implies$ $\mathcal{F}$ and $\mathcal{F}^{\mathrm{r}}$ are $\mathbb{P}_{Z(\theta_1)}$-Glivenko-Cantelli,

$$
\begin{aligned}
|\hat{\Delta}_2(\mathcal{F})| &\leq \sup_{f \in \mathcal{F}} ||(\hat{\mathbb{P}}_{Z(\theta_1)} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}(\theta_1)})f| - |(\mathbb{P}_{Z(\theta_1)} - \mathbb{P}_{Z^{\mathrm{r}}(\theta_1)})f|| \\
&\leq \sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{Z(\theta_1)} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}(\theta_1)})f - (\mathbb{P}_{Z(\theta_1)} - \mathbb{P}_{Z^{\mathrm{r}}(\theta_1)})f| \\
&\leq \sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{Z(\theta_1)} - \mathbb{P}_{Z(\theta_1)})f| + \sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{Z(\theta_1)} - \mathbb{P}_{Z(\theta_1)})f^{\mathrm{r}}| \\
&= \sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{Z(\theta_1)} - \mathbb{P}_{Z(\theta_1)})f| + \sup_{f \in \mathcal{F}^{\mathrm{r}}} |(\hat{\mathbb{P}}_{Z(\theta_1)} - \mathbb{P}_{Z(\theta_1)})f| \\
&= o_{\mathrm{Pr}^{\circ}}(1). \tag{D.2}
\end{aligned}
$$

Hence,

$$
\widehat{\mathrm{KS}}_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f| = \sup_{f \in \mathcal{F}} |(\mathbb{P}_{Z(\theta_1)} - \mathbb{P}_{Z^{\mathrm{r}}(\theta_1)})f| + o_{\mathrm{Pr}^{\circ}}(1). \tag{D.3}
$$

Since $\mathcal{F}$ is measure determining, $\sup_{f \in \mathcal{F}} |(\mathbb{P}_{Z(\theta_1)} - \mathbb{P}_{Z^{\mathrm{r}}(\theta_1)})f| = 0 \implies \mathbb{P}_{Z(\theta_1)} = \mathbb{P}_{Z^{\mathrm{r}}(\theta_1)}$. Therefore, under the alternative hypothesis, $\sup_{f \in \mathcal{F}} |(\mathbb{P}_{Z(\theta_1)} - \mathbb{P}_{Z^{\mathrm{r}}(\theta_1)})f| > 0$. It follows by (D.3) that, under the alternative, $n^{1/2} \widehat{\mathrm{KS}}_{\mathcal{F}}$ is unbounded w.p.a.1. $\qquad \square$

Proof of Theorem 5.2. Let $\Psi(f) := |(\mathbb{P}_{Z(\theta_1)} - \mathbb{P}_{Z^{\mathrm{r}}(\theta_1)})f|$ for notational convenience. Then, with $\hat{\Delta}_1$ and $\hat{\Delta}_2$ as defined in the proof of Theorem 5.1,

$$
\hat{\mathrm{R}}_{\max} := \sup_{f \in \hat{\mathcal{F}}_1} |(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}})f| = \sup_{f \in \hat{\mathcal{F}}_1} \Psi(f) + \hat{\Delta}_1(\hat{\mathcal{F}}_1) + \hat{\Delta}_2(\hat{\mathcal{F}}_1).
$$

The argument leading to (D.1) shows that

$$
\begin{aligned}
|\hat{\Delta}_1(\hat{\mathcal{F}}_1)| &\leq \sup_{f \in \hat{\mathcal{F}}_1} |(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{Z(\theta_1)})f| + \sup_{f \in \hat{\mathcal{F}}_1} |(\hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}(\theta_1)})f| \\
&\leq \sup_{f \in \mathcal{F}_1} |(\hat{\mathbb{P}}_{\hat{Z}} - \hat{\mathbb{P}}_{Z(\theta_1)})f| + \sup_{f \in \mathcal{F}_1} |(\hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}} - \hat{\mathbb{P}}_{Z^{\mathrm{r}}(\theta_1)})f| \qquad (\hat{\mathcal{F}}_1 \subset \mathcal{F}_1) \\
&= o_{\mathrm{Pr}^{\circ}}(1),
\end{aligned}
$$

Similarly, the argument leading to (D.2) shows that

$$
|\hat{\Delta}_2(\hat{\mathcal{F}}_1)| \leq \sup_{f \in \mathcal{F}_1} |(\hat{\mathbb{P}}_{Z(\theta_1)} - \mathbb{P}_{Z(\theta_1)})f| + \sup_{f \in \mathcal{F}_1^{\mathrm{r}}} |(\hat{\mathbb{P}}_{Z(\theta_1)} - \mathbb{P}_{Z(\theta_1)})f| = o_{\mathrm{Pr}^{\circ}}(1).
$$

Therefore, letting $\hat{\Delta}_3 := \sup_{f \in \hat{\mathcal{F}}_1} \Psi(f) - \sup_{f \in \mathcal{F}_1} \Psi(f)$,

$$
\hat{\mathrm{R}}_{\max} = \sup_{f \in \hat{\mathcal{F}}_1} \Psi(f) + o_{\mathrm{Pr}^{\circ}}(1) = \sup_{f \in \mathcal{F}_1} \Psi(f) + \hat{\Delta}_3 + o_{\mathrm{Pr}^{\circ}}(1).
$$

Since $\sup_{f \in \mathcal{F}} \Psi(f) > 0$ under the alternative hypothesis, $n^{1/2} \hat{\mathrm{R}}_{\max}$ will be unbounded under the alternative (hence consistent) provided $\hat{\Delta}_3 = o_{\mathrm{Pr}}(1)$. To show the latter, let $\epsilon > 0$ and observe

that $\hat{\Delta}_3 \leq \epsilon$ is always true because $\hat{\mathcal{F}}_1 \subset \mathcal{F}_1$. Next, by definition of the supremum, there exists $g_\epsilon \in \mathcal{F}_1$ such that $\sup_{f \in \mathcal{F}_1} \Psi(f) - \epsilon < \Psi(g_\epsilon)$. For $\eta > 0$, consider the event

$$\hat{E}_\eta := \{\exists h_\eta \in \hat{\mathcal{F}}_1 : \|h_\eta - g_\epsilon\|_{2,\mathbb{P}_{A(\theta_1)}} \vee \|h_\eta^{\mathrm{r}} - g_\epsilon^{\mathrm{r}}\|_{2,\mathbb{P}_{A(\theta_1)}} < \eta\}.$$

If $\hat{E}_\eta$ is true, then there exists $h_\eta \in \hat{\mathcal{F}}_1$ such that $\|h_\eta - g_\epsilon\|_{2,\mathbb{P}_{A(\theta_1)}} + \|h_\eta^{\mathrm{r}} - g_\epsilon^{\mathrm{r}}\|_{2,\mathbb{P}_{A(\theta_1)}} < 2\eta$. But,

$$\sup_{f \in \mathcal{F}_1} \Psi(f) - \epsilon < \Psi(g_\epsilon) \leq |\Psi(g_\epsilon) - \Psi(h_\eta)| + \Psi(h_\eta) \leq |\Psi(g_\epsilon) - \Psi(h_\eta)| + \sup_{f \in \hat{\mathcal{F}}_1} \Psi(f)$$

and

$$|\Psi(g_\epsilon) - \Psi(h_\eta)| \leq |(\mathbb{P}_{Z(\theta_1)} - \mathbb{P}_{Z^{\mathrm{r}}(\theta_1)})(g_\epsilon - h_\eta)| \qquad \text{(reverse triangle inequality)}$$
$$\leq |\mathbb{P}_{Z(\theta_1)}(g_\epsilon - h_\eta)| + |\mathbb{P}_{Z(\theta_1)}(g_\epsilon^{\mathrm{r}} - h_\eta^{\mathrm{r}})|$$
$$\leq \|g_\epsilon - h_\eta\|_{2,\mathbb{P}_{A(\theta_1)}} + \|g_\epsilon^{\mathrm{r}} - h_\eta^{\mathrm{r}}\|_{2,\mathbb{P}_{A(\theta_1)}} \qquad \text{(Jensen)}.$$

Hence, choosing $\eta \leq \epsilon$,

$$\hat{E}_\eta \implies \sup_{f \in \hat{\mathcal{F}}_1} \Psi(f) - \epsilon < 2\eta + \sup_{f \in \hat{\mathcal{F}}_1} \Psi(f) \iff \hat{\Delta}_3 \geq -\epsilon - 2\eta \geq -3\epsilon \implies |\hat{\Delta}_3| \leq 3\epsilon.$$

As we proved (F.1), we can show that $\lim_{n \to \infty} \Pr(\hat{E}_\eta) = 1$ under Assumptions 3.1, 3.2(ii), and 3.4 with (θ_0, ε) replaced by $(\theta_1, \varepsilon(\theta_1))$. [Note that Lemma F.3 holds irrespective of whether the null hypothesis is true or not.] It follows that $\lim_{n \to \infty} \Pr(|\hat{\Delta}_3| \leq 3\epsilon) = 1$ which is equivalent to $\hat{\Delta}_3 = o_{\Pr}(1)$ because ϵ was arbitrary. $\qquad \square$

Proof of Lemma 5.2. The approach (and notation) here is very similar to the proof of Lemma 3.2, the main difference being that we now use empirical process results that allow the underlying data generating measures to depend upon n. Recall that $\hat{\mathbb{X}}_{\theta_n}(f) := n^{1/2}(\hat{\mathbb{P}}_{Z(\theta_n)} - \mathbb{P}_{Z(\theta_n)})(f - f^{\mathrm{r}}) + \langle f - f^{\mathrm{r}}, m'_{A_n}\rangle_{\mathbb{P}_{A_n}} n^{1/2}(\hat{\theta} - \theta_n)$, where $m_{A_n} := (\partial_1 \log p_{A_n})\dot{\mu}_0$ and $f \in \mathcal{F}$. To prove that $(\hat{\mathbb{X}}_{\theta_n} + \Gamma_h)\mathcal{F}$ is asymptotically tight, it suffices to show (V&W, Section 2.8.3) that $\lim_{n \to \infty} \sup_{f,g \in \mathcal{F}} |\|f - g\|_{2,\mathbb{P}_{A_n}} - \|f - g\|_{2,\mathbb{P}_{A_0}}| = 0$ and, for all $\epsilon > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \Pr^{\circ}(\sup_{f,g \in \mathcal{F}_{\delta,\mathbb{P}_{A_n}}} |(\hat{\mathbb{X}}_{\theta_n} + \Gamma_h)(f - g)| > \epsilon) = 0. \qquad (D.4)$$

The first condition follows immediately because for all $f, g \in \mathcal{F}$,

$$(\|f - g\|_{2,\mathbb{P}_{A_n}} - \|f - g\|_{2,\mathbb{P}_{A_0}})^2 \leq |\|f - g\|_{2,\mathbb{P}_{A_n}} - \|f - g\|_{2,\mathbb{P}_{A_0}}| \times |\|f - g\|_{2,\mathbb{P}_{A_n}} + \|f - g\|_{2,\mathbb{P}_{A_0}}|$$
$$= |\|f - g\|_{2,\mathbb{P}_{A_n}}^2 - \|f - g\|_{2,\mathbb{P}_{A_0}}^2|$$
$$= n^{-1/2} \int (f - g)^2 h \, d\mathbb{P}_{A_0}$$
$$\leq 2M_{\mathcal{F}}^2 \|h\|_\infty n^{-1/2}.$$

To show (D.4), begin by observing that

$$\Pr^\circ(\sup_{f,g\in\mathcal{F}_{\delta,\mathbb{P}_{A_n}}} |(\hat{\mathbb{X}}_{\theta_n} + \Gamma_h)(f-g)| > 3\epsilon) \le t_{\delta,n}(1) + t_{\delta,n}(2) + t_{\delta,n}(3),$$

where $t_{\delta,n}(1) := \Pr^\circ(\sup_{f,g\in\mathcal{F}_{\delta,\mathbb{P}_{A_n}}} |n^{1/2}(\hat{\mathbb{P}}_{Z(\theta_n)} - \mathbb{P}_{Z(\theta_n)})((f-f^{\mathrm{r}}) - (g-g^{\mathrm{r}}))| > \epsilon),$

$$t_{\delta,n}(2) := \Pr^\circ(\sup_{f,g\in\mathcal{F}_{\delta,\mathbb{P}_{A_n}}} |\langle (f-g)-(f^{\mathrm{r}}-g^{\mathrm{r}}), m'_{A_n}\rangle_{\mathbb{P}_{A_n}} n^{1/2}(\hat{\theta} - \theta_n)| > \epsilon),$$

and $t_{\delta,n}(3) := \mathbb{1}(\sup_{f,g\in\mathcal{F}_{\delta,\mathbb{P}_{A_n}}} |\Gamma_h(f-g)| > \epsilon)$ because $\Gamma_h(f-g)$ is nonstochastic. Now,

$$\lim_{\delta\to 0}\limsup_{n\to\infty} t_{\delta,n}(1) \le \lim_{\delta\to 0}\limsup_{n\to\infty}\Pr^\circ(\sup_{f,g\in(\mathcal{F}-\mathcal{F}^{\mathrm{r}})_{\delta,\mathbb{P}_{A_n}}} |n^{1/2}(\hat{\mathbb{P}}_{Z(\theta_n)} - \mathbb{P}_{Z(\theta_n)})(f-g)| > \epsilon) = 0$$

because $\mathcal{F} - \mathcal{F}^{\mathrm{r}}$ is asymptotically equicontinuous uniformly in $(\mathbb{P}_{A_n})$ which follows because $\mathcal{F} - \mathcal{F}^{\mathrm{r}}$ is Donsker and pre-gaussian uniformly in $(\mathbb{P}_{A_n})$ by Assumptions 5.3(ii) and (iii); cf. V&W (Section 2.8.3). Next, for the remainder of the proof, let $f, g \in \mathcal{F}_{\delta,\mathbb{P}_{A_n}}$. By Assumption 5.2(ii),

$$\|\langle (f-g)-(f^{\mathrm{r}}-g^{\mathrm{r}}), m_{A_n}\rangle_{\mathbb{P}_{A_n}}\| \le (\|f-g\|_{2,\mathbb{P}_{A_n}} + \|f^{\mathrm{r}}-g^{\mathrm{r}}\|_{2,\mathbb{P}_{A_n}}) \|v_{A_n}\|_\infty \|(\dot{\mu}'_0\dot{\mu}_0)^{1/2}\|_{2,\mathbb{P}_X}.$$

Moreover, recalling that $\mathbb{P}_Z = \mathbb{P}_{Z^{\mathrm{r}}}$ was assumed to create the local alternatives,

$$
\begin{aligned}
\|f^{\mathrm{r}}-g^{\mathrm{r}}\|^2_{2,\mathbb{P}_{A_n}} &= \int (f-g)^2(-u,w)(1+n^{-1/2}h(u,w))\,\mathbb{P}_{A_0}(du,dx,dw) \\
&= \int (f-g)^2(-u,w)(1+n^{-1/2}h(u,w))\,\mathbb{P}_Z(du,dw) && \text{(marginal integration)} \\
&= \int (f-g)^2(u,w)(1+n^{-1/2}h(-u,w))\,\mathbb{P}_Z(du,dw) && \text{(change of variables)} \\
&= \int (f-g)^2(u,w)(\frac{1+n^{-1/2}h(-u,w)}{1+n^{-1/2}h(u,w)})\,\mathbb{P}_{A_n}(du,dx,dw) && \text{(marginal integration)} \\
&\le (\frac{1+n^{-1/2}\|h\|_\infty}{1-n^{-1/2}\|h\|_\infty}) \|f-g\|^2_{2,\mathbb{P}_{A_n}} && \text{(for } n \ge \|h\|^2_\infty) \\
&\le (\frac{1+n^{-1/2}\|h\|_\infty}{1-n^{-1/2}\|h\|_\infty}) \delta^2.
\end{aligned}
$$

Hence, $\lim_{\delta \to 0} \limsup_{n \to \infty} t_{\delta,n}(2) = 0$ because $\|n^{1/2}(\hat{\theta} - \theta_n)\|$ is bounded in probability by Assumption 5.5. Finally, by marginal integration and Jensen,

$$
\begin{aligned}
|\Gamma_h(f-g)| \leq \int |(f-g)(h-h^{\mathrm{r}})| \, d\mathbb{P}_{A_0} &= \int \frac{|(f-g)(h-h^{\mathrm{r}})|}{1+n^{-1/2}h} \, d\mathbb{P}_{A_n} \\
&\leq \frac{2\|h\|_\infty}{1-n^{-1/2}\|h\|_\infty} \int |f-g| \, d\mathbb{P}_{A_n} \quad (\text{for } n \geq \|h\|_\infty^2) \\
&\leq \frac{2\|h\|_\infty}{1-n^{-1/2}\|h\|_\infty} \|f-g\|_{2,\mathbb{P}_{A_n}} \\
&\leq \frac{2\|h\|_\infty}{1-n^{-1/2}\|h\|_\infty} \delta.
\end{aligned}
$$

Hence, $\lim_{\delta \to 0} \limsup_{n \to \infty} t_{\delta,n}(3) = 0$ as well. Therefore, (D.4) holds. □

Proof of Lemma 5.3. In the proof of Lemma 3.3, replace references to Corollary 3.1 by Corollary 5.1, Lemma 3.2 by Lemma 5.2, and $\mathbb{X}_0$ by $\mathbb{X}_0 + \Gamma_h$. The argument leading to (B.19) goes through because: (a) the denseness result in Lemma F.3 continues to hold under H_{1n} and Assumptions 5.1, 5.2(ii), 5.4; (b) almost all sample paths of $(\mathbb{X}_0 + \Gamma_h)\mathcal{F}_1$ are uniformly $\|\cdot\|_{2,\mathbb{P}_{A_0}}$-continuous; and (c) $\sup_{f \in \mathcal{F}_1}(\mathbb{X}_0 + \Gamma_h)f$ is continuously distributed. □

Appendix E. Justification behind simulating Powell's model

As mentioned earlier, the main technical difficulty in applying our test to Powell's model stems from the fact that the symmetrically censored residual defined in (6.2) has bounded support conditional on the regressors. In this section we describe how the limiting distribution of $\hat{\mathrm{R}}_{\max}$ can be obtained under (6.1). We begin by showing how to verify (B.2) and (B.3), and thus Lemma 3.1, for Powell's model. Since the proof of Lemma 4.1 is very similar to that of Lemma 3.1, the same argument can be used to prove the validity of the simulated critical values as well.

Define $V(\theta) := Y - X'\theta$ if $Y - X'\theta < X'\theta$ and $X'\theta$ otherwise so that $V(\hat{\theta}) = \hat{V}$ and $V(\theta_0) = V$. To simplify the algebraic details, we assume that $X'\theta > 0$ w.p.1 for each θ in a small enough neighborhood of θ_0 (although Powell only requires $\Pr(X'\theta > 0) > 0$, the presence of the indicator function $\mathbb{1}(X'\theta > 0)$ to accommodate this everywhere makes the calculations much more tedious). Define $B(V, X, \theta, \theta_0) := V(\theta) - V$ and let $f \in \mathcal{F}_1$ so that $f = \mathbb{1}_{(-\infty,s] \times (-\infty,t]}$ for some $(s,t) \in \mathbb{R} \times \mathbb{R}^{\dim(X)}$. Since $f(V(\theta), X) = f(V + B(V, X, \theta, \theta_0), X) =: f^\theta(V, X)$, $f^\theta(v, x) = \mathbb{1}_{(-\infty,s] \times (-\infty,t]}(v + B(v, x, \theta, \theta_0), x) = \mathbb{1}_{(-\infty,s-B(v,x,\theta,\theta_0)]}(v)\mathbb{1}_{(-\infty,t]}(x)$. Furthermore, as $\sup_{\tilde{\theta} \in \overline{B(\theta_0, \|\theta-\theta_0\|)}} |V(\tilde{\theta}) - V(\theta_0)| \leq \|X\| \|\theta - \theta_0\|$ by equation A.11 of Powell's paper, $((s-$

$B(v, x, \theta, \theta_0)) \wedge s, (s - B(v, x, \theta, \theta_0)) \vee s] \subset (s - \|x\|\|\theta - \theta_0\|, s + \|x\|\|\theta - \theta_0\|]$. Hence,

$$\|f^{\hat\theta} - f^{\theta_0}\|_{2,\mathbb{P}_{V,X}}^2 \leq \int \int \mathbb{1}_{(s-\|x\|\|\hat\theta-\theta_0\|, s+\|x\|\|\hat\theta-\theta_0\|]}(v) \mathbb{P}_{V|X=x}(dv) \mathbb{P}_X(dx)$$

$$\leq \int \int \mathbb{1}_{(s-\|x\|\|\hat\theta-\theta_0\|, s+\|x\|\|\hat\theta-\theta_0\|]}(v) \, \mathrm{pdf}_{U^0|X=x}(dv) \mathbb{P}_X(dx) \qquad \text{(by (6.3))}$$

$$+ \int \mathbb{1}_{(s-\|x\|\|\hat\theta-\theta_0\|, s+\|x\|\|\hat\theta-\theta_0\|]}(-x'\theta_0) \, \mathbb{P}_X(dx)$$

$$+ \int \mathbb{1}_{(s-\|x\|\|\hat\theta-\theta_0\|, s+\|x\|\|\hat\theta-\theta_0\|]}(x'\theta_0) \, \mathbb{P}_X(dx)$$

$$=: a_1^{\hat\theta}(s) + a_2^{\hat\theta}(s) + a_3^{\hat\theta}(s).$$

Assume that the conditional cdf of $U^0|X = x$ is Lipschitz on $\mathbb{R}$ with Lipschitz constant $\zeta(x)$ satisfying $\int \zeta(x)\|x\| \, \mathbb{P}_X(dx) < \infty$. Then, $\sup_{s\in\mathbb{R}} a_1^{\hat\theta}(s) \leq 2\|\hat\theta - \theta_0\| \int \zeta(x)\|x\| \, \mathbb{P}_X(dx)$. Similarly, under certain conditions it can be shown that $\sup_{s\in\mathbb{R}} a_2^{\hat\theta}(s) \vee \sup_{s\in\mathbb{R}} a_3^{\hat\theta}(s) \leq \text{const.} \|\hat\theta - \theta_0\|$. [For instance, if the conditional distribution of $Q_2 := X'\theta_0$ given $Q_1 := \|X\|$ is absolutely continuous with respect to the Lebesgue measure such that $\sup_{q_2 \in \mathbb{R}} \mathrm{pdf}_{Q_2|Q_1=q_1}(q_2) \leq \xi(q_1)$ for some ξ satisfying $\int \xi(q_1) q_1 \, \mathbb{P}_{Q_1}(dq_1) < \infty$, then $a_3^{\hat\theta}(s) = \int \mathbb{1}_{(s-q_1\|\hat\theta-\theta_0\|, s+q_1\|\hat\theta-\theta_0\|]}(q_2) \, \mathbb{P}_{Q_2,Q_1}(dq_2, dq_1) \leq \int \xi(q_1) \int \mathbb{1}_{(s-q_1\|\hat\theta-\theta_0\|, s+q_1\|\hat\theta-\theta_0\|]}(q_2) \, dq_2 \, \mathbb{P}_{Q_1}(dq_1) = 2\|\hat\theta - \theta_0\| \int \xi(q_1) q_1 \, \mathbb{P}_{Q_1}(dq_1)$ and the desired conclusion is obtained as the right hand side does not depend upon s.] It follows that $\sup_{f\in\mathcal{F}_1} \|f^{\hat\theta} - f^{\theta_0}\|_{2,\mathbb{P}_{V,X}} \leq \text{const.} \|\hat\theta - \theta_0\|^{1/2}$. Consequently, the argument leading to (B.5) goes through unchanged so that we obtain (B.2) for Powell's model, i.e.,

$$\sup_{f\in\mathcal{F}_1} |n^{1/2}(\hat{\mathbb{P}}_{V,X} - \mathbb{P}_{V,X})(f^{\hat\theta} - f^{\theta_0})| = o_{\mathrm{Pr}^\circ}(1)$$

$$\sup_{f\in\mathcal{F}_1} |n^{1/2}(\hat{\mathbb{P}}_{V,X} - \mathbb{P}_{V,X})(f^{\mathrm{r}\hat\theta} - f^{\mathrm{r}\theta_0})| = o_{\mathrm{Pr}^\circ}(1). \tag{E.1}$$

[The second result follows from the first upon replacing f by f^{r}.] To verify (B.3) for Powell's model, begin by observing that

$$B(V, X, \hat\theta, \theta_0) = \begin{cases} X'(\hat\theta - \theta_0) & \text{if } Y - X'\hat\theta \geq X'\hat\theta \text{ and } Y - X'\theta_0 \geq X'\theta_0 \\ X'(\theta_0 - \hat\theta) & \text{if } Y - X'\hat\theta < X'\hat\theta \text{ and } Y - X'\theta_0 < X'\theta_0 \\ X'\hat\theta - Y + X'\theta_0 & \text{if } Y - X'\hat\theta \geq X'\hat\theta \text{ and } Y - X'\theta_0 < X'\theta_0 \\ Y - X'\hat\theta - X'\theta_0 & \text{if } Y - X'\hat\theta < X'\hat\theta \text{ and } Y - X'\theta_0 \geq X'\theta_0. \end{cases}$$

The last two cases are determined by the events $0 < 2X'\hat\theta \leq Y < 2X'\theta_0$ and $0 < 2X'\theta_0 \leq Y < 2X'\hat\theta$, the probabilities of which go to zero as $\hat\theta \to \theta_0$. Consequently, $B(V, X, \hat\theta, \theta_0) = \pm X'(\hat\theta - \theta_0)$ w.p.a.1 because the SCLS estimator is strongly consistent for θ_0. Furthermore, as will become clear from subsequent results, $\mathbb{P}_{V,X} f^{\hat\theta}$ depends upon $\hat\theta$ only through $\hat\theta - \theta_0$ which in large samples behaves like a gaussian random vector centered at zero, i.e., asymptotically, the distribution

of $\hat{\theta} - \theta_0$ is symmetric about the origin. Therefore, without loss of generality, the probability calculations for the remainder of this section are conditional on the first two cases and we can replace $B(V, X, \hat{\theta}, \theta_0)$ by $\Delta(X, \hat{\theta}, \theta_0) = X'(\hat{\theta} - \theta_0)$ mimicking our earlier notation. So let $r(x) := \mathbb{P}_{U^0|X=x}(U^0 \leq -x'\theta_0) = \mathbb{P}_{U^0|X=x}(U^0 \geq x'\theta_0)$. Then, by (6.3),

$$
\begin{aligned}
\mathbb{P}_{V,X} f^{\hat{\theta}} &= \int_{\mathrm{supp}(X)} \int_{[-x'\theta_0, x'\theta_0]} f(v + \Delta(x, \hat{\theta}, \theta_0), x) \mathbb{P}_{V|X=x}(dv)\, \mathbb{P}_X(dx) \\
&= \int \int_{-x'\theta_0}^{x'\theta_0} f(v + \Delta(x, \hat{\theta}, \theta_0), x) p_{U^0|X=x}(v)\, dv\, \mathbb{P}_X(dx) \\
&\quad + \int f(-x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0), x) r(x)\, \mathbb{P}_X(dx) + \int f(x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0), x) r(x)\, \mathbb{P}_X(dx) \\
&= \int \int_{-x'\theta_0 + \Delta(x, \theta, \theta_0)}^{x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0)} f(t, x) p_{U^0|X=x}(t - \Delta(x, \hat{\theta}, \theta_0))\, dt\, \mathbb{P}_X(dx) \\
&\quad + \int f(-x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0), x) r(x)\, \mathbb{P}_X(dx) + \int f(x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0), x) r(x)\, \mathbb{P}_X(dx) \\
&= \int \int_{-x'\theta_0}^{x'\theta_0} f(t, x) p_{U^0|X=x}(t - \Delta(x, \hat{\theta}, \theta_0)) dt\, \mathbb{P}_X(dx) \\
&\quad + \int \int_{x'\theta_0}^{x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0)} f(t, x) p_{U^0|X=x}(t - \Delta(x, \hat{\theta}, \theta_0))\, dt\, \mathbb{P}_X(dx) \\
&\quad - \int \int_{-x'\theta_0}^{-x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0)} f(t, x) p_{U^0|X=x}(t - \Delta(x, \hat{\theta}, \theta_0))\, dt\, \mathbb{P}_X(dx) \\
&\quad + \int f(-x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0), x) r(x)\, \mathbb{P}_X(dx) + \int f(x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0), x) r(x)\, \mathbb{P}_X(dx).
\end{aligned}
$$

Therefore,

$$\mathbb{P}_{V,X}(f^{\hat{\theta}} - f^{\theta_0}) = \int \int_{-x'\theta_0}^{x'\theta_0} f(t,x)[p_{U^0,X}(t - \Delta(x,\hat{\theta},\theta_0),x)/p_{U^0,X}(t,x) - 1]\,\mathbb{P}_{U^0,X}(dt,dx)$$

$$+ \int \int_{x'\theta_0}^{x'\theta_0 + \Delta(x,\hat{\theta},\theta_0)} f(t,x)p_{U^0|X=x}(t - \Delta(x,\hat{\theta},\theta_0))\,dt\,\mathbb{P}_X(dx)$$

$$- \int \int_{-x'\theta_0}^{-x'\theta_0 + \Delta(x,\hat{\theta},\theta_0)} f(t,x)p_{U^0|X=x}(t - \Delta(x,\hat{\theta},\theta_0))\,dt\,\mathbb{P}_X(dx)$$

$$+ \int f(-x'\theta_0 + \Delta(x,\hat{\theta},\theta_0),x)r(x)\,\mathbb{P}_X(dx)$$

$$+ \int f(x'\theta_0 + \Delta(x,\hat{\theta},\theta_0),x)r(x)\,\mathbb{P}_X(dx)$$

$$- \int f(-x'\theta_0,x)r(x)\,\mathbb{P}_X(dx)$$

$$- \int f(x'\theta_0,x)r(x)\,\mathbb{P}_X(dx)$$

$$=: b_1^{\hat{\theta}}(f) + b_2^{\hat{\theta}}(f) - b_3^{\hat{\theta}}(f) + b_4^{\hat{\theta}}(f) + b_5^{\hat{\theta}}(f) - b_6(f) - b_7(f).$$

As with (B.6), we can show that $b_1^{\hat{\theta}}(f) = -\langle f, \mathbb{1}_{(-X'\theta_0,X'\theta_0)}m'_{U^0,X}\rangle_{\mathbb{P}_{U^0,X}}(\hat{\theta} - \theta_0) + o(\|\hat{\theta} - \theta_0\|)$ uniformly in $f \in \mathcal{F}_1$, where $m_{U^0,X} := (\partial_1 \log p_{U^0,X})X$. Therefore

$$\sup_{f \in \mathcal{F}_1} |\mathbb{P}_{V,X}(f^{\hat{\theta}} - f^{\theta_0}) + \langle f, \mathbb{1}_{(-X'\theta_0,X'\theta_0)}m'_{U^0,X}\rangle_{\mathbb{P}_{U^0,X}}(\hat{\theta} - \theta_0)$$

$$- (b_2^{\hat{\theta}}(f) - b_3^{\hat{\theta}}(f) + b_4^{\hat{\theta}}(f) + b_5^{\hat{\theta}}(f) - b_6(f) - b_7(f))| = o(\|\hat{\theta} - \theta_0\|).$$

Similarly, replacing f by f^{r},

$$\sup_{f \in \mathcal{F}_1} |\mathbb{P}_{V,X}(f^{\mathrm{r}\hat{\theta}} - f^{\mathrm{r}\theta_0}) + \langle f^{\mathrm{r}}, \mathbb{1}_{(-X'\theta_0,X'\theta_0)}m'_{U^0,X}\rangle_{\mathbb{P}_{U^0,X}}(\hat{\theta} - \theta_0)$$

$$- (b_2^{\hat{\theta}}(f^{\mathrm{r}}) - b_3^{\hat{\theta}}(f^{\mathrm{r}}) + b_4^{\hat{\theta}}(f^{\mathrm{r}}) + b_5^{\hat{\theta}}(f^{\mathrm{r}}) - b_6(f^{\mathrm{r}}) - b_7(f^{\mathrm{r}}))| = o(\|\hat{\theta} - \theta_0\|).$$

Therefore, since $b_6(f) = b_7(f^{\mathrm{r}})$ and $b_7(f) = b_6(f^{\mathrm{r}})$,

$$\sup_{f \in \mathcal{F}_1} |\mathbb{P}_{V,X}(f^{\hat{\theta}} - f^{\mathrm{r}\hat{\theta}}) + \langle f - f^{\mathrm{r}}, \mathbb{1}_{(-X'\theta_0,X'\theta_0)}m'_{U^0,X}\rangle_{\mathbb{P}_{U^0,X}}(\hat{\theta} - \theta_0)$$

$$- b_2^{\hat{\theta}}(f - f^{\mathrm{r}}) + b_3^{\hat{\theta}}(f - f^{\mathrm{r}}) - (b_4^{\hat{\theta}}(f) - b_5^{\hat{\theta}}(f^{\mathrm{r}})) - (b_5^{\hat{\theta}}(f) - b_4^{\hat{\theta}}(f^{\mathrm{r}}))| = o(\|\hat{\theta} - \theta_0\|). \quad \text{(E.2)}$$

By (E.1) and (E.2), $n^{1/2}(\hat{\mathbb{P}}_{\hat{V},X} - \hat{\mathbb{P}}_{\hat{V}^{\mathrm{r}},X})f = \hat{\mathbb{X}}_0(f) + o_{\mathrm{Pr}}(1)$, where

$$\hat{\mathbb{X}}_0(f) := n^{1/2}(\hat{\mathbb{P}}_{V,X} - \mathbb{P}_{V,X})(f - f^{\mathrm{r}}) + \langle f - f^{\mathrm{r}}, \mathbb{1}_{(-X'\theta_0,X'\theta_0)}m'_{U^0,X}\rangle_{\mathbb{P}_{U^0,X}}n^{1/2}(\hat{\theta} - \theta_0)$$

$$- n^{1/2}(b_2^{\hat{\theta}} - b_3^{\hat{\theta}})(f - f^{\mathrm{r}}) - n^{1/2}(b_4^{\hat{\theta}}(f) - b_5^{\hat{\theta}}(f^{\mathrm{r}})) - n^{1/2}(b_5^{\hat{\theta}}(f) - b_4^{\hat{\theta}}(f^{\mathrm{r}})).$$

Consequently, $n^{1/2}(\hat{\mathbb{P}}_{\hat{V},X} - \hat{\mathbb{P}}_{\hat{V}^{\mathrm{r}},X})$ converges in distribution provided $\hat{\mathbb{X}}_0$ is asymptotically tight. Since tightness of $n^{1/2}(\hat{\mathbb{P}}_{V,X} - \mathbb{P}_{V,X})(f - f^{\mathrm{r}}) - \langle f - f^{\mathrm{r}}, \mathbb{1}_{(-X'\theta_0, X'\theta_0)} m'_{U^0,X} \rangle_{\mathbb{P}_{U^0,X}} n^{1/2}(\hat{\theta} - \theta_0)$ follows directly from Lemmas 3.2 and 3.3, it suffices to show that the remaining terms are asymptotically tight. Assume that $p_{U^0|X}$ is differentiable with derivative $\dot{p}_{U^0|X}$ satisfying $\sup_{t \in \mathbb{R}} |\dot{p}_{U^0|X=x}(t)| \le c_1(x)$ and $\int c_1(x) \|x\|^2 \mathbb{P}_X(dx) < \infty$. Then, uniformly in $f \in \mathcal{F}_1$, $b_2^{\hat{\theta}}(f) = \int \int_{x'\theta_0}^{x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0)} f(u,x) p_{U^0|X=x}(u) \, du \, \mathbb{P}_X(dx) + o(\|\hat{\theta} - \theta_0\|)$. [A similar result holds for $b_3^{\theta}(f)$ as well.] Therefore,

$$(b_2^{\hat{\theta}} - b_3^{\hat{\theta}})(f - f^{\mathrm{r}}) = \int \int_{x'\theta_0 - \Delta(x, \hat{\theta}, \theta_0)}^{x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0)} (f(u,x) - f^{\mathrm{r}}(u,x)) p_{U^0|X=x}(u) \, du \, \mathbb{P}_X(dx) + o(\|\hat{\theta} - \theta_0\|) \quad \text{(E.3)}$$

holds uniformly in $f \in \mathcal{F}_1$. To prove that $(b_2^{\hat{\theta}} - b_3^{\hat{\theta}})(\cdot)$ is tight, it suffices to show the tightness of the process $\{\int \int_{x'\theta_0 - \Delta(x, \hat{\theta}, \theta_0)}^{x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0)} f(u,x) p_{U^0|X=x}(u) \, du \, \mathbb{P}_X(dx) : f \in \mathcal{F}_1\}$. So let $f, g \in \mathcal{F}_1$. Then,

$$\int \int_{x'\theta_0 - \Delta(x, \hat{\theta}, \theta_0)}^{x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0)} (f(u,x) - g(u,x)) p_{U^0|X=x}(u) \, du \, \mathbb{P}_X(dx)$$

$$= \int \frac{1}{2\Delta(x, \hat{\theta}, \theta_0)} \int_{x'\theta_0 - \Delta(x, \hat{\theta}, \theta_0)}^{x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0)} (f(u,x) - g(u,x)) p_{U^0|X=x}(u) \, du \, 2\Delta(x, \hat{\theta}, \theta_0) \, \mathbb{P}_X(dx)$$

$$\le 2\|\hat{\theta} - \theta_0\| \int H_x(f - g)(x'\theta_0) \|x\| \, \mathbb{P}_X(dx)$$

$$\le 2\|\hat{\theta} - \theta_0\| (\int H_x^2(f - g)(x'\theta_0) \, \mathbb{P}_X(dx))^{1/2} (\int \|x\|^2 \, \mathbb{P}_X(dx))^{1/2},$$

where $H_x(f)(t) := \sup_{r>0} \frac{1}{2r} \int_{t-r}^{t+r} |f(u,x)| p_{U^0|X=x}(u) \, du$ is the Hardy-Littlewood maximal function with respect to the Lebesgue measure. If $\sup_{(x,t) \in \mathrm{supp}(X) \times \mathrm{supp}(X'\theta_0)} p_{X|X'\theta_0=t}(x) < \infty$ and $\sup_{t \in \mathrm{supp}(X'\theta_0)} p_{X'\theta_0}(t) < \infty$, then

$$\int H_x^2(f - g)(x'\theta_0) \, \mathbb{P}_X(dx) = \int (\int H_x^2(f - g)(t) \, p_{X|X'\theta_0=t}(x) \, dx) \, p_{X'\theta_0}(t) \, dt$$

$$\le \text{const.} \int \int H_x^2(f - g)(t) \, dt \, dx$$

$$\le \text{const.} \int \int (f - g)^2(t, x) p_{U^0|X=x}^2(t) \, dt \, dx$$

$$\le \text{const.} \|f - g\|_{2, \mathbb{P}_{U^0,X}}^2,$$

where the second inequality follows by the boundedness of the Hardy-Littlewood operator (Stein, 1970, Theorem I.1) and the third by assuming that $\sup_{(t,x) \in \mathbb{R} \times \mathrm{supp}(X)} p_{U^0|X=x}(t) < \infty$ and $\inf_{x \in \mathrm{supp}(X)} p_X(x) > 0$. Hence,

$$|\int \int_{x'\theta_0 - \Delta(x, \hat{\theta}, \theta_0)}^{x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0)} (f(u,x) - g(u,x)) p_{U^0|X=x}(u) \, du \, \mathbb{P}_X(dx)| \le \text{const.} \|f - g\|_{2, \mathbb{P}_{U^0,X}} \|\hat{\theta} - \theta_0\|;$$

consequently, the process $\{n^{1/2}(b_2^{\hat{\theta}} - b_3^{\hat{\theta}})(f - f^{\mathrm{r}}) : f \in \mathcal{F}_1\}$ is asymptotically tight (cf. the argument at the end of the proof of Lemma 3.2). We can use the same argument to show that $n^{1/2}(b_2^{\hat{\theta}} - b_3^{\hat{\theta}})$ is linear in $n^{1/2}(\hat{\theta} - \theta_0)$. Indeed, by dominated convergence and the Lebesgue differentiation theorem (Stein, 1970, Corollary I.1) applied to (E.3),

$$n^{1/2}(b_2^{\hat{\theta}} - b_3^{\hat{\theta}})(f - f^{\mathrm{r}}) = 2\langle (f - f^{\mathrm{r}})(X'\theta_0, X), p_{U^0|X}(X'\theta_0)X' \rangle_{\mathbb{P}_X} n^{1/2}(\hat{\theta} - \theta_0)$$
$$+ o(\|n^{1/2}(\hat{\theta} - \theta_0)\|) \qquad \text{unif. in } f \in \mathcal{F}_1.$$

Finally, under certain conditions it can be shown that the map $t \mapsto \int f(x'\theta_0 + t)r(x)\mathbb{P}_X(dx)$ is uniformly differentiable in the sense that there exists $g \in L_2(\mathbb{P}_{U^0, X})$ such that

$$\int f(x'\theta_0 + \Delta(x, \hat{\theta}, \theta_0), x)r(x)\mathbb{P}_X(dx)$$
$$= \int f(x'\theta_0, x)r(x)\mathbb{P}_X(dx) + \langle f, g' \rangle_{\mathbb{P}_{U^0, X}}(\hat{\theta} - \theta_0) + o(\|\hat{\theta} - \theta_0\|) \qquad \text{unif. in } f \in \mathcal{F}_1.$$

[To get some intuition behind this condition, consider the case when f takes only one argument and belongs to the class of trapezoidal functions. Let $q(u) := \mathbb{E}[r(X)|X'\theta_0 = u]p_{X'\theta_0}(u)$ so that $\int f(x'\theta_0 + t)r(x)\mathbb{P}_X(dx) = \int f(u + t)q(u)\,du$. By a Taylor expansion, $\int f(u + t)q(u)\,du = \int f(u)q(u)\,du + t\int f'(u)q(u)\,du$ because the second derivative of trapezoidal functions is zero almost everywhere. If the density of $X'\theta_0$ vanishes at the boundary of its support then (doing integration by parts) $\int f'(u)q(u)\,du = -\int f(u)S(u)\,\mathbb{P}_{U^0}(du)$ with $S(u) := q'(u)/p_{U^0}(u)$. Consequently, $\int f(x'\theta_0 + t)r(x)\mathbb{P}_X(dx) = \int f(x'\theta_0)r(x)\mathbb{P}_X(dx) - \langle f, S \rangle_{\mathbb{P}_{U^0}} t$. The corresponding result when f is an indicator function now follows because indicators can be approximated arbitrarily well by trapezoids.] From this condition we have that

$$n^{1/2}(b_5^{\hat{\theta}}(f) - b_4^{\hat{\theta}}(f^{\mathrm{r}})) = n^{1/2}(b_4^{\hat{\theta}}(f) - b_5^{\hat{\theta}}(f^{\mathrm{r}})) = 2\langle f, g' \rangle_{\mathbb{P}_{U^0, X}} n^{1/2}(\hat{\theta} - \theta_0) + o(\|n^{1/2}(\hat{\theta} - \theta_0)\|)$$

uniformly in $f \in \mathcal{F}_1$, implying that $n^{1/2}(b_4^{\hat{\theta}}(f) - b_5^{\hat{\theta}}(f^{\mathrm{r}}))$ and $n^{1/2}(b_5^{\hat{\theta}}(f) - b_4^{\hat{\theta}}(f^{\mathrm{r}}))$ are asymptotically tight as well. Therefore, combining results, $\{\hat{\mathbb{X}}_0(f) : f \in \mathcal{F}_1\}$ is asymptotically tight implying that $\{n^{1/2}(\hat{\mathbb{P}}_{\hat{V}, X} - \hat{\mathbb{P}}_{\hat{V}^{\mathrm{r}}, X})f : f \in \mathcal{F}_1\}$ converges in distribution in $\ell^\infty(\mathcal{F}_1)$. The limiting process is gaussian because the finite dimensional distributions of $\{\hat{\mathbb{X}}_0(f) : f \in \mathcal{F}_1\}$ are gaussian due to the fact that

$$\hat{\mathbb{X}}_0(f) = n^{1/2}(\hat{\mathbb{P}}_{V, X} - \mathbb{P}_{V, X})(f - f^{\mathrm{r}}) + \langle f - f^{\mathrm{r}}, \mathbb{1}_{(-X'\theta_0, X'\theta_0)} m'_{U^0, X} \rangle_{\mathbb{P}_{U^0, X}} n^{1/2}(\hat{\theta} - \theta_0)$$
$$- 2\langle (f - f^{\mathrm{r}})(X'\theta_0, X), p_{U^0|X}(X'\theta_0)X' \rangle_{\mathbb{P}_X} n^{1/2}(\hat{\theta} - \theta_0) - 4\langle f, g' \rangle_{\mathbb{P}_{U^0, X}} n^{1/2}(\hat{\theta} - \theta_0)$$

and $n^{1/2}(\hat{\theta} - \theta_0)$ is asymptotically linear (cf. equation A.15 in Powell's paper). Applying the denseness result (Lemma F.3), it follows that the limiting distribution of $n^{1/2}\hat{R}_{\max}$ is the norm of the aforementioned process.

Appendix F. Auxiliary results

Lemma F.1. *Under the conditions of Theorem* 5.1,

$$\sup_{f\in\mathcal{F}}|(\hat{\mathbb{P}}_{\hat{Z}}-\hat{\mathbb{P}}_{Z(\theta_1)})f|=o_{\mathrm{Pr}^\circ}(1)\quad\&\quad\sup_{f\in\mathcal{F}}|(\hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}}-\hat{\mathbb{P}}_{Z^{\mathrm{r}}(\theta_1)})f|=o_{\mathrm{Pr}^\circ}(1).$$

Proof of Lemma F.1. Since the proof of Lemma 3.1 only requires the existence of $\mathrm{plim}(\hat{\theta})$, the argument leading to (B.1) can be replicated word for word with (θ_0,ε) replaced by $(\theta_1,\varepsilon(\theta_1))$ and A_0 replaced by $A(\theta_1)$ to show that

$$\sup_{f\in\mathcal{F}}|(\hat{\mathbb{P}}_{\hat{Z}}-\hat{\mathbb{P}}_{Z(\theta_1)})f-\langle f,(\partial_1\log p_{A(\theta_1)})\dot{\mu}_1'\rangle_{\mathbb{P}_{A(\theta_1)}}(\hat{\theta}-\theta_1)|=o_{\mathrm{Pr}^\circ}(n^{-1/2})+o_{\mathrm{Pr}}(\|\hat{\theta}-\theta_1\|).$$

Therefore, $\sup_{f\in\mathcal{F}}|(\hat{\mathbb{P}}_{\hat{Z}}-\hat{\mathbb{P}}_{Z(\theta_1)})f|=o_{\mathrm{Pr}^\circ}(1)$. The second result follows from the first upon noting that $(\hat{\mathbb{P}}_{\hat{Z}^{\mathrm{r}}}-\hat{\mathbb{P}}_{Z^{\mathrm{r}}(\theta_1)})f=(\hat{\mathbb{P}}_{\hat{Z}}-\hat{\mathbb{P}}_{Z(\theta_1)})f^{\mathrm{r}}$. □

Lemma F.2. *Under Assumption* 3.1 *and* 3.2(ii), $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *satisfy Assumption* 3.3(v) *with* $q(t)\propto t^{1/2}$.

Proof of Lemma F.2. If $f\in\mathcal{F}_1$, then $f=\mathbb{1}_{(-\infty,\tau]\times(-\infty,v]}$ for some $(\tau,v)\in\mathbb{R}\times\mathbb{R}^{\dim(W)}$ and

$$f(u-\Delta(x,\theta,\theta_0),w)=\mathbb{1}_{(-\infty,\tau]\times(-\infty,v]}(u-\Delta(x,\theta,\theta_0),w)=\mathbb{1}_{(-\infty,\tau+\Delta(x,\theta,\theta_0)]}(u)\mathbb{1}_{(-\infty,v]}(w).$$

Since $|\mathbb{1}_{(-\infty,a]}-\mathbb{1}_{(-\infty,b]}|=\mathbb{1}_{(a\wedge b,a\vee b]}$ and $a\vee b-a\wedge b=|a-b|$, it follows that

$$\int(f(u-\Delta(x,\theta,\theta_0),w)-f(u,w))^2\,\mathbb{P}_{\varepsilon|X=x,W=w}(du)$$

$$=\int_{((\tau+\Delta(x,\theta,\theta_0))\wedge\tau,(\tau+\Delta(x,\theta,\theta_0))\vee\tau]}\mathbb{P}_{\varepsilon|X=x,W=w}(du)\,\mathbb{1}_{(-\infty,v]}(w)$$

$$\le\zeta(x,w)|\Delta(x,\theta,\theta_0)|\qquad\qquad\text{(by Assumption 3.2(ii))}$$

$$\le\zeta(x,w)(\|\dot{\mu}_0(x)\|\,\|\theta-\theta_0\|+|\rho(x,\theta,\theta_0)|)\qquad\text{(by Assumption 3.1)}$$

$$\le\zeta(x,w)(\|\dot{\mu}_0(x)\|\,\|\theta-\theta_0\|+\sup_{\tilde{\theta}\in\overline{B(\theta_0,\|\theta-\theta_0\|)}}|\rho(x,\tilde{\theta},\theta_0)|).$$

Hence, by Cauchy-Schwarz and Assumption 3.1, for all $\epsilon>0$,

$$\int(f(u-\Delta(x,\theta,\theta_0),w)-f(u,w))^2\,\mathbb{P}_{A_0}(du,dx,dw)$$

$$\le\|\zeta\|_{2,\mathbb{P}_{X,W}}(\|(\dot{\mu}_0'\dot{\mu}_0)^{1/2}\|_{2,\mathbb{P}_X}\,\|\theta-\theta_0\|+\|\sup_{\tilde{\theta}\in\overline{B(\theta_0,\|\theta-\theta_0\|)}}|\rho(\cdot,\tilde{\theta},\theta_0)|\|_{2,\mathbb{P}_X})$$

$$\le\|\zeta\|_{2,\mathbb{P}_{X,W}}(\|(\dot{\mu}_0'\dot{\mu}_0)^{1/2}\|_{2,\mathbb{P}_X}+\epsilon)\|\theta-\theta_0\|.$$

Therefore, since the right hand side does not depend upon (τ,v),

$$\sup_{f\in\mathcal{F}_1}\|f(\cdot-\Delta(X,\theta,\theta_0),\cdot)-f(\cdot,\cdot)\|_{2,\mathbb{P}_{A_0}}\le\|\zeta\|_{2,\mathbb{P}_{X,W}}^{1/2}(\|(\dot{\mu}_0'\dot{\mu}_0)^{1/2}\|_{2,\mathbb{P}_X}+\epsilon)^{1/2}\|\theta-\theta_0\|^{1/2};$$

i.e., $\mathcal{F}_1$ satisfies Assumption 3.3(v) with $q(t) \propto t^{1/2}$. Since $f \in \mathcal{F}_1 \implies f^{\mathrm{r}} = \mathbb{1}_{[-\tau,\infty)\times(-\infty,v]}$ and $\mathbb{P}_{\varepsilon|X,W}(\{-\tau\}) = 0$, the above argument shows that Assumption 3.3(v) is also satisfied with f^{r} and $q(t) \propto t^{1/2}$. The same reasoning works for $\mathcal{F}_2$ as well. $\qquad\square$

The following result may be of independent interest.

Lemma F.3. *Let Assumptions* 3.1, 3.2(ii), *and* 3.4 *hold. Then,* $\hat{\mathcal{F}}_1$ *is dense in* $\mathcal{F}_1$ *in the following sense: For each* $f \in \mathcal{F}_1$ *and* $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr(\exists \hat{f} \in \hat{\mathcal{F}}_1 : \|\hat{f} - f\|_{2,\mathbb{P}_{A_0}} \vee \|\hat{f}^{\mathrm{r}} - f^{\mathrm{r}}\|_{2,\mathbb{P}_{A_0}} < \epsilon) = 1. \tag{F.1}$$

Proof of Lemma F.3. We make use of the following fact (van der Vaart, 1998, Theorem 19.1): Given $\epsilon > 0$ and any cdf F, there exists a partition $\{-\infty =: t_0 < t_1 < \ldots < t_K := \infty\}$ of $\mathbb{R}$ such that $F(t_l-) - F(t_{l-1}) < \epsilon$ for each l (points at which F jumps more than ϵ belong to the partition). Hence, letting F_ε and $F_{W^{(i)}}$ denote the marginal cdfs of ε and $W^{(i)}$ and working coordinate by coordinate, we can create a partition of $\mathbb{R} \times \mathbb{R}^{\dim(W)}$ denoted by $\mathcal{Z}_\pi :=$ $\{\tau_0 < \tau_1 < \ldots < \tau_K\} \times \times_{i=1}^{\dim(W)}\{\vartheta_0^{(i)} < \vartheta_1^{(i)} < \ldots < \vartheta_{K_i}^{(i)}\}$ such that $F_\varepsilon(\tau_l) - F_\varepsilon(\tau_{l-1}) < \epsilon$ and $F_{W^{(i)}}(\vartheta_l^{(i)}-) - F_{W^{(i)}}(\vartheta_{l-1}^{(i)}) < \epsilon$ for each l (remember that ε is assumed to be continuously distributed). For the remainder of the proof, let $c := 1 + \dim(W)$ and $d := \dim(W)$.

Fix $f \in \mathcal{F}_1$. Then, $f = \mathbb{1}_{(-\infty,u]\times(-\infty,v]}$ for some $(u,v) \in \mathbb{R} \times \mathbb{R}^d$ and $f^{\mathrm{r}} = \mathbb{1}_{[-u,\infty)\times(-\infty,v]}$. We begin by showing that, given $\mathcal{Z}_\pi$, there exist functions f_π and $\tilde{f}_\pi$ such that

$$\mathrm{d}(f, f_\pi; f^{\mathrm{r}}, \tilde{f}_\pi) := \|f - f_\pi\|_{2,\mathbb{P}_{A_0}} \vee \|f^{\mathrm{r}} - \tilde{f}_\pi\|_{2,\mathbb{P}_{A_0}} < (c\epsilon)^{1/2}. \tag{F.2}$$

Since $\mathcal{Z}_\pi$ partitions $\mathbb{R} \times \mathbb{R}^d$, we can find $z_\pi := (\tau_l, \vartheta_{l_1}^{(1)}, \ldots, \vartheta_{l_d}^{(d)}) \in \mathcal{Z}_\pi$ such that $(u,v) \in$ $[\tau_l, \tau_{l+1}) \times \times_{i=1}^d[\vartheta_{l_i}^{(i)}, \vartheta_{l_i+1}^{(i)})$. Similarly, there exists $\tilde{z}_\pi := (\tau_{\tilde{l}}, \vartheta_{l_1}^{(1)}, \ldots, \vartheta_{l_d}^{(d)}) \in \mathcal{Z}_\pi$ such that $(-u,v) \in [\tau_{\tilde{l}}, \tau_{\tilde{l}+1}) \times \times_{i=1}^d[\vartheta_{l_i}^{(i)}, \vartheta_{l_i+1}^{(i)})$. Now, if $(A_i)_{i=1}^k$ and $(B_i)_{i=1}^k$ are subsets of $\mathbb{R}$ then using the fact that $(\cap_i A_i) \triangle (\cap_i B_i) \subset \cup_i (A_i \triangle B_i)$, where $A_i \triangle B_i := (A_i \setminus B_i) \cup (B_i \setminus A_i)$ is the symmetric difference of A_i and B_i, it is straightforward to show that $\mathbb{1}_{(\times_{i=1}^k A_i)\triangle(\times_{i=1}^k B_i)} \leq \sum_{i=1}^k \mathbb{1}_{A_i \triangle B_i}$. Hence, letting $\vartheta := (\vartheta_{l_1}^{(1)}, \ldots, \vartheta_{l_d}^{(d)})$, $f_\pi := \mathbb{1}_{(-\infty,\tau_l]\times(-\infty,\vartheta]}$, and $\tilde{f}_\pi := \mathbb{1}_{[\tau_{\tilde{l}},\infty)\times(-\infty,\vartheta]}$, we have that

$$|f - f_\pi|^2 = |\mathbb{1}_{(-\infty,u]\times(-\infty,v]} - \mathbb{1}_{(-\infty,\tau_l]\times(-\infty,\vartheta]}|^2 = \mathbb{1}_{(-\infty,u]\times(-\infty,v]\triangle(-\infty,\tau_l]\times(-\infty,\vartheta]}$$

$$\leq \mathbb{1}_{(-\infty,u]\triangle(-\infty,\tau_l]} + \sum_{i=1}^d \mathbb{1}_{(-\infty,v^{(i)}]\triangle(-\infty,\vartheta_{l_i}^{(i)}]}$$

$$= \mathbb{1}_{(u\wedge\tau_l, u\vee\tau_l]} + \sum_{i=1}^d \mathbb{1}_{(v^{(i)}\wedge\vartheta_{l_i}^{(i)}, v^{(i)}\vee\vartheta_{l_i}^{(i)}]}$$

$$\leq \mathbb{1}_{(\tau_l,\tau_{l+1})} + \sum_{i=1}^d \mathbb{1}_{(\vartheta_{l_i}^{(i)},\vartheta_{l_i+1}^{(i)})}. \tag{F.3}$$

Similarly,

$$|f^{\mathrm{r}} - \tilde{f}_\pi|^2 = \mathbb{1}_{[-u,\infty)\times(-\infty,v]\triangle[\tau_{\tilde{l}},\infty)\times(-\infty,\vartheta]} \leq \mathbb{1}_{[-u\wedge\tau_{\tilde{l}},-u\vee\tau_{\tilde{l}})} + \sum_{i=1}^{d} \mathbb{1}_{(v^{(i)}\wedge\vartheta_{l_i}^{(i)},v^{(i)}\vee\vartheta_{l_i}^{(i)}]}$$

$$\leq \mathbb{1}_{[\tau_{\tilde{l}},\tau_{\tilde{l}+1})} + \sum_{i=1}^{d} \mathbb{1}_{(\vartheta_{l_i}^{(i)},\vartheta_{l_i+1}^{(i)})}.$$

Let $\mathrm{supp}(W) := \times_{i=1}^{d}[a^{(i)}, b^{(i)}]$ with the convention that $[-\infty, \infty] := (-\infty, \infty)$, $[-\infty, \cdot] := (-\infty, \cdot]$, and $[\cdot, \infty] := [\cdot, \infty)$ so that coordinates of W are allowed to have unbounded support under this notation. By (F.3),

$$\|f - f_\pi\|_{2,\mathbb{P}_{A_0}}^2 \leq \int_{\tau_l}^{\tau_{l+1}} d\mathbb{P}_\varepsilon + \sum_{i=1}^{d} \int \mathbb{1}_{(\vartheta_{l_i}^{(i)}\vee a^{(i)},\vartheta_{l_i+1}^{(i)}\wedge b^{(i)})} d\mathbb{P}_{W^{(i)}} \tag{F.4}$$

$$\leq F_\varepsilon(\tau_{l+1}) - F_\varepsilon(\tau_l) + \sum_{i=1}^{d} F_{W^{(i)}}(\vartheta_{l_i+1}^{(i)} \wedge b^{(i)}-) - F_{W^{(i)}}(\vartheta_{l_i}^{(i)} \vee a^{(i)}) < c\epsilon.$$

Similarly, using the fact that $F_\varepsilon(\tau_{\tilde{l}+1}) - F_\varepsilon(\tau_{\tilde{l}}) < \epsilon$,

$$\|f^{\mathrm{r}} - \tilde{f}_\pi\|_{2,\mathbb{P}_{A_0}}^2 < c\epsilon. \tag{F.5}$$

Therefore, (F.2) is proved. Next, note that since $\vartheta_{l_i}^{(i)}$ and $\vartheta_{l_i+1}^{(i)}$ enter (F.4) only via $\vartheta_{l_i}^{(i)} \vee a^{(i)}$ and $\vartheta_{l_i+1}^{(i)} \wedge b^{(i)}$, we can assume without loss of generality that each $v^{(i)} \in [\vartheta_{l_i}^{(i)} \vee a^{(i)}, \vartheta_{l_i+1}^{(i)} \wedge b^{(i)})$ so that $(u, v) \in \mathcal{E}_\pi \times \mathcal{W}_\pi \subset \mathbb{R} \times \mathrm{supp}(W)$ and $(-u, v) \in \tilde{\mathcal{E}}_\pi \times \mathcal{W}_\pi$ with $\mathcal{E}_\pi := [\tau_l, \tau_{l+1})$, $\tilde{\mathcal{E}}_\pi := [\tau_{\tilde{l}}, \tau_{\tilde{l}+1})$, $\mathcal{W}_\pi := \times_{i=1}^{d}[\vartheta_{l_i}^{(i)} \vee a^{(i)}, \vartheta_{l_i+1}^{(i)} \wedge b^{(i)})$, $\mathbb{P}_{\varepsilon,W}(\mathcal{E}_\pi \times \mathcal{W}_\pi) > 0$, and $\mathbb{P}_{\varepsilon,W}(\tilde{\mathcal{E}}_\pi \times \mathcal{W}_\pi) > 0$. Next, replacing (u, v) by $Z_j := (\varepsilon_j, W_j)$ and $(-u, v)$ by $Z_j^{\mathrm{r}} := (-\varepsilon_j, W_j)$ in the argument leading to (F.5), we get that for $f_j := \mathbb{1}_{(-\infty,Z_j]} = \mathbb{1}_{(-\infty,\varepsilon_j]\times(-\infty,W_j]}$,

$$Z_j \in \mathcal{E}_\pi \times \mathcal{W}_\pi =: \mathcal{S}_\pi \quad \& \quad Z_j^{\mathrm{r}} \in \tilde{\mathcal{E}}_\pi \times \mathcal{W}_\pi =: \tilde{\mathcal{S}}_\pi \implies \mathrm{d}(f_j, f_\pi; f_j^{\mathrm{r}}, \tilde{f}_\pi) < (c\epsilon)^{1/2}. \tag{F.6}$$

Recall that $\hat{Z}_j := (\hat{\varepsilon}_j, W_j)$ and $\hat{Z}_j^{\mathrm{r}} := (-\hat{\varepsilon}_j, W_j)$ and let $\hat{f}_j := \mathbb{1}_{(-\infty,\hat{Z}_j]} = \mathbb{1}_{(-\infty,\hat{\varepsilon}_j]\times(-\infty,W_j]}$. To prove (F.1), it suffices to show that

$$\lim_{n\to\infty} \mathrm{Pr}(\cup_{j=1}^{n}\{\mathrm{d}(\hat{f}_j, f_j; \hat{f}_j^{\mathrm{r}}, f_j^{\mathrm{r}}) \leq (c\epsilon)^{1/2} \ \& \ (Z_j, Z_j^{\mathrm{r}}) \in \mathcal{S}_\pi \times \tilde{\mathcal{S}}_\pi\}) = 1. \tag{F.7}$$

This is because (F.6), (F.7), and the triangle inequality imply that the event

$$K_n := \{\mathrm{d}(\hat{f}_j, f_\pi; \hat{f}_j^{\mathrm{r}}, \tilde{f}_\pi) < 2(c\epsilon)^{1/2} \text{ for some } j \leq n\}$$

holds w.p.a.1. Moreover,

$$\mathrm{d}(\hat{f}_j, f; \hat{f}_j^{\mathrm{r}}, f^{\mathrm{r}}) \leq \mathrm{d}(\hat{f}_j, f_\pi; \hat{f}_j^{\mathrm{r}}, \tilde{f}_\pi) + (c\epsilon)^{1/2} \qquad \text{(by (F.2))}$$

$$< 3(c\epsilon)^{1/2} \text{ for some } j \leq n \qquad \text{(if } K_n \text{ is true.)}$$

Therefore, since K_n holds w.p.a.1,

$$1 = \lim_{n\to\infty} \Pr(K_n) \leq \lim_{n\to\infty} \Pr(\mathrm{d}(\hat{f}_j, f; \hat{f}_j^{\mathrm{r}}, f^{\mathrm{r}}) < 3(c\epsilon)^{1/2} \text{ for some } j \leq n)$$

and (F.1) follows because ϵ was arbitrary.

So we now prove (F.7). Since $\hat{\varepsilon}_j = \varepsilon_j - \Delta(X_j, \hat{\theta}, \theta_0)$ and

$$|\hat{f}_j - f_j| = |\mathbb{1}_{(-\infty, \hat{Z}_j]} - \mathbb{1}_{(-\infty, Z_j]}| \leq |\mathbb{1}_{(-\infty, \hat{\varepsilon}_j]} - \mathbb{1}_{(-\infty, \varepsilon_j]}| = \mathbb{1}_{(\hat{\varepsilon}_j \wedge \varepsilon_j, \hat{\varepsilon}_j \vee \varepsilon_j]},$$

we have that

$$\int |\hat{f}_j - f_j|^2 \, d\mathbb{P}_{\varepsilon|X=x, W=w} \leq \int_{\hat{\varepsilon}_j \wedge \varepsilon_j}^{\hat{\varepsilon}_j \vee \varepsilon_j} d\mathbb{P}_{\varepsilon|X=x, W=w}$$

$$\leq \zeta(x, w)|\Delta(X_j, \hat{\theta}, \theta_0)| \qquad \text{(by Ass. 3.2(ii))}$$

$$\leq \zeta(x, w)(\|\dot{\mu}_0(X_j)\| \, \|\hat{\theta} - \theta_0\| + |\rho(X_j, \hat{\theta}, \theta_0)|) \qquad \text{(by Ass. 3.1)}$$

$$\leq \zeta(x, w)(\|\dot{\mu}_0(X_j)\| \, \|\hat{\theta} - \theta_0\| + \sup_{\tilde{\theta} \in B(\theta_0, \|\hat{\theta} - \theta_0\|)} |\rho(X_j, \tilde{\theta}, \theta_0)|).$$

Similarly,

$$|\hat{f}_j^{\mathrm{r}} - f_j^{\mathrm{r}}| = |\mathbb{1}_{[-\hat{\varepsilon}_j, \infty) \times (-\infty, W_j]} - \mathbb{1}_{[-\varepsilon_j, \infty) \times (-\infty, W_j]}| \leq |\mathbb{1}_{[-\hat{\varepsilon}_j, \infty)} - \mathbb{1}_{[-\varepsilon_j, \infty)}| = \mathbb{1}_{[-\hat{\varepsilon}_j \wedge -\varepsilon_j, -\hat{\varepsilon}_j \vee -\varepsilon_j)}$$

implies that

$$\int |\hat{f}_j^{\mathrm{r}} - f_j^{\mathrm{r}}|^2 \, d\mathbb{P}_{\varepsilon|X=x, W=w} \leq \zeta(x, w)(\|\dot{\mu}_0(X_j)\| \, \|\hat{\theta} - \theta_0\| + \sup_{\tilde{\theta} \in B(\theta_0, \|\hat{\theta} - \theta_0\|)} |\rho(X_j, \tilde{\theta}, \theta_0)|).$$

Hence,

$$\mathrm{d}^2(\hat{f}_j, f_j; \hat{f}_j^{\mathrm{r}}, f_j^{\mathrm{r}}) \leq \|\zeta\|_{2, \mathbb{P}_{X, W}}(\|\dot{\mu}_0(X_j)\| \, \|\hat{\theta} - \theta_0\| + \sup_{\tilde{\theta} \in B(\theta_0, \|\hat{\theta} - \theta_0\|)} |\rho(X_j, \tilde{\theta}, \theta_0)|). \qquad \text{(F.8)}$$

Consequently, for each $r, \eta > 0$,

$$\{\mathrm{d}^2(\hat{f}_j, f_j; \hat{f}_j^{\mathrm{r}}, f_j^{\mathrm{r}}) > r\} \cap \{\|\hat{\theta} - \theta_0\| < \eta\}$$

$$\subset \{\|\zeta\|_{2, \mathbb{P}_{X, W}}(\|\dot{\mu}_0(X_j)\| \eta + \sup_{\tilde{\theta} \in \overline{B(\theta_0, \eta)}} |\rho(X_j, \tilde{\theta}, \theta_0)|) > r\},$$

which, by (F.8), implies that for each j,

$$(\{\mathrm{d}^2(\hat{f}_j, f_j; \hat{f}_j^{\mathrm{r}}, f_j^{\mathrm{r}}) > r\} \cup \{(Z_j, Z_j^{\mathrm{r}}) \notin \mathcal{S}_\pi \times \tilde{\mathcal{S}}_\pi\}) \cap \{\|\hat{\theta} - \theta_0\| < \eta\}$$

$$\subset \{\|\zeta\|_{2, \mathbb{P}_{X, W}}(\|\dot{\mu}_0(X_j)\| \eta + \sup_{\tilde{\theta} \in \overline{B(\theta_0, \eta)}} |\rho(X_j, \tilde{\theta}, \theta_0)|) > r\} \cup \{(Z_j, Z_j^{\mathrm{r}}) \notin \mathcal{S}_\pi \times \tilde{\mathcal{S}}_\pi\}.$$

Hence, we have that

$$\cap_{j=1}^{n}\left(\{\mathrm{d}^{2}(\hat{f}_{j},f_{j};\hat{f}_{j}^{\mathrm{r}},f_{j}^{\mathrm{r}})>r\}\cup\{(Z_{j},Z_{j}^{\mathrm{r}})\notin\mathcal{S}_{\pi}\times\tilde{\mathcal{S}}_{\pi}\}\right)\cap\{\|\hat{\theta}-\theta_{0}\|<\eta\}$$
$$\subset\cap_{j=1}^{n}\{\|\zeta\|_{2,\mathbb{P}_{X,W}}(\|\dot{\mu}_{0}(X_{j})\|\,\eta+\sup_{\tilde{\theta}\in\overline{B(\theta_{0},\eta)}}|\rho(X_{j},\tilde{\theta},\theta_{0})|)>r\}\cup\{(Z_{j},Z_{j}^{\mathrm{r}})\notin\mathcal{S}_{\pi}\times\tilde{\mathcal{S}}_{\pi}\}.$$

Therefore, since the observations are iid and $p:=\Pr((Z_{j},Z_{j}^{\mathrm{r}})\in\mathcal{S}_{\pi}\times\tilde{\mathcal{S}}_{\pi})>0$ for all j,

$$\Pr(\cap_{j=1}^{n}(\{\mathrm{d}^{2}(\hat{f}_{j},f_{j};\hat{f}_{j}^{\mathrm{r}},f_{j}^{\mathrm{r}})>r\}\cup\{(Z_{j},Z_{j}^{\mathrm{r}})\notin\mathcal{S}_{\pi}\times\tilde{\mathcal{S}}_{\pi}\})\cap\{\|\hat{\theta}-\theta_{0}\|<\eta\})$$
$$\leq(\Pr(\|\zeta\|_{2,\mathbb{P}_{X,W}}(\|\dot{\mu}_{0}(X_{1})\|\,\eta+\sup_{\tilde{\theta}\in\overline{B(\theta_{0},\eta)}}|\rho(X_{1},\tilde{\theta},\theta_{0})|)>r)+1-p)^{n}$$
$$\leq(r^{-1}\|\zeta\|_{2,\mathbb{P}_{X,W}}(\mathbb{E}\|\dot{\mu}_{0}(X_{1})\|\,\eta+\mathbb{E}\sup_{\tilde{\theta}\in\overline{B(\theta_{0},\eta)}}|\rho(X_{1},\tilde{\theta},\theta_{0})|)+1-p)^{n}$$
$$\leq(r^{-1}\|\zeta\|_{2,\mathbb{P}_{X,W}}(\|(\dot{\mu}_{0}'\dot{\mu}_{0})^{1/2}\|_{2,\mathbb{P}_{X}}+1)\eta+1-p)^{n},$$

where the second inequality is by Chebychev and the third by Jensen and Assumption 3.1. Let $r:=c\epsilon$ and choose η small enough so that $r^{-1}\|\zeta\|_{2,\mathbb{P}_{X,W}}(\|(\dot{\mu}_{0}'\dot{\mu}_{0})^{1/2}\|_{2,\mathbb{P}_{X}}+1)\eta+1-p<1$. Then,

$$\sum_{n=1}^{\infty}\Pr(\cap_{j=1}^{n}(\{\mathrm{d}^{2}(\hat{f}_{j},f_{j};\hat{f}_{j}^{\mathrm{r}},f_{j}^{\mathrm{r}})>c\epsilon\}\cup\{(Z_{j},Z_{j}^{\mathrm{r}})\notin\mathcal{S}_{\pi}\times\tilde{\mathcal{S}}_{\pi}\})\cap\{\|\hat{\theta}-\theta_{0}\|<\eta\})<\infty.$$

Hence, by Borel-Cantelli,

$$\Pr(\cap_{j=1}^{n}(\{\mathrm{d}^{2}(\hat{f}_{j},f_{j};\hat{f}_{j}^{\mathrm{r}},f_{j}^{\mathrm{r}})>c\epsilon\}\cup\{(Z_{j},Z_{j}^{\mathrm{r}})\notin\mathcal{S}_{\pi}\times\tilde{\mathcal{S}}_{\pi}\})\cap\{\|\hat{\theta}-\theta_{0}\|<\eta\}\text{ i.o.})=0$$
$$\Longleftrightarrow$$
$$\Pr(\cup_{j=1}^{n}(\{\mathrm{d}^{2}(\hat{f}_{j},f_{j};\hat{f}_{j}^{\mathrm{r}},f_{j}^{\mathrm{r}})\leq c\epsilon\}\cap\{(Z_{j},Z_{j}^{\mathrm{r}})\in\mathcal{S}_{\pi}\times\tilde{\mathcal{S}}_{\pi}\})\cup\{\|\hat{\theta}-\theta_{0}\|\geq\eta\}\text{ a.b.f.m }n)=1,$$

where "a.b.f.m n" is short for "all but finitely many n." Therefore, since $\lim_{n\to\infty}\Pr(\|\hat{\theta}-\theta_{0}\|\geq\eta)=0$ by Assumption 3.4, it follows that

$$\lim_{n\to\infty}\Pr(\cup_{j=1}^{n}\{\mathrm{d}(\hat{f}_{j},f_{j};\hat{f}_{j}^{\mathrm{r}},f_{j}^{\mathrm{r}})\leq(c\epsilon)^{1/2}\ \&\ (Z_{j},Z_{j}^{\mathrm{r}})\in\mathcal{S}_{\pi}\times\tilde{\mathcal{S}}_{\pi}\})=1.\qquad\square$$

References

AHMAD, I. A., AND Q. LI (1997): "Testing symmetry of an unknown density function by kernel method," *Journal of Nonparametric Statistics*, 7, 279–293.

ANDREWS, D. W. K. (1997): "A conditional Kolmogorov test," *Econometrica*, 65, 1097–1128.

BAI, J., AND S. NG (2001): "A consistent test for conditional symmetry of time series," *Journal of Econometrics*, 103, 225–258.

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

CARD, D., AND D. HYSLOP (1996): "Does inflation "grease the wheels of the labor market"?," NBER Working Paper 5538.

CHAY, K., AND J. POWELL (2001): "Semiparametric censored regression models," *Journal of Economic Perspectives*, 15, 29–42.

CHRISTOFIDES, L. N., AND T. STENGOS (2001): "A non-parametric test of the symmetry of PSID wage-change distributions," *Economics Letters*, 71, 363–368.

DAVYDOV, Y., M. LIFSHITS, AND N. SMORODINA (1998): *Local properties of distributions of stochastic functionals.* American Mathematical Society.

DELGADO, M. A., M. A. DOMÍNGUEZ, AND P. LAVERGNE (2006): "Consistent tests of conditional moment restrictions," *Annales d'Economie et de Statistique*, 81, 33–67.

DELGADO, M. A., AND J. C. ESCANCIANO (2007): "Nonparametric tests for conditional symmetry in dynamic models," *Journal of Econometrics*, 141, 652–682.

FAN, Y., AND R. GENCAY (1995): "A consistent nonparametric test of symmetry in linear regression models," *Journal of the American Statistical Association*, 90, 551–557.

HÁJEK, J. (1972): "Local asymptotic minimax and admissibility in estimation," in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 175–194. University of California Press, Berkeley.

HANSEN, B. E. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica*, 64, 413–430.

JACOBY, H. G., R. MURGAI, AND S. U. REHMAN (2004): "Monopoly power and distribution in fragmented markets: The case of groundwater," *Review of Economic Studies*, 71, 783–808.

KHMALADZE, E. V. (1993): "Goodness of fit problem and scanning innovation martingales," *Annals of Statistics*, 21(2), 798–829.

KHMALADZE, E. V., AND H. KOUL (2004): "Martingale transforms goodness-of-fit tests in regression models," *Annals of Statistics*, 32(3), 995–1034.

KOPCZUK, W. (2007): "Bequest and tax planning: Evidence from estate tax returns," *Quarterly Journal of Economics*, 122, 1801–1854.

LEVITT, S. (1996): "How do senators vote? Disentangling the role of voter preferences, party affiliation, and senator ideology," *The American Economic Review*, 86, 425–441.

NEUMEYER, N., AND H. DETTE (2007): "Testing for symmetric error distribution in nonparametric regression models," *Statistica Sinica*, 17, 775–795.

NEUMEYER, N., H. DETTE, AND E.-R. NAGEL (2005): "A note on testing symmetry of the error distribution in linear regression models," *Nonparametric Statistics*, 17, 697–715.

NEWEY, W. K. (1988): "Efficient estimation of Tobit models under conditional symmetry," in *Nonparametric and semiparametric methods in econometrics and statistics. Proceedings of the fifth international symposium in Economic Theory and Econometrics*, ed. by W. A. Barnett, J. Powell, and G. Tauchen, pp. 291–336. Cambridge University Press.

NEWEY, W. K., AND J. L. POWELL (1987): "Assymetric least squares estimation and testing," *Econometrica*, 55, 819–847.

Powell, J. L. (1986a): "Censored regression quantiles," *Journal of Econometrics*, 32, 143–155.

——— (1986b): "Symmetrically trimmed least squares estimation for tobit models," *Econometrica*, 54, 1435–1460.

Randles, R. H., and D. A. Wolfe (1979): *Introduction to the theory of nonparametric statistics*. John Wiley & Sons.

Stein, E. (1970): *Singular integrals and differentiability properties of functions*. Princeton University Press.

Stengos, T., and X. Wu (2004): "Information-theoretic distribution tests with application to symmetry and normality," `http://ssrn.com/abstract=512902`.

Su, J. Q., and L. Wei (1991): "A lack-of-fit test for the mean function in a generalized linear model," *Journal of the American Statistical Association*, 86, 420–426.

van der Vaart, A. W. (1998): *Asymptotic statistics*. Cambridge university press.

van der Vaart, A. W., and J. A. Wellner (1996): *Weak convergence and empirical processes: With applications to statistics*. Springer-Verlag.

——— (2007): "Empirical processes indexed by estimated functions," in *Asymptotics: Particles, Processes and Inverse Problems*, vol. 55 of *IMS Lecture Notes-Monograph Series*, pp. 234–252.

Zheng, J. X. (1998): "Consistent specification testing for conditional symmetry," *Econometric Theory*, 14, 139–149.

Department of Economics, University of Connecticut, Storrs, CT 06269.
E-mail address: `tao.chen@uconn.edu`

Department of Economics, University of Connecticut, Storrs, CT 06269, USA and Department of Economics, Universität Mannheim, 68131 Mannheim, Germany.
E-mail address: `gtxmwtnt@gmail.com`