FACULTY OF LAW, ECONOMICS AND FINANCE

CREA Discussion Paper 2014-04

Center for Research in Economic Analysis
University of Luxembourg

A simple consistent test of conditional symmetry in symmetrically trimmed tobit models

available online : http://wwwfr.uni.lu/recherche/fdef/crea/publications2/discussion_papers

Tao Chen, Université of Waterloo, Ontario, Canada
Gautam Tripathi, CREA, Université du Luxembourg

February, 2013

For editorial correspondence, please contact: crea@uni.lu
University of Luxembourg
Faculty of Law, Economics and Finance
162A, avenue de la Faïencerie
L-1511 Luxembourg

UNIVERSITÉ DU LUXEMBOURG

A SIMPLE CONSISTENT TEST OF CONDITIONAL SYMMETRY IN SYMMETRICALLY TRIMMED TOBIT MODELS

TAO CHEN AND GAUTAM TRIPATHI

Abstract. We propose a "weighted and sample-size adjusted" Kolmogorov-Smirnov type statistic to test the assumption of conditional symmetry maintained in the symmetrically trimmed least-squares (STLS) approach of Powell (1986b), which is widely used to estimate censored or truncated regression models without making distributional assumptions. The statistic proposed here is consistent and computationally easy to implement because, unlike traditional Kolmogorov-Smirnov statistics, it is not optimized over an uncountable set. Moreover, it does not require any nonparametric smoothing, although we test the validity of a conditional feature. We also propose a bootstrap procedure to obtain the p-values and critical values that are required to carry out the test in practical applications. Results from a simulation study suggest that our test can work very well even in small to moderately sized samples. As an empirical illustration, we apply our test to two datasets that have been used in the literature to estimate censored regression models using Powell's STLS approach, to check whether the assumption of conditional symmetry is supported by these datasets.

1. introduction

In a seminal paper, Powell (1986b) showed how to identify and estimate censored or truncated regression models, commonly referred to as "censored tobit" or "truncated tobit" models, respectively, in the econometrics literature, 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!). Most empirical applications of Powell's "symmetric trimming approach" that we are aware of, e.g., Levitt (1996), Chay and Powell (2001), Jolliffe (2002), and Jacoby, Murgai, and Rehman (2004), who estimate censored regression models, and Kopczuk (2007), who estimates a truncated regression model, to name a few, simply assume the aforementioned conditional symmetry condition. Some empirical studies, e.g., those in Newey (1987, Section 5) and Lee (1995, Section 4), do report a specification test for conditional symmetry, although the test statistics they employ cannot detect all deviations from conditional symmetry, regardless of the sample size used to construct the test statistics. A failure to reject conditional symmetry thus reveals nothing about the shape of the conditional distribution, even if the sample size is infinitely large. However, since the assumption of conditional symmetry is necessary for Powell's symmetric trimming approach to go through, it

Source file: powell-symm-2a.tex. Compiled: 17[th] February, 2014 at 10:02am.

is important to have a test that can detect, with probability approaching one (w.p.a.1) in large enough samples, any deviation from conditional symmetry. Such a test can be used to determine whether the conditional symmetry assumption maintained in Powell (1986b) is supported by a large enough dataset or not.

Although there has been continued interest in testing conditional symmetry in linear and nonlinear regression models, cf., e.g., Fan and Gencay (1995), Ahmad and Li (1997), Zheng (1998), Bai and Ng (2001), Hyndman and Yao (2002), Neumeyer, Dette, and Nagel (2005), Delgado and Escanciano (2007), Neumeyer and Dette (2007), Chen and Tripathi (2013), and the references therein, not much attention seems to have been paid towards testing the assumption of conditional symmetry maintained in Powell (1986b).[1] This constitutes a big gap in the literature because the assumption of conditional symmetry underlying Powell's symmetrically trimmed least-squares (STLS) estimators (widely used in applications for estimating censored tobit or truncated tobit models, as mentioned earlier) is necessary for the consistency and asymptotic normality of the STLS estimators as well as for the identification of the parameters these estimators are supposed to be estimating. The objective of our paper is to fill this gap.

In Section 4 of his 1987 paper cited earlier, Newey proposed a test of the conditional symmetry assumption in Powell's STLS approach for estimating censored-tobit models. However, since the Hausman type statistic proposed by him only tests some of the moment conditions implied by conditional symmetry, Newey's test is not consistent against all deviations from conditional symmetry and so has zero power to detect certain alternatives. This is confirmed by the simulation evidence in Section 6. Similarly, Lee (1995, p. 193) also contains a test based on testing a finite number of moment conditions implied by conditional symmetry, although Lee's moment conditions are different from those proposed by Newey. There is also a brief mention, but no formal test, in Powell (1986a, p. 155) suggesting a quantile matching approach to test for conditional symmetry in censored regression models.

In this paper, we propose a Kolmogorov-Smirnov type statistic to test the assumption of conditional symmetry maintained in Powell's STLS approach. The statistic proposed here, extending recent work by Chen and Tripathi (2013), henceforth abbreviated as CT, is consistent, computationally simple to implement because, unlike traditional Kolmogorov-Smirnov statistics, it is not based on optimization over an uncountable set, and it does not require any

[1]The results in the previously cited papers are not directly applicable for testing the assumption of conditional symmetry maintained in Powell's approach because of a fundamental difference between the model errors for trimmed and non-trimmed regression models. Namely, whereas the model errors in the linear or nonlinear regression models considered in the aforementioned papers are assumed to be continuously distributed with full support on $\mathbb{R}$, the residuals in trimmed regression models are not continuously distributed with full support. For instance, the "symmetrically censored residual" that forms the basis for Powell's approach for estimating censored regression models has compact support with mass-points at the boundary and the support depends upon the parameter of interest (cf. Section 2 for details).

nonparametric smoothing although the test is for the validity of a conditional feature. Since the limiting distribution of the test statistic turns out to be non-standard and non-pivotal, we also propose a bootstrap procedure to obtain the p-values and critical values that are required to carry out the test in practical applications. Our bootstrap approach is a variation of the well known wild bootstrap procedure that takes symmetric trimming into account. Results from a simulation study suggest that our test can work very well even in small to moderately sized samples. As an empirical illustration, we obtain some interesting findings by applying our test to two datasets, that have been used in the literature to estimate censored regression models using Powell's STLS approach, to check whether the assumption of conditional symmetry is supported by these datasets.

The remainder of the paper is organized as follows. In Section 2, we describe Powell's STLS approach for estimating censored regression models. In Section 3, we demonstrate how the symmetrically censored residuals can be used to consistently test for the conditional symmetry of the model error in censored regression models, outline the large sample properties of the proposed test statistic, and show how to obtain its critical values by employing the bootstrap. Section 4 shows how our statistic can be used for testing conditional symmetry in truncated regression models, and Section 5 describes how to handle models with a general, i.e., not necessarily zero, censoring or truncation threshold. Finite sample properties of the proposed statistic for testing conditional symmetry in censored and truncated regression models are examined in Section 6, the empirical applications are in Section 7, and Section 8 concludes the paper. Technical details are in the appendix.

2. Powell's symmetric trimming approach for censored regression models

Consider the linear regression model

$$Y_0 = X'\theta_0 + U_0, \tag{2.1}$$

where the scalar response Y_0 is latent[2] and X is a (column) vector of explanatory variables. Conditional on X, the model error U_0 is assumed to be continuously distributed on $\mathbb{R}$ with full support, meaning that $\mathrm{Law}(U_0|X)$ has Lebesgue density $p_{U_0|X}$ that is positive on $\mathbb{R}$.

Instead of observing Y_0, the researcher only observes its left censored (by 0) version

$$Y := \max(Y_0, 0) := \begin{cases} Y_0 & \text{if } Y_0 > 0 \\ 0 & \text{if } Y_0 \le 0 \end{cases} = \begin{cases} X'\theta_0 + U_0 & \text{if } X'\theta_0 + U_0 > 0 \\ 0 & \text{if } X'\theta_0 + U_0 \le 0. \end{cases} \tag{2.2}$$

We focus on left censoring because in our simulation experiments we replicate Powell's design. Extension to right censoring follows mutatis mutandis.

[2] Powell uses a "$*$" superscript to denote latent variables, but we avoid this because in our paper a "$*$" superscript indicates "bootstrapped" random variables, cf. Section 3.3.

Powell (1986b) showed that if the model error U_0 in the latent model (2.1) is symmetrically distributed about the origin conditional on the regressors, i.e., if

$$H_0: \quad U_0|X \stackrel{\mathrm{d}}{=} -U_0|X$$

is true, where "$\stackrel{\mathrm{d}}{=}$" is shorthand for "equal in distribution", 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.

To understand Powell's symmetric trimming approach, let $U := Y - X'\theta_0$ be the residual from the censored regression. Hence,

$$(2.2) \iff U = \max(U_0, -X'\theta_0) = \begin{cases} U_0 & \text{if } U_0 > -X'\theta_0 \\ -X'\theta_0 & \text{if } U_0 \leq -X'\theta_0. \end{cases}$$

Given X, the random variable U takes values in the interval $[-X'\theta_0, \infty)$ with a mass-point at $-X'\theta_0$, i.e., $\Pr(U = -X'\theta_0|X) > 0$.

Following Powell, define V to be the version of U right-censored at $X'\theta_0$, i.e.,

$$V := \min(U, X'\theta_0) := \begin{cases} U & \text{if } U < X'\theta_0 \\ X'\theta_0 & \text{if } U \geq X'\theta_0. \end{cases}$$

The definition of U implies that

$$V = \begin{cases} X'\theta_0 & \text{if } U_0 \geq X'\theta_0 \\ U_0 & \text{if } -X'\theta_0 < U_0 < X'\theta_0 \\ -X'\theta_0 & \text{if } U_0 \leq -X'\theta_0. \end{cases} \tag{2.3}$$

Given X, the "symmetrically censored residual" V only takes values in the interval $[-X'\theta_0, X'\theta_0]$, i.e., $\mathrm{Law}(V|X)$ has compact support $[-X'\theta_0, X'\theta_0]$, with mass-points at $\pm X'\theta_0$. If $X'\theta_0 \leq 0$, then either $[-X'\theta_0, X'\theta_0]$ is an empty interval or $V = 0$, neither of which gives any information about the distribution of U_0. Therefore, for $\mathrm{Law}(V|X)$ to be informative about $\mathrm{Law}(U_0|X)$, it is necessary to assume that $X'\theta_0 > 0$ (Powell, 1986b, p. 1440).

Let $x \in \mathrm{supp}(X)$ be such that $x'\theta_0 > 0$. From (2.3), it is clear that the conditional density function of $V|X = x$ is given by[3]

$$\begin{aligned} \mathrm{pdf}_{V|X=x}(v) := p_{U_0|X=x}(v)\mathbb{1}_{(-x'\theta_0, x'\theta_0)}(v) + \Pr(U_0 \leq -x'\theta_0|X = x)\mathbb{1}_{\{-x'\theta_0\}}(v) \\ + \Pr(U_0 \geq x'\theta_0|X = x)\mathbb{1}_{\{x'\theta_0\}}(v), \qquad v \in \mathbb{R}, \end{aligned} \tag{2.4}$$

[3]Since the conditional distribution of $V|X = x$ has mass-points at $\pm x'\theta_0$, the function $v \mapsto \mathrm{pdf}_{V|X=x}(v)$ is a density with respect to a dominating measure that is a mixture of the Lebesgue measure on $\mathbb{R}$ and the counting measure on $\{-x'\theta_0, x'\theta_0\}$. In contrast, $v \mapsto p_{U_0|X=x}(v)$ is a density with respect to the Lebesgue measure because, given X, the error U_0 in the latent model is assumed to be continuously distributed on $\mathbb{R}$.

where $\mathbb{1}_A$ is the indicator function of the set A, i.e., $\mathbb{1}_A(a) = 1$ if $a \in A$, zero otherwise. It is immediate from (2.4), in fact from (2.3) itself, that $\mathrm{pdf}_{V|X=x}(v) = p_{U_0|X=x}(v)$ for all $v \in (-x'\theta_0, x'\theta_0)$, i.e.,

$$V|X \overset{\mathrm{d}}{=} U_0|X \text{ on } (-X'\theta_0, X'\theta_0). \tag{2.5}$$

Moreover, $v \mapsto \mathrm{pdf}_{V|X=x}(v)$ is an even function on its support $[-x'\theta_0, x'\theta_0]$ if and only if $p_{U_0|X=x}(v) = p_{U_0|X=x}(-v)$ for all $v \in (-x'\theta_0, x'\theta_0)$ and $\Pr(U_0 \leq -x'\theta_0|X = x) = \Pr(U_0 \geq x'\theta_0|X = x)$. It follows that $V|X$ is symmetrically distributed about the origin on its support $[-X'\theta_0, X'\theta_0]$ if and only if $U_0|X$ is symmetrically distributed on $[-X'\theta_0, X'\theta_0]$, i.e.,

$$V|X \overset{\mathrm{d}}{=} -V|X \iff U_0|X \overset{\mathrm{d}}{=} -U_0|X \text{ on } [-X'\theta_0, X'\theta_0]. \tag{2.6}$$

This equivalence can be used to estimate θ_0. Indeed, since

$$H_0 \implies \mathbb{E}[V\mathbb{1}_{(0,\infty)}(X'\theta_0)|X] = 0 \text{ w.p.1} \implies \mathbb{E}[XV\mathbb{1}_{(0,\infty)}(X'\theta_0)] = 0,$$

where the first implication follows by (2.6), Powell used the last moment condition to estimate θ_0. More importantly for us, we can employ the symmetrically censored residual V to test the conditional symmetry of U_0. However, the equivalence in (2.6) makes it clear that V can only be used to test the validity of H_0 on the support of $\mathrm{Law}(V|X)$. In other words, any test that uses the symmetrically censored residuals to check the conditional symmetry of $U_0|X$ will be able to detect departures from conditional symmetry only on the interval $[-X'\theta_0, X'\theta_0]$.

Therefore, we henceforth focus on testing H_0 against the alternative hypothesis that

$$\tilde{H}_0 : \quad U_0|X \overset{\mathrm{d}}{=} -U_0|X \text{ on } [-X'\theta_0, X'\theta_0]$$

is false. The expression for $\tilde{H}_0^{\complement}$, the negation of $\tilde{H}_0$, is given in the proof of Proposition 3.1.

3. THE TEST

3.1. Motivating the test statistic.

By (2.6),

$$\tilde{H}_0 \iff V|X \overset{\mathrm{d}}{=} -V|X \iff (V, X) \overset{\mathrm{d}}{=} (-V, X),$$

where the second equivalence can be shown as in the proof of equation (2.1) of CT. Therefore, H_0 can be tested by contrasting the empirical distributions of (V, X) and $(-V, X)$. However, recalling the discussion after (2.3), $\mathrm{Law}(V|X)$ is informative about $\mathrm{Law}(U_0|X)$ only for those X for which $X'\theta_0 > 0$. Therefore, the empirical distributions of (V, X) and $(-V, X)$ should be based on only the informative V, i.e., those corresponding to $X'\theta_0 > 0$. This is done as follows.

First, use the given sample $\mathscr{Y} \times \mathscr{X} := \{(Y_j, X_j) : j = 1, \ldots, n\}$ to obtain Powell's SCLS estimator $\hat{\theta}$. Next, extract from $\mathscr{Y} \times \mathscr{X}$ those (Y, X) observation pairs for which $X'\hat{\theta} > 0$.[4]

[4]Note that $\Pr(X'\theta_0 > 0)$ may be positive even if the unconditional probability of censoring $\Pr(Y_0 \leq 0) = \Pr(Y = 0)$ is larger than $1/2$ (e.g., in the application in Section 7.2, the proportion of censored observations is more than 50%). Indeed, let $q(X) := \Pr(Y_0 \leq 0|X) = \Pr(U_0 \leq -X'\theta_0|X)$ be the conditional probability

This leads to the subsample $\hat{\mathcal{Y}} \times \mathcal{X} := \{(Y_j, X_j) : X_j'\hat{\theta} > 0,\ j = 1, \dots, \hat{N}\}$, where $\hat{N} := \sum_{j=1}^n \mathbb{1}_{(0,\infty)}(X_j'\hat{\theta})$ is the "effective" sample size. The subsample $\hat{\mathcal{Y}} \times \mathcal{X}$ is informative in the sense that the symmetrically censored residual $\hat{V}$ constructed using this subsample, cf. (3.1) below, has non-empty support $[-X'\hat{\theta}, X'\hat{\theta}]$ and is thus informative about the distribution of $U_0|X$. Use the $\hat{\mathcal{Y}} \times \mathcal{X}$ subsample to construct the set of observations $\hat{\mathcal{V}} \times \mathcal{X} := \{(\hat{V}_j, X_j) : j = 1, \dots, \hat{N}\}$ with

$$\hat{V}_j := \begin{cases} \hat{U}_j & \text{if } \hat{U}_j < X_j'\hat{\theta} \\ X_j'\hat{\theta} & \text{if } \hat{U}_j \geq X_j'\hat{\theta} \end{cases} \quad \& \quad \hat{U}_j := Y_j - X_j'\hat{\theta}, \qquad j = 1, \dots, \hat{N}. \tag{3.1}$$

The statistic for testing H_0 against the alternative that $\tilde{H}_0$ is false is then defined to be

$$\hat{\mathrm{T}}_{\max} := \hat{N}^{1/2}\hat{\mathrm{R}}_{\max}, \tag{3.2}$$

where

$$\hat{\mathrm{R}}_{\max} := \max_{(v,x)\in\hat{\mathcal{V}}\times\mathcal{X}} \left|\frac{1}{\hat{N}}\sum_{j=1}^{\hat{N}} \mathbb{1}_{(-\infty,v]\times(-\infty,x]}(\hat{V}_j, X_j) - \frac{1}{\hat{N}}\sum_{j=1}^{\hat{N}} \mathbb{1}_{(-\infty,v]\times(-\infty,x]}(-\hat{V}_j, X_j)\right| \tag{3.3}$$

is a measure of the distance between the empirical distributions of $(\hat{V}, X)$ and $(-\hat{V}, X)$ based on the "informative" subsample $\hat{\mathcal{V}} \times \mathcal{X}$.[5] The discrepancy measure $\hat{\mathrm{R}}_{\max}$ resembles the usual Kolmogorov-Smirnov contrast for testing whether two samples come from the same distribution although, unlike in traditional Kolmogorov-Smirnov type statistics, the scale-factor $\hat{N}$ in (3.2) is data dependent (implying that $\hat{\mathrm{T}}_{\max}$ is based on data dependent weights, cf. Section 3.2) and the optimization in (3.3) is over a finite set. The latter feature implies that $\hat{\mathrm{T}}_{\max}$ can be computed very quickly, unlike traditional Kolmogorov-Smirnov type statistics requiring optimization over an uncountable set which can be very demanding in practical applications.

There are three major differences between $\hat{\mathrm{T}}_{\max}$ and the test statistic proposed by CT. First, the scale-factor $\hat{N}^{1/2}$ in its definition is random, unlike the non-random scaling $n^{1/2}$ used in CT. To appreciate the second difference, notice that we can write

$$\hat{\mathrm{R}}_{\max} = \frac{1}{\hat{N}/n} \max_{(v,x)\in\hat{\mathscr{V}}\times\mathscr{X}} \Big|\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{(0,\infty)}(X_j'\hat{\theta})\mathbb{1}_{(-\infty,v]\times(-\infty,x]}(\hat{V}_j, X_j) \\ - \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{(0,\infty)}(X_j'\hat{\theta})\mathbb{1}_{(-\infty,v]\times(-\infty,x]}(-\hat{V}_j, X_j)\Big|,$$

---

that Y_0 is censored, given X. Then, assuming that $\mathrm{cdf}_{U_0|X}$ is strictly increasing on $\mathbb{R}$, $\Pr(X'\theta_0 > 0) = \Pr(\mathrm{cdf}_{U_0|X}(-X'\theta_0) < \mathrm{cdf}_{U_0|X}(0)) \overset{H_0}{=} \Pr(q(X) < 1/2)$. Therefore, $1/2 < \Pr(Y_0 \leq 0) = \mathbb{E}q(X)$ does not rule out $\Pr(X'\theta_0 > 0) > 0$, provided that $\Pr(q(X) \geq 1/2) < 1$. An analogous argument holds if $\tilde{H}_0$ is false, provided that θ_0 is suitably redefined in this case (cf. the discussion after (A.2)).

[5]The half-closed interval $(-\infty, v]$ in (3.3) is a subset of $\mathbb{R}$ because $\hat{V}_j$ is a scalar, whereas $(-\infty, x] := (-\infty, x^{(1)}] \times \dots \times (-\infty, x^{(\dim(X))}] \subset \mathbb{R}^{\dim(X)}$ because $X := (X^{(1)}, \dots, X^{(\dim(X))})_{\dim(X)\times 1}$ is a vector.

where $\hat{\mathscr{V}} \times \mathscr{X} := \{(\hat{V}_j, X_j) : j = 1, \ldots, n\}$ denotes the "full" sample of $(\hat{V}, X)$ observation pairs. Hence, unlike the statistic of CT, $\hat{\mathrm{T}}_{\max}$ is a "weighted" Kolmogorov-Smirnov type statistic because of the data-dependent weights $\mathbb{1}_{(0,\infty)}(X_1'\hat{\theta}), \ldots, \mathbb{1}_{(0,\infty)}(X_n'\hat{\theta})$ appearing in $\hat{\mathrm{R}}_{\max}$. That is why $\hat{\mathrm{T}}_{\max}$ is referred to as a "weighted and sample-size adjusted" statistic in the abstract. Finally, the residuals $\hat{V}$ in $\hat{\mathrm{R}}_{\max}$ differ from those in CT's statistic (denoted there by $\hat{\varepsilon}$) in that their population analog V is supported on a compact interval with mass-points at its boundary, whereas the population analog of $\hat{\varepsilon}$ (denoted by ε in CT) is assumed to be continuously distributed with full support on $\mathbb{R}$.

3.2. **Large sample properties.** Despite the differences between $\hat{\mathrm{T}}_{\max}$ and the statistic proposed by CT, the technical arguments in CT can be extended to show the following result under certain regularity conditions (cf. Appendix A for details).[6]

Proposition 3.1. *If H_0 is true, then $\hat{\mathrm{T}}_{\max}$ converges in distribution to a nonnegative random variable T_∞. If $\tilde{H}_0$ is false, then $\hat{\mathrm{T}}_{\max}$ converges to ∞ w.p.a.1.*

For $\alpha \in (0, 1)$, let c_α denote the $(1-\alpha)$ quantile of T_∞ under H_0. The null hypothesis H_0 is rejected in favor of $\tilde{H}_0^{\complement}$ at $100\alpha\%$ level of significance if $\hat{\mathrm{T}}^{\mathrm{obs}}_{\max} > c_\alpha$, where $\hat{\mathrm{T}}^{\mathrm{obs}}_{\max}$ is the observed value of $\hat{\mathrm{T}}_{\max}$. The test "Reject H_0 if $\hat{\mathrm{T}}^{\mathrm{obs}}_{\max} > c_\alpha$" has the correct level of significance because $\Pr(\hat{\mathrm{T}}_{\max} > c_\alpha; H_0) \to \alpha$ by Proposition 3.1. Moreover, $\Pr(\hat{\mathrm{T}}_{\max} > c_\alpha; \tilde{H}_0^{\complement}) \to 1$ implies that this test will detect w.p.a.1 any deviation from conditional symmetry on the interval $[-X'\theta_0, X'\theta_0]$. This is the sense in which our test is "consistent".

A brief word on interpreting the test outcomes. From Table 1, it is clear that the test will reject H_0 either because it makes a mistake, or because H_0 is false ((b) and (c) both imply that H_0 is false). Hence, after controlling the probability of making a type I mistake, i.e., the size of the test, a rejection of H_0 by the test can be attributed to H_0 being false. Similarly, the test will fail to reject H_0 either because $\tilde{H}_0$ is true ((a) and (c) both imply that $\tilde{H}_0$ is true, even though (c) cannot be detected by any test), or because it made a mistake. Since the test is consistent, i.e., the probability of making a type II mistake is arbitrarily small in large enough samples, a failure to reject H_0 by the test can be attributed to at least $\tilde{H}_0$ being true.

TABLE 1. Interpreting the test outcomes.

State of nature		Outcome	
		Test rejects H_0	Test fails to reject H_0
(a)	H_0 is true	type I mistake	correct decision
(b)	$\tilde{H}_0$ is false	correct decision	type II mistake
(c)	H_0 is false but $\tilde{H}_0$ is true	correct decision	(c) is not detectable

[6] Unless specified otherwise, all limits are taken as the sample size $n \to \infty$.

3.3. **Bootstrapping the critical values.** In practice, the test "Reject H_0 if $\hat{\mathrm{T}}_{\max}^{\mathrm{obs}} > c_\alpha$" cannot be implemented as stated because the quantiles of T_∞, hence, the critical value c_α, cannot be tabulated in advance.[7] Instead, as we show next, c_α can be estimated by using the bootstrap. The "bootstrapped critical value" $c_{\alpha,B}^*$ possesses two important properties. First, irrespective of whether H_0 is true or not, it is bounded in probability. Second, it converges to c_α whenever H_0 is true. These properties ensure that the "bootstrapped test" "Reject H_0 if $\hat{\mathrm{T}}_{\max}^{\mathrm{obs}} > c_{\alpha,B}^*$" is consistent and has size α. Indeed, since $\hat{\mathrm{T}}_{\max} \xrightarrow{\mathrm{p}} \infty$ if $\tilde{H}_0$ is false, the first property of $c_{\alpha,B}^*$ ensures that $\hat{\mathrm{T}}_{\max}$ will reject a false $\tilde{H}_0$ w.p.a.1; the second property ensures the correct size.

The use of the bootstrap to estimate critical values for specification tests has been investigated earlier in the literature, cf. the references in Section 4 of CT. In this section, we enlarge this literature by showing how to handle symmetrically trimmed tobit models.

The basic idea is to construct random variables $V_1^*, \ldots, V_n^*$ such that each V^* satisfies two properties: (i) $\mathrm{Law}(V^*|X) = \mathrm{Law}(-V^*|X)$ whether H_0 is true or not, and (ii) $\mathrm{Law}(V^*|X) = \mathrm{Law}(V|X)$ if H_0 is true. The statistic $\hat{\mathrm{T}}_{\max}$ applied to the bootstrap sample $\{(V_j^*, X_j) : j = 1, \ldots, n\}$ then yields its bootstrap analog $\hat{\mathrm{T}}_{\max}^*$.[8] Since conditional symmetry holds in the bootstrap universe by construction, irrespective of whether H_0 in the population is true or not, the statistic $\hat{\mathrm{T}}_{\max}^*$, which tests for conditional symmetry in the bootstrap sample, is well behaved even if H_0 is false. Of course, if H_0 is true then $V^*|X \overset{\mathrm{d}}{=} V|X$. Consequently, $\mathrm{Law}(\hat{\mathrm{T}}_{\max}^*; H_0) = \mathrm{Law}(\hat{\mathrm{T}}_{\max}; H_0)$ and both converge to $\mathrm{Law}(\mathrm{T}_\infty; H_0)$ as $n \to \infty$. In short, $\hat{\mathrm{T}}_{\max}^*$ is bounded in probability whether H_0 is true or not, and it converges in distribution to T_∞ if H_0 is true. Therefore, it makes sense to estimate the quantiles of $\mathrm{Law}(\mathrm{T}_\infty; H_0)$ by those of $\hat{\mathrm{T}}_{\max}^*$. This is straightforward to do by constructing, say, B bootstrap samples and calculating $\hat{\mathrm{T}}_{\max}^*$ for each sample. This yields B draws $\{\hat{\mathrm{T}}_{\max}^*(1), \ldots, \hat{\mathrm{T}}_{\max}^*(B)\}$ from the distribution of $\hat{\mathrm{T}}_{\max}^*$ whose $(1-\alpha)$th empirical quantile is a consistent estimator of c_α, the $(1-\alpha)$th quantile of $\mathrm{Law}(\mathrm{T}_\infty; H_0)$, as both $B, n \to \infty$.

To motivate how V^* is constructed, recall that a random variable R is said to have the Rademacher, or symmetric Bernoulli, distribution if R takes the values -1 and 1, each with probability 0.5. Let RU be the "Rademacherized" version of U, where U is the censored version of U_0. Note that $\mathrm{Law}(RU|X)$ has support $\mathbb{R}$ with mass-points at $\pm X'\theta_0$. Since censoring destroys information, the distribution of U_0 cannot be recovered, i.e., identified, from the distribution of its censored version U. However, the distribution of $U_0|X$ can be identified from the distribution of its Rademacherized censored version $RU|X$, provided $U_0|X$ is symmetrically

[7] The random variable T_∞ is a functional of a gaussian process whose distribution varies from application to application because it depends upon U_0, Y, and X, cf. the proof of Proposition 3.1.

[8] In practice, V^* has to be replaced by its estimator $\hat{V}^*$ in order to construct $\hat{\mathrm{T}}_{\max}^*$.

distributed. Indeed, as shown in the appendix,

$$H_0 \implies \text{pdf}_{RU|X=x}(u) = \begin{cases} 0.5 p_{U_0|X=x}(u) & \text{if } u \in (-\infty, -x'\theta_0) \cup (x'\theta_0, \infty) \\ p_{U_0|X=x}(u) & \text{if } u \in (-x'\theta_0, x'\theta_0) \\ 0.5 \Pr(U_0 \leq -x'\theta_0 | X = x) & \text{if } u \in \{-x'\theta_0, x'\theta_0\}. \end{cases} \tag{3.4}$$

Thus, $RU|X$ can be thought of as the bootstrap analog of $U_0|X$ under H_0. Hence, it is reasonable to define V^* as the symmetrically censored version of RU, i.e.,

$$V^* := \begin{cases} X'\theta_0 & \text{if } RU > X'\theta_0 \\ RU & \text{if } -X'\theta_0 < RU < X'\theta_0 \\ -X'\theta_0 & \text{if } RU < -X'\theta_0. \end{cases}$$

It is clear that $V^*|X \overset{\text{d}}{=} -V^*|X$ by construction whether H_0 is true or not. Moreover, when H_0 is true, on the interval $(-X'\theta_0, X'\theta_0)$ we have that $V^*|X \overset{\text{d}}{=} RU|X \overset{\text{d}}{=} U_0|X \overset{\text{d}}{=} V|X$, where the first "$\overset{\text{d}}{=}$" follows from the definition of V^*, the second from (3.4), and the last from (2.5). Therefore, V^* satisfies the properties mentioned earlier.

It is useful to note that V^* can be obtained in two steps in a more familiar format. First, let $Y_0^* := X'\theta_0 + RU$. Since, given X, RU is the bootstrap analog of U_0 under H_0, it follows that, given X, Y_0^* is the bootstrap analog of Y_0 under H_0. Following this reasoning, given X, $Y^* := \max(Y_0^*, 0)$, $U^* := Y^* - X'\theta_0 = \max(RU, -X'\theta_0)$, and $V^* = \min(U^*, X'\theta_0)$ are the bootstrap analogs of Y, U, and V, respectively, under H_0. As will soon be apparent, the knowledge that Y^*, U^*, and V^* are the bootstrap analogs of Y, U, and V, respectively, facilitates their construction.

As mentioned earlier, V^* has to be replaced by its estimator $\hat{V}^*$ in order to obtain $\hat{\text{T}}^*_{\text{max}}$. While constructing $\hat{V}^*$, it is important to keep in mind that since $\hat{V}^*$ is the bootstrap analog of $\hat{V}$ and randomness in $\hat{V}$ comes both from the data and the estimation uncertainty in $\hat{\theta}$, these sources of randomness should be replicated in $\hat{V}^*$ as well.

Therefore, in practice, the bootstrapped test is implemented as follows.

(i) Use (3.2) to calculate $\hat{\text{T}}^{\text{obs}}_{\text{max}}$.

(ii) Generate $R_1, \ldots, R_{\hat{N}} \overset{\text{iid}}{=}$ Rademacher. Use the subsample $\hat{\mathcal{Y}} \times \mathcal{X}$ to let $\hat{Y}_j^* := \max(X_j'\hat{\theta} + R_j\hat{U}_j, 0)$, $j = 1, \ldots, \hat{N}$, denote the analog of the censored outcome Y_j in the bootstrap universe. This yields the bootstrap sample $\{(\hat{Y}_j^*, X_j) : j = 1, \ldots, \hat{N}\}$.

(iii) Use the bootstrap sample and Powell's SCLS procedure to re-estimate θ_0. Denote this estimator by $\hat{\theta}^*$ and use it to obtain

$$\hat{U}_j^* := \hat{Y}_j^* - X_j'\hat{\theta}^*, \qquad \hat{V}_j^* := \begin{cases} \hat{U}_j^* & \text{if } \hat{U}_j^* < X_j'\hat{\theta}^* \\ X_j'\hat{\theta}^* & \text{if } \hat{U}_j^* \geq X_j'\hat{\theta}^* \end{cases}, \qquad j = 1, \ldots, \hat{N}.$$

Employing $\hat{\theta}^*$ to construct $\hat{V}^*$, the bootstrap analog of $\hat{V}$, replicates the estimation uncertainty of $\hat{\theta}$ in $\hat{V}$. Next, from $\{(\hat{V}_j^*, X_j) : j = 1, \ldots, \hat{N}\}$ extract those $(\hat{V}^*, X)$ pairs for which $X'\hat{\theta}^* > 0$. This leads to the subsample of size $\hat{N}^* := \sum_{j=1}^{\hat{N}} \mathbb{1}_{(0,\infty)}(X_j'\hat{\theta}^*)$, denoted by $\hat{\mathcal{V}}^* \times \mathcal{X} := \{(\hat{V}_j^*, X_j) : X_j'\hat{\theta}^* > 0,\ j = 1, \ldots, \hat{N}^*\}$, ensuring that only the $\hat{V}^*$ having non-empty support $[-X'\hat{\theta}^*, X'\hat{\theta}^*]$ are used to construct the bootstrap analog of $\hat{\mathrm{T}}_{\max}$.

(iv) Define $\hat{\mathrm{T}}^*_{\max} := \hat{N}^{*1/2}\hat{\mathrm{R}}^*_{\max}$ to be the bootstrap analog of $\hat{\mathrm{T}}_{\max}$ with

$$\hat{\mathrm{R}}^*_{\max} := \max_{(v,x)\in\hat{\mathcal{V}}^*\times\mathcal{X}} \Big|\frac{1}{\hat{N}^*}\sum_{j=1}^{\hat{N}^*} \mathbb{1}_{(-\infty,v]\times(-\infty,x]}(\hat{V}_j^*, X_j) - \frac{1}{\hat{N}^*}\sum_{j=1}^{\hat{N}^*} \mathbb{1}_{(-\infty,v]\times(-\infty,x]}(-\hat{V}_j^*, X_j)\Big|.$$

(v) Repeat (ii)–(iv) B times to get $\{\hat{\mathrm{T}}^*_{\max}(b) : b = 1, \ldots, B\}$, a sample of B observations from the distribution of $\hat{\mathrm{T}}^*_{\max}$. Find $c^*_{\alpha,B}$, the $(1-\alpha)$th empirical quantile of this sample.

(vi) Reject H_0 at $100\alpha\%$ level of significance if $\hat{\mathrm{T}}^{\mathrm{obs}}_{\max} > c^*_{\alpha,B}$. Alternatively, use the p-value given by $\sum_{b=1}^{B} \mathbb{1}_{(\hat{\mathrm{T}}^{\mathrm{obs}}_{\max},\infty)}(\hat{\mathrm{T}}^*_{\max}(b))/B$ to make the appropriate decision.

The asymptotic behavior of $\hat{\mathrm{T}}^*_{\max}$ is summarized in the following proposition.[9]

Proposition 3.2. *If H_0 is true, then $\hat{\mathrm{T}}^*_{\max}$ converges in distribution to T_∞ as $B, n \to \infty$. If $\tilde{H}_0$ is false, then $\hat{\mathrm{T}}^*_{\max}$ is bounded in probability.*

These properties of $\hat{\mathrm{T}}^*_{\max}$ ensure that the bootstrapped test "Reject H_0 in favor of $\tilde{H}_0^{\complement}$ at $100\alpha\%$ level of significance if $\hat{\mathrm{T}}^{\mathrm{obs}}_{\max} > c^*_{\alpha,B}$" has the correct size, and is consistent in the sense explained earlier, as $B, n \to \infty$.

4. Testing conditional symmetry in truncated regression models

In this section, we describe how $\hat{\mathrm{T}}_{\max}$ can be used for testing conditional symmetry in truncated regression models. The latent regression model is specified in (2.1). However, instead of observing Y_0, we now observe its truncated version

$$Y := \begin{cases} Y_0 & \text{if } Y_0 > 0 \\ \text{unobserved} & \text{otherwise.} \end{cases}$$

Consequently, the residual from the truncated regression is

$$U := Y - X'\theta_0 = \begin{cases} U_0 & \text{if } U_0 > -X'\theta_0 \\ \text{unobserved} & \text{otherwise.} \end{cases}$$

Powell (1986b) showed that if H_0 is true, then θ_0 is identified and the symmetrically truncated least-squares estimator $\hat{\theta}$, defined via equation 2.7 of his paper, converges almost surely to θ_0.

[9]The proof of Proposition 3.2 is very similar to the proof of Proposition 3.1 (compare, e.g., the proofs of Lemmas 3.3 and 4.3 in CT), and is therefore omitted.

Let V denote the retruncated, from the right by $X'\theta_0$, version of U, i.e.,

$$V := \begin{cases} U & \text{if } U < X'\theta_0 \\ \text{missing} & \text{otherwise} \end{cases} = \begin{cases} U_0 & \text{if } -X'\theta_0 < U_0 < X'\theta_0 \\ \text{missing} & \text{otherwise.} \end{cases}$$

As before, since the conditional distribution of $V|X$ has support $(-X'\theta_0, X'\theta_0)$, it is necessary to assume that $X'\theta_0$ is positive to ensure that the support is not empty. Notice that[10]

$$V|X \stackrel{\mathrm{d}}{=} -V|X \text{ on } (-X'\theta_0, X'\theta_0) \iff U_0|X \stackrel{\mathrm{d}}{=} -U_0|X \text{ on } (-X'\theta_0, X'\theta_0).$$

Again, it is clear the symmetrically truncated residuals can only be used to test H_0 against the alternative that $U_0|X$ is not symmetric on the interval $(-X'\theta_0, X'\theta_0)$.

We can use $\hat{\mathrm{T}}_{\max}$ to test H_0, although $\hat{\mathrm{T}}_{\max}$ needs to be slightly modified because we are now dealing with truncated, rather than censored, variables. The modification required is straightforward, and consists of replacing the censored variables in the definition of $\hat{\mathrm{T}}_{\max}$ with their truncated counterparts. In particular, we now let

$$\hat{V}_j := \begin{cases} \hat{U}_j & \text{if } \hat{U}_j < X_j'\hat{\theta} \\ \text{missing} & \text{otherwise} \end{cases}, \qquad j = 1, \dots, n,$$

denote the symmetrically truncated version of $\hat{U}_j := Y_j - X_j'\hat{\theta}$, where $\hat{\theta}$ is Powell's symmetrically truncated least-squares estimator.[11] Let $\hat{N} := \sum_{j=1}^{n} \mathbb{1}(\hat{V}_j \text{ is not missing and } X_j'\hat{\theta} > 0)$.[12] The subsample $\hat{\mathcal{V}} \times \mathcal{X} := \{(\hat{V}_j, X_j) : \hat{V}_j \text{ is not missing and } X_j'\hat{\theta} > 0,\ j = 1, \dots, \hat{N}\}$ is used to construct $\hat{\mathrm{R}}_{\max}$ and $\hat{\mathrm{T}}_{\max}$ as defined in (3.3) and (3.2), respectively.

To bootstrap the test, steps (i)–(vi) in Section 3.3 are followed with censored variables replaced by their truncated versions. Specifically, using the subsample $\hat{\mathcal{V}} \times \mathcal{X}$, let

$$\hat{Y}_j^* := \begin{cases} X_j'\hat{\theta} + R_j\hat{U}_j & \text{if } X_j'\hat{\theta} + R_j\hat{U}_j > 0 \\ \text{missing} & \text{otherwise} \end{cases}, \qquad j = 1, \dots, \hat{N}.$$

The bootstrap sample is the collection of the non-missing $\hat{Y}^*$ with their corresponding X, i.e., $\{(\hat{Y}_j^*, X_j) : \hat{Y}_j^* \text{ is not missing},\ j = 1, \dots, \hat{N}_1^*\}$, where $\hat{N}_1^* := \sum_{j=1}^{\hat{N}} \mathbb{1}(\hat{Y}_j^* \text{ is not missing})$. Let $\hat{\theta}^*$ denote Powell's symmetrically truncated least-squares estimator obtained using the bootstrap

[10]Indeed, let $r(x) := \int_{-x'\theta_0}^{x'\theta_0} \mathrm{pdf}_{U_0|X=x}(t)\,dt$. Then, since $\mathrm{pdf}_{V|X=x}(t) := \mathrm{pdf}_{U_0|X=x}(t)\mathbb{1}_{(-x'\theta_0, x'\theta_0)}(t)/r(x)$, $t \in \mathbb{R}$, the conditional density $\mathrm{pdf}_{V|X=x}$ is symmetric on the interval $(-x'\theta_0, x'\theta_0)$ if and only if $\mathrm{pdf}_{U_0|X=x}(t) = \mathrm{pdf}_{U_0|X=x}(-t)$ for all $t \in (-x'\theta_0, x'\theta_0)$.

[11]The algorithm of Santos Silva (2001) is designed to implement the SCLS estimator for censored regression models. Therefore, for the simulation results reported in Section 6.2, $\hat{\theta}$ was obtained by minimizing the objective function given in Powell (1986b, Equation 2.7).

[12]The indicator of event A is written as $\mathbb{1}(A)$, which takes the value 1 if A is true, zero otherwise.

sample, and define the bootstrap analog of $\hat{U}$ to be $\hat{U}_j^* := \hat{Y}_j^* - X_j'\hat{\theta}^*$, $j = 1, \ldots, \hat{N}_1^*$. The symmetrically truncated version of $\hat{U}_j^*$ is given by

$$\hat{V}_j^* := \begin{cases} \hat{U}_j^* & \text{if } \hat{U}_j^* < X_j'\hat{\theta}^* \\ \text{missing} & \text{otherwise} \end{cases}, \qquad j = 1, \ldots, \hat{N}_1^*.$$

Finally, let $\hat{\mathcal{V}}^* \times \mathcal{X} := \{(\hat{V}_j^*, X_j) : \hat{V}_j^* \text{ is not missing and } X_j'\hat{\theta}^* > 0,\ j = 1, \ldots, \hat{N}^*\}$, where $\hat{N}^* := \sum_{j=1}^{\hat{N}_1^*} \mathbb{1}(\hat{V}_j^* \text{ is not missing and } X_j'\hat{\theta}^* > 0)$. The $\hat{\mathcal{V}}^* \times \mathcal{X}$ subsample, which only contains those $\hat{V}^*$ having non-empty support $(-X'\hat{\theta}^*, X'\hat{\theta}^*)$, is used to construct $\hat{\mathrm{T}}_{\max}^*$ and estimate the critical values or p-values as described in Section 3.3.

5. Handling a general censoring or truncation threshold

Powell (1986b) only considers models where the censoring or truncation threshold is zero. However, it is not difficult to adapt Powell's treatment to the case when the threshold may be non-zero. In this section, we describe how our test can be applied if the censoring or truncation threshold is not zero. Throughout this section, we will denote the censoring or truncation threshold by a known constant c, which may or may not be zero. The definitions of the latent regression model and the test statistic $\hat{\mathrm{T}}_{\max}$ are not affected by c. As described below, the only change is in the manner in which the observed response Y, the symmetrically censored/truncated residual V, and the hypothesis $\tilde{H}_0$, are defined.

5.1. **Censoring.**

We observe the left censored (by c) version of Y_0, namely,

$$Y := \begin{cases} Y_0 & \text{if } Y_0 > c \\ c & \text{if } Y_0 \le c \end{cases} = \begin{cases} X'\theta_0 + U_0 & \text{if } X'\theta_0 + U_0 > c \\ c & \text{if } X'\theta_0 + U_0 \le c. \end{cases}$$

Consequently, the residual from the censored regression is

$$U := Y - X'\theta_0 = \begin{cases} U_0 & \text{if } U_0 > c - X'\theta_0 \\ c - X'\theta_0 & \text{if } U_0 \le c - X'\theta_0. \end{cases}$$

Let V denote the version of U, recensored from the right by $X'\theta_0 - c = -(c - X'\theta_0)$, i.e.,

$$V := \begin{cases} U & \text{if } U < X'\theta_0 - c \\ X'\theta_0 - c & \text{if } U \ge X'\theta_0 - c \end{cases} = \begin{cases} X'\theta_0 - c & \text{if } U_0 \ge X'\theta_0 - c \\ U_0 & \text{if } c - X'\theta_0 < U_0 < X'\theta_0 - c \\ c - X'\theta_0 & \text{if } U_0 \le c - X'\theta_0. \end{cases}$$

Clearly, $\mathrm{Law}(V|X)$ has non-empty support if $X'\theta_0 - c > 0$. In this case, $V|X$ is symmetrically distributed about the origin on its support $[c - X'\theta_0, X'\theta_0 - c]$ if and only if $U_0|X$ is symmetrically distributed on $[c - X'\theta_0, X'\theta_0 - c]$. Hence, $\hat{\mathrm{T}}_{\max}$ can consistently test H_0 against the alternative hypothesis that "$\tilde{H}_0 : U_0|X \overset{\mathrm{d}}{=} -U_0|X$ on $[c - X'\theta_0, X'\theta_0 - c]$" is false. The definition of $\hat{\mathrm{T}}_{\max}$

does not change, and it can be implemented as described in Section 3.1 with V as defined above. The procedure for bootstrapping the test goes through as in Section 3.3 with

$$\hat{Y}^* := \max(X'\hat{\theta} + R\hat{U}, c) \quad \text{and} \quad \hat{V}^* := \begin{cases} \hat{U}^* & \text{if } \hat{U}^* < X'\hat{\theta}^* - c \\ X'\hat{\theta}^* - c & \text{if } \hat{U}^* \geq X'\hat{\theta}^* - c. \end{cases}$$

5.2. **Truncation.** Instead of observing Y_0, we observe its truncated version

$$Y := \begin{cases} Y_0 & \text{if } Y_0 > c \\ \text{unobserved} & \text{otherwise.} \end{cases}$$

Consequently, the residual from the truncated regression is

$$U := Y - X'\theta_0 = \begin{cases} U_0 & \text{if } U_0 > c - X'\theta_0 \\ \text{unobserved} & \text{otherwise.} \end{cases}$$

Let V denote the version of U, retruncated from the right by $X'\theta_0 - c$, i.e.,

$$V := \begin{cases} U & \text{if } U < X'\theta_0 - c \\ \text{missing} & \text{otherwise} \end{cases} = \begin{cases} U_0 & \text{if } c - X'\theta_0 < U_0 < X'\theta_0 - c \\ \text{missing} & \text{otherwise.} \end{cases}$$

The conditional distribution of $V|X$ has non-empty support $(c - X'\theta_0, X'\theta_0 - c)$ if $X'\theta_0 - c > 0$. In this case, $V|X \overset{\text{d}}{=} -V|X$ on $(c - X'\theta_0, X'\theta_0 - c)$ if and only if $U_0|X \overset{\text{d}}{=} -U_0|X$ on $(c - X'\theta_0, X'\theta_0 - c)$. Hence, $\hat{\mathrm{T}}_{\max}$ will consistently test H_0 again the alternative hypothesis that "$U_0|X \overset{\text{d}}{=} -U_0|X$ on $(c - X'\theta_0, X'\theta_0 - c)$" is false. The definition of $\hat{\mathrm{T}}_{\max}$ does not change, and it can be implemented as described in Section 4 with V as defined above. Similarly, the bootstrap in Section 4 goes through with the only changes being that now

$$\hat{Y}^* := \begin{cases} X'\hat{\theta} + R\hat{U} & \text{if } X'\hat{\theta} + R\hat{U} > c \\ \text{missing} & \text{otherwise} \end{cases} \quad \text{and} \quad \hat{V}^* := \begin{cases} \hat{U}^* & \text{if } \hat{U}^* < X'\hat{\theta}^* - c \\ \text{missing} & \text{otherwise.} \end{cases}$$

6. SIMULATION STUDY

We now investigate the finite sample properties of $\hat{\mathrm{T}}_{\max}$. To keep the running time manageable, the results in Tables 2–6 are based on 500 simulations and 1000 bootstrap replications per simulation for estimating the critical values. Code for the simulation experiments, and the applications in Section 7, is written in `R` and is available from the authors.

6.1. **Censored regression.** We use the modified Newton-Raphson algorithm of Santos Silva (2001) to implement the SCLS estimator, which converges rapidly even when the number of regressors is large (as in the applications in Section 7).

6.1.1. *Empirical size of* $\hat{\mathrm{T}}_{\max}$. The behavior of $\hat{\mathrm{T}}_{\max}$, when H_0 is true, is reported in Table 2 for various sample sizes and four specifications of the latent model $Y_0 := \theta_0 + \theta_1 X + U_0$, where $X \stackrel{\mathrm{d}}{=} \mathrm{Unif}(-1.7, 1.7)$ and $U_0 \stackrel{\mathrm{d}}{=} N(0, \sigma^2(X))$. In each specification, $(\theta_0, \theta_1) := (1, 1)$ corresponds to medium (approx. 25%) censoring and $(\theta_0, \theta_1) := (0, 1)$ to high (approx. 50%) censoring. The first two designs, which emulate a conditionally homoscedastic U_0 by letting $\sigma^2(X) := 1$, are identical to designs 4 and 1 of Powell (1986b, Section 4). By contrast, in the next two designs $\sigma^2(X) := 1 + 0.24255X$ so that U_0 is conditionally heteroscedastic. This "increasing scedastic function" is the one specified in Case 2 of Santos Silva (2001, Section 3).

The results in Table 2 suggest that our test works well when H_0 is true. When U_0 is conditionally homoscedastic, namely, the first two designs, for each sample size the rejection probabilities are statistically no different from their nominal values at 5% level of significance, i.e., the t-statistics constructed using the Monte Carlo standard errors are bounded by 2 in absolute value. When U_0 is conditionally heteroscedastic, i.e., in the last two designs, the test appears to over-reject a little bit when the sample size $n \leq 200$. This is not entirely unexpected because heteroscedasticity increases the variability of the SCLS estimator. However, the rejection rates again become statistically equal to their nominal values as the sample size becomes moderately large, namely, $n = 400$ when censoring is medium and $n = 600$ when censoring is high. Further increasing the sample size does not significantly change the rejection rates, as is evident from the results for $n = 800$. In short, the empirical size of $\hat{\mathrm{T}}_{\max}$ appears to be fairly robust to the degree of censoring and the scale of U_0, at least in moderately large samples.

6.1.2. *Empirical power of* $\hat{\mathrm{T}}_{\max}$. To examine the power of $\hat{\mathrm{T}}_{\max}$ in finite samples, the latent response Y_0 was generated with the asymmetric distribution $U_0 \stackrel{\mathrm{d}}{=} \mathrm{LogNormal}(\mu, \sigma^2)$, and the parameters $\theta_0, \theta_1, \mu, \sigma^2$ were chosen to produce three approximate levels of censoring, namely, 3% (low), 25% (medium), and 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 in Table 3 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 3 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, $\hat{\mathrm{T}}_{\max}$ appears to have good power for moderate sample sizes even when censoring is high and V is not very skewed.

6.1.3. *Power of* $\hat{\mathrm{T}}_{\max}$ *vs. the power of Newey's statistic.* As mentioned in Section 1, a test for conditional symmetry in censored regression models has been proposed by Newey (1987, Section 4). However, Newey's statistic can only test the validity of some of the moment conditions implied by conditional symmetry rather than the shape restriction of conditional symmetry itself. Consequently, unlike $\hat{\mathrm{T}}_{\max}$, Newey's statistic is not consistent, i.e., it cannot detect all deviations from conditional symmetry on the interval $[-X'\theta_0, X'\theta_0]$ even in large samples. To illustrate this, we performed a small simulation experiment to compare the power of $\hat{\mathrm{T}}_{\max}$ with that of Newey's statistic.

Before commenting on the results of this experiment, we briefly review how Newey's statistic is constructed (we use our notation). Newey bases his test on the observation that

$$H_0 \implies \begin{cases} \mathbb{E}[\tilde{X} V \mathbb{1}_{(0,\infty)}(\tilde{X}'\beta_0)] = 0 & (6.1a) \\ \mathbb{E}[\tilde{X} V^3 \mathbb{1}_{(0,\infty)}(\tilde{X}'\beta_0)] = 0, & (6.1b) \end{cases}$$

where $\tilde{X} := (1, X)_{2\times 1}$ and $\beta_0 := (\theta_0, \theta_1)_{2\times 1}$. Let $\check{\beta}$ denote the optimal GMM estimator of β_0 based on (6.1a) and (6.1b). Note that Powell's SCLS estimator, denoted here by $\hat{\beta}$, solves (6.1a), whereas (6.1b) is an additional implication of conditional symmetry. Hence, $\check{\beta}$ is more efficient than $\hat{\beta}$. In the test proposed by him, Newey contrasts $\hat{\beta}$ with $\check{\beta}$ using the statistic $\hat{T}_{\text{Newey}} := n(\hat{\beta} - \check{\beta})'(\hat{\Omega} - \check{\Omega})^{-1}(\hat{\beta} - \check{\beta})$, where $\hat{\Omega}$ (resp. $\check{\Omega}$) is the estimated variance of $\hat{\beta}$ (resp. $\check{\beta}$), and both variances are estimated using the same estimator ($\hat{\beta}$ in the simulations) so that their difference is positive semi-definite.[13] When H_0 is true, $\hat{\beta}$ and $\check{\beta}$ are both consistent estimators of β_0 and, hence, $\hat{T}_{\text{Newey}}$ is asymptotically distributed as a $\chi^2_{\dim(\beta_0)}$ random variable. By contrast, since Newey's statistic actually tests the hypothesis that $\operatorname{plim}(\hat{\beta}) = \operatorname{plim}(\check{\beta})$ rather than H_0, there is no guarantee that $\hat{T}_{\text{Newey}}$ will reject H_0 when $\tilde{H}_0$ is false. Indeed, consider a data generating process for which $\mathbb{E}[V|X] = 0$ and $\mathbb{E}[V^3|X] = 0$ hold w.p.1, but $\mathbb{E}[V^5|X] \neq 0$ with positive probability so that $V|X$ is not symmetrically distributed about the origin and, thus, $\tilde{H}_0$ is false. For such data generating processes, $\hat{T}_{\text{Newey}}$ will continue to be asymptotically $\chi^2_{\dim(\beta_0)}$ because $\hat{\beta}$ and $\check{\beta}$ remain consistent due to the fact that the moment conditions in (6.1a) and (6.1b) are still valid. In short, $\hat{T}_{\text{Newey}}$ will have trivial power, i.e., power no greater than its size, against certain deviations from conditional symmetry even in infinitely large samples.

We compare the power of $\hat{T}_{\text{Newey}}$ with $\hat{\mathrm{T}}_{\max}$ for Design 1, cf. Table 3 for its specification, with two slight modifications. Namely, the support of X is $(-0.5, 0.5)$ instead of $(-1.7, 1.7)$ and $\theta_0 := 0.05$ instead of 1. Shrinking the support of X helps us construct the appropriate alternative (cf. below for details) and setting $\theta_0 := 0.05$ ensures that the censoring probability remains unchanged (approx. 3%) when $\operatorname{supp}(X) := (-0.5, 0.5)$. We only consider Design 1,

[13] Newey actually uses a one-step approximation of $\check{\beta}$ to construct $\hat{T}_{\text{Newey}}$, cf. (4.5) of his paper, which is what we use in the simulations. Since $\hat{T}_{\text{Newey}}$ only tests the validity of a finite number of moment conditions implied by conditional symmetry, it does not lead to a consistent test of conditional symmetry even if there is no censoring. Indeed, if there is no censoring, then Y_0 is fully observed and $V := U_0$ because there is no need to recensor the residuals. Hence, in the absence of censoring, $\hat{T}_{\text{Newey}}$ reduces to comparing the OLS estimator with a GMM estimator that is more efficient under conditional symmetry. Clearly, this contrast is not strong enough to detect all deviations from conditional symmetry.

because departures from conditional symmetry are easiest to detect for this design (recall that Design 1 entails low censoring and medium skewness). Consequently, if $\hat{T}_{\text{Newey}}$ has trivial power for the most favorable design, its performance cannot improve as the designs become less favorable.

A word about how the data was simulated for this experiment. Ideally, the observations should be generated by choosing distributions for U_0 and X such that the recensored residual V satisfies $\mathbb{E}[V|X] = 0$ w.p.1 and $\mathbb{E}[V^3|X] = 0$ w.p.1, but $\mathbb{E}[V^5|X] \neq 0$ with positive probability. However, this is difficult to do in the present setting because V is a non-linear transformation of U_0 and X. Instead, an easier approach is to proceed as follows. First, we generate $\hat{V}$ using the modified Design 1. Next, we transform $\hat{V}$ into a new random variable $\dot{V}$ that has the same support as $\hat{V}$ (because we want $\dot{V}$ to resemble $\hat{V}$) and, conditional on the X data, its first and third moments are zero but the fifth moment is non-zero.[14] Consequently, conditional on the X data, $\dot{V}$ is not symmetrically distributed about the origin although it satisfies (6.1a) and (6.1b). For instance, when $\text{supp}(X) = (-0.5, 0.5)$ and $n = 200$, we have $\sum_{j=1}^{n} \tilde{X}_j \dot{V}_j \mathbb{1}_{(0,\infty)}(\tilde{X}_j' \hat{\beta})/n = \left[\begin{smallmatrix} -0.0016 \\ -0.0001 \end{smallmatrix}\right]$ and $\sum_{j=1}^{n} \tilde{X}_j \dot{V}_j^3 \mathbb{1}_{(0,\infty)}(\tilde{X}_j' \hat{\beta})/n = \left[\begin{smallmatrix} -0.0003 \\ -0.0002 \end{smallmatrix}\right]$.[15] Therefore, we use $\dot{V}$ to estimate (6.1a) and (6.1b), i.e., obtain $\check{\beta}$, and construct $\hat{T}_{\text{Newey}}$ and $\hat{\mathrm{T}}_{\max}$.

Table 4 displays the power of $\hat{T}_{\text{Newey}}$ vs. $\hat{\mathrm{T}}_{\max}$ for several sample sizes. The results are as expected. The rejection probabilities for $\hat{T}_{\text{Newey}}$ decrease steadily as the sample size increases and are practically equal to their nominal values for $n = 500$, indicating that $\hat{T}_{\text{Newey}}$ has trivial power. In sharp contrast, the rejection rates for $\hat{\mathrm{T}}_{\max}$ are always much bigger than their nominal values and increase steadily as n gets larger. The reason why the empirical power of $\hat{\mathrm{T}}_{\max}$ in Table 4 appears to be lower than the numbers reported in Table 3 is due to the fact that the skewness of $\dot{V}$ is zero (recall that the first and third moments of $\dot{V}$, conditional on the X data, are zero by construction). Thus, deviations from conditional symmetry are harder to detect for $\dot{V}$ than for the alternatives used in Table 3.

6.2. **Truncated regression.** Since the focus of this paper is on censored regression models, we only present limited simulation results for the truncated case. The behavior of $\hat{\mathrm{T}}_{\max}$ when H_0 is true and Y_0 is truncated is reported in Table 5 for a design with medium (25%) truncation and

[14]This is done as follows. Independently of the (Y_0, X, U_0) observations, we generate a Rademacher random variable R and another independent discrete random variable W with support $\{-1/2, -1/3, 1/5, 1/2\}$ whose probability distribution is chosen so that $\mathbb{E}W = \mathbb{E}W^3 = 0$, but $\mathbb{E}W^5 \neq 0$. Letting $\dot{V} := 0.5\hat{V}R + \hat{V}W$, it is clear that $\text{supp}(\dot{V}) = \text{supp}(\hat{V})$. Furthermore, $\mathbb{E}R = \mathbb{E}W = \mathbb{E}W^3 = 0 \neq \mathbb{E}W^5$ plus the fact that R is independent of W implies that the first and third moments of $\dot{V}$, though not its fifth moment, are zero.

[15]The fact that these sample averages are very close to zero is influenced by the fact that the support of X is not too large. Indeed, as the support of X gets bigger, the probability that it takes extreme values increases implying that $\sum_{j=1}^{n} \tilde{X}_j \dot{V}_j \mathbb{1}_{(0,\infty)}(\tilde{X}_j' \hat{\beta})/n$ and $\sum_{j=1}^{n} \tilde{X}_j \dot{V}_j^3 \mathbb{1}_{(0,\infty)}(\tilde{X}_j' \hat{\beta})/n$ may not be close to zero. Choosing the support of X to be $(-0.5, 0.5)$ ensures enough variation in X and at the same time leads to a $\dot{V}$ which possesses the desired properties.

a conditionally homoscedastic U_0. This is Design (a) in Table 2, the only difference being that the censoring mechanism used for generating Y is replaced by truncation. The results are quite good. Except for one case, corresponding to $n = 200$ and nominal size = 10%, the rejection rates in Table 5 appear numerically to be only slightly worse than those for Design (a) in Table 2, due to the fact that truncation causes more information loss than censoring. Nonetheless, for $n \leq 200$ (excluding the case mentioned earlier), the rejection probabilities are statistically no different from their nominal values at 5% level of significance. Reassuringly, the exception too falls in line when the sample size increases to $n = 300$ suggesting that the empirical size of $\hat{\mathrm{T}}_{\max}$ is close to its nominal value in moderately sized samples.

The power of $\hat{\mathrm{T}}_{\max}$ when Y_0 is truncated is reported in Table 6 for a design with 27% truncation and conditionally homoscedastic U_0, which is just the truncated version of Design 2 in Table 3. The rejection rates in Table 6 are very similar to those for Design 2 in Table 3, suggesting that $\hat{\mathrm{T}}_{\max}$ has good power to detect deviations from conditional symmetry in a neighborhood of zero even in truncated regression models.

7. APPLICATIONS

We now apply $\hat{\mathrm{T}}_{\max}$ to two empirical datasets to test the assumption of conditional symmetry in the context of estimating censored regression models. The first dataset was used by Lee (1996, Appendix, Section 8) to estimate (2.2), referred to as a "type 1" censored tobit model in the terminology of Amemiya (1985, Section 10.2), under conditional symmetry. The second dataset we examine was used by Jolliffe (2002) to estimate a selection model, what Amemiya (Section 10.8) calls a "type 3" censored tobit model, under conditional symmetry. Both datasets are publicly available: Lee's dataset is printed on pages 256–260 of his book, and Jolliffe's dataset can be downloaded from `http://go.worldbank.org/94GH3XWAR1`.

7.1. Application 1: Type 1 censored tobit under conditional symmetry.

In Section 8 of the appendix of his book, Lee estimates a regression model of female labor supply (Y_0) which is censored from below at 0. In his model, $Y \geq 0$ denotes hours worked by married females, X is a 7×1 vector containing the following determinants of labor supply: other household income, wife's age, wife's schooling, husband's employment status, number of preschool children, number of primary school children, and whether or not house is on mortgage,[16] and U_0 is assumed to be conditionally symmetric given the regressors. The dataset that Lee employs to estimate this model using Powell's SCLS estimator, consists of 200 observations from the 1987 Michigan Panel Study of Income Dynamics (PSID). The proportion of censored observations in the dataset is 22%, which corresponds to a moderate level of censoring, and the estimated $\Pr(X'\theta_0 > 0)$ is 0.975.

[16]The full list of variables, and other details about his dataset, can be found in Lee (1996, p. 235).

Before applying $\hat{\mathrm{T}}_{\max}$ to Lee's dataset, we look at the QQ plot of $\hat{V}$ versus $-\hat{V}$. It can be seen from Figure 1 (a) that the QQ plot is very close to the 45° line in a neighborhood of zero, suggesting that the marginal distribution of $\hat{V}$ may be symmetric in a neighborhood of the origin.[17] Of course, this graphical evidence regarding the marginal distribution of $\hat{V}$ is not particularly informative about testing conditional symmetry, which is a conditional feature. Therefore, we applied $\hat{\mathrm{T}}_{\max}$ to this dataset, obtaining a p-value of 0.063. Hence, at 5% level of significance, Lee's dataset supports the assumption that $U_0|X$ is symmetrically distributed about the origin on $[-X'\theta_0, X'\theta_0]$. However, at 10% level of significance, which is routinely acceptable in typical economic applications, H_0 is rejected.

The fact that H_0 is rejected at 10%, but not 5%, level of significance suggests, albeit inconclusively, that H_0 should be rejected in practice. The standard advice to resolve this dilemma is to get more data, in the hope that the additional observations will help reduce sampling uncertainty and thereby provide more accurate and unambiguous inference. Fortunately, this prescription is straightforward to implement here because the dataset in Lee's book is part of a much larger sample employed by Lee (1995). The larger dataset, downloadable from the Journal of Applied Econometrics website, has 3382 observations and the proportion of censored observations is 25% (Lee, 1995, Table I). The estimated $\Pr(X'\theta_0 > 0)$ is 0.952, where X is now a 14×1 vector of regressors in the RF model of Lee (1995, Section 3, p. 191); cf. Table II in his paper for the list of explanatory variables in X. We applied $\hat{\mathrm{T}}_{\max}$ to this dataset to test the assumption of conditional symmetry maintained in Lee's RF model, and obtained a p-value of 0. Therefore, in the larger sample, the assumption of conditional symmetry maintained in Lee's paper is rejected decisively at all levels of significance.

7.2. **Application 2: Type 3 censored tobit under conditional symmetry.** A type 3 censored model is one for which selection into the sample is determined by another censored variable. For instance, consider the system of equations

$$W_0 = Z'\gamma_0 + \varepsilon \qquad \text{(wage equation)}$$
$$Y_0 = X'\theta_0 + U_0, \qquad \text{(selection equation)}$$

where (W_0, Y_0) are not observed. Instead, we observe their censored versions

$$W := \begin{cases} W_0 & \text{if } Y_0 > 0 \\ 0 & \text{otherwise} \end{cases} \qquad \& \qquad Y := \begin{cases} Y_0 & \text{if } Y_0 > 0 \\ 0 & \text{otherwise.} \end{cases}$$

[17] Since the marginal distribution of $\hat{V}$ appears to be symmetric about the origin, in Figure 1 (b) we plot the quantiles of the standardized version of $\hat{V}$ versus the quantiles of a standard normal random variable. This QQ plot suggests that the marginal distribution of $\hat{V}$ is gaussian, which helps explain the finding in Lee (1996, p. 249) that the SCLS point estimates are very close to the maximum likelihood estimates (obtained under the assumptions that U_0 is normally distributed and independent of X).

Clearly, this generalizes the standard, or type 1, censored regression model (2.2). The parametric approach for estimating (γ_0, θ_0) is to assume that (ε, U_0) is jointly normal and independent of (W, X) and then either do a two-step selection correction (Heckman, 1976), or full information maximum likelihood (Amemiya, 1985, Section 10.8.1). However, Honoré, Kyriazidou, and Udry (1997) showed that (γ_0, θ_0) could be consistently estimated without making distributional assumptions and without assuming independence of the model errors and regressors, thus allowing for conditionally heteroscedastic errors, provided it was assumed that the model errors are symmetrically distributed about the origin conditional on the regressors, i.e., $(\varepsilon, U_0)|Z, X \stackrel{d}{=} (-\varepsilon, -U_0)|Z, X$.

Using data from the 1995 Bulgaria Integrated Household Survey (BIHS) consisting of 2182 females and 2067 males, Jolliffe (2002) applied the approach of Honoré et al. to estimate wage regressions separately for women and men. By comparing these two wage regressions, Jolliffe could estimate the gender wage gap in Bulgaria as a function of the conditioning variables in the wage equation. A non-negligible wage gap (that Jolliffe found) is evidence of gender discrimination in the Bulgarian labor market, which was an important concern in the (then) ongoing accession talks between Bulgaria and the European Union (Jolliffe, p. 277).

In Jolliffe's model, W_0 denotes unobserved wage offers, W is hourly wage in log, Y_0 is unobserved labor supply, and Y is log of hours in wage work. The fraction of individuals reporting zero log hours is 59.1% (females) and 52.5% (males), and the estimated $\Pr(X'\theta_0 > 0)$ is 0.477 (females) and 0.499 (males). The 31×1 vector X contains determinants of labor supply including, e.g., schooling dummies, experience, sectors of employment, region dummies, number of children, various incomes, and an intercept. Cf. Table 1 in Jolliffe's paper for a full list of the variables in X. Similarly, Z is a 27×1 vector of wage determining variables listed in Table 2 of Jolliffe's paper.

For the remainder of this section, we focus our attention on testing the conditional symmetry of U_0, the error term in the selection model. We begin by looking at the QQ plot of $\hat{V}$ versus $-\hat{V}$. The plots in Figure 2 suggest that $\hat{V}$ is not symmetrically distributed about the origin for both subsamples, i.e., females and males. However, this evidence is purely graphical and should not be taken as implying that U_0 is not conditionally symmetric. Indeed, applying $\hat{\mathrm{T}}_{\max}$ to the data used for estimating the selection equation, we get p-values of 0.171 for the female subsample and 0.01 for the male subsample. This suggests that the female subsample supports the assumption that $U_0|X$ is symmetrically distributed about the origin on $[-X'\theta_0, X'\theta_0]$ at 5%, or even 10%, level of significance. However, H_0 appears to be strongly rejected for the male subsample.

What are the implications of this specification test for the estimates in Jolliffe's paper? If $U_0|X$ is not symmetric about zero, then $(\varepsilon, U_0)|Z, X$ cannot be symmetric about the origin, implying that the estimator of Honoré et al. is not consistent for (γ_0, θ_0). This raises questions about the estimates for the male subsample. On the other hand, symmetry of $U_0|X$ about zero does not imply symmetry of $(\varepsilon, U_0)|Z, X$ about the origin. Therefore, although the female subsample supports the assumption that $U_0|X$ is symmetrically distributed about the origin on $[-X'\theta_0, X'\theta_0]$ at 5% and 10% levels of significance, we cannot conclude the same for $(\varepsilon, U_0)|Z, X$. Hence, whether $(\varepsilon, U_0)|Z, X$ is symmetric about the origin for the female subsample, and consequently the validity of the estimates for the female subsample, remains an open question and a topic for future research.

8. CONCLUSION

We have shown how to consistently test the assumption of conditional symmetry maintained in the symmetrically trimmed least-squares approach of Powell (1986b), which is used for estimating tobit models without making distributional assumptions. The Kolmogorov-Smirnov type statistic we propose is easy to implement because it requires neither nonparametric smoothing nor optimization over an uncountable set. Simulation results suggest that our test statistic has good size and power in small to moderately sized samples and that its performance does not appear to be too sensitive to the degree of censoring or truncation. An application of our test to two empirical datasets yields some interesting findings.

ACKNOWLEDGEMENTS

We are grateful to Dean Jolliffe for helping us recreate the samples used in his paper, and to Joao Santos Silva for drawing our attention to a reference that we had missed.

APPENDIX A. TECHNICAL DETAILS

The observations $(Y_1, X_1), \ldots, (Y_n, X_n)$ are assumed to be independently and identically distributed (i.i.d.) and all limits are taken as the sample size $n \to \infty$.

The following notation is used throughout the appendix. Let $\mathbb{P}_Z$ be the distribution of a random vector Z, and $\hat{\mathbb{P}}_Z$ denote the empirical measure induced by n observations on Z, i.e., $\hat{\mathbb{P}}_Z := n^{-1} \sum_{j=1}^{n} \delta_{Z_j}$, where δ_{Z_j} is the point measure at Z_j. Integrals are frequently written in linear functional notation, e.g., $\mathbb{P}_Z h := \int h(z)\, \mathbb{P}_Z(dz)$ and $\hat{\mathbb{P}}_Z h := n^{-1} \sum_{j=1}^{n} h(Z_j)$. The set of real-valued functions of Z that are square-integrable with respect to $\mathbb{P}_Z$ is denoted by $L_2(\mathbb{P}_Z)$. The $L_2(\mathbb{P}_Z)$ inner product and norm are $\langle h_1, h_2 \rangle_{\mathbb{P}_Z} := \int h_1(z) h_2(z)\, \mathbb{P}_Z(dz)$ and $\|h\|_{2,\mathbb{P}_Z} := \langle h, h \rangle_{\mathbb{P}_Z}^{1/2}$, respectively. The Euclidean norm is $\|\cdot\|$, and $B(\theta, \epsilon)$ is the open ball of radius ϵ centered at θ (its $\|\cdot\|$-closure is $\overline{B(\theta, \epsilon)}$).

In order to prove Proposition 3.1, we need an asymptotic approximation of $\hat{\mathrm{T}}_{\max}$ that holds whether H_0 is true or not. We begin by describing this approximation.

Proposition A.1. *Let* $\mathcal{F}_1 := \{\mathbb{1}_{(-\infty,s]\times(-\infty,t]} : (s,t) \in \mathbb{R} \times \mathbb{R}^{\dim(X)}\}$*. Then, under the assumptions stated in its proof, and irrespective of whether* H_0 *is true or false,*

$$\hat{\mathrm{T}}_{\max} = (1 + o_{\mathrm{p}}(1)) \sup_{f\in\mathcal{F}_1} n^{1/2}|\hat{\mathbb{P}}_{U_0,X} g_f^{\hat{\theta}} - \hat{\mathbb{P}}_{U_0,X} g_{f^{\mathrm{r}}}^{\hat{\theta}}|,$$

where g_f^θ *is a* $\{0,1\}$*-valued function of* (U_0, X) *defined in* (A.1)*, and* f^{r} *denotes* f *with its first argument reflected about the origin, i.e.,* $f^{\mathrm{r}}(U_0, X) := f(-U_0, X)$*.*

Therefore, irrespective of whether H_0 is true or false, the asymptotic behavior of $\hat{\mathrm{T}}_{\max}$ is determined by the asymptotic behavior of

$$\hat{\mathrm{T}}_{\sup} := \sup_{f\in\mathcal{F}_1} n^{1/2}|\hat{\mathbb{P}}_{U_0,X} g_f^{\hat{\theta}} - \hat{\mathbb{P}}_{U_0,X} g_{f^{\mathrm{r}}}^{\hat{\theta}}|,$$

Hence, to prove Proposition 3.1, it suffices to derive the asymptotic distribution of $\hat{\mathrm{T}}_{\sup}$. For this, we need the following result.

Proposition A.2. *Under the assumptions stated in its proof, and without assuming that* $U_0|X$ *is conditionally symmetric about the origin, there exists a neighborhood of* θ_0*, and a function* $D : \mathcal{F}_1 \times \mathbb{R}^{\dim(\theta_0)} \to \mathbb{R}^{\dim(\theta_0)}$*, such that for all* θ *in this neighborhood,*

$$\sup_{f\in\mathcal{F}_1} |\hat{\mathbb{P}}_{U_0,X} g_f^{\theta} - \hat{\mathbb{P}}_{U_0,X} g_f^{\theta_0} - D'(f,\theta_0)(\theta - \theta_0)| = o_{\mathrm{p}}(n^{-1/2}) + o(\|\theta - \theta_0\|)$$

$$\sup_{f\in\mathcal{F}_1} |\hat{\mathbb{P}}_{U_0,X} g_{f^{\mathrm{r}}}^{\theta} - \hat{\mathbb{P}}_{U_0,X} g_{f^{\mathrm{r}}}^{\theta_0} - D'(f^{\mathrm{r}},\theta_0)(\theta - \theta_0)| = o_{\mathrm{p}}(n^{-1/2}) + o(\|\theta - \theta_0\|).$$

This result shows how the processes $\{\hat{\mathbb{P}}_{U_0,X} g_f^{\theta} : f \in \mathcal{F}_1\}$ and $\{\hat{\mathbb{P}}_{U_0,X} g_{f^{\mathrm{r}}}^{\theta} : f \in \mathcal{F}_1\}$ can be linearized about θ_0. Since conditional symmetry of $U_0|X$ is not required to prove Proposition A.2, it can be used to characterize the asymptotic behavior of $\hat{\mathrm{T}}_{\sup}$ even when $\tilde{H}_0$ is false.

We are now ready to prove Proposition 3.1.

Proof of Proposition 3.1. Following the discussion after Proposition A.1, it suffices to look at $\hat{\mathrm{T}}_{\sup}$. We first show that, under H_0, $\hat{\mathrm{T}}_{\sup}$ converges in distribution to a random variable T_∞. Begin by recalling that the SCLS estimator $\hat{\theta}$ is $n^{1/2}$-consistent for θ_0 under H_0. Hence, by Proposition A.2, $\hat{\mathrm{T}}_{\sup} = \sup_{f\in\mathcal{F}_1} |\hat{\mathbb{X}}_0(f)| + o_{\mathrm{p}}(1)$ with

$$\hat{\mathbb{X}}_0(f) := n^{1/2}\hat{\mathbb{P}}_{U_0,X}(g_f^{\theta_0} - g_{f^{\mathrm{r}}}^{\theta_0}) + (D(f,\theta_0) - D(f^{\mathrm{r}},\theta_0))' n^{1/2}(\hat{\theta} - \theta_0).$$

Under H_0, $\{\hat{\mathbb{P}}_{U_0,X}(g_f^{\theta_0} - g_{f^{\mathrm{r}}}^{\theta_0}) : f \in \mathcal{F}_1\}$ is a mean zero gaussian process and $n^{1/2}(\hat{\theta} - \theta_0)$ is asymptotically linear (cf. equation A.15 in Powell's paper). Hence, $\{\hat{\mathbb{X}}_0(f) : f \in \mathcal{F}_1\}$ is asymptotically tight, implying that it converges in distribution, in the set of bounded functions from $\mathcal{F}_1 \to \mathbb{R}$, to some tight process $\{\mathbb{X}_0(f) : f \in \mathcal{F}_1\}$. The limiting process $\mathbb{X}_0$ is also gaussian because the finite dimensional distributions of $\{\hat{\mathbb{X}}_0(f) : f \in \mathcal{F}_1\}$ are gaussian. It follows that, under H_0, $\hat{\mathrm{T}}_{\sup}$ converges in distribution to the random variable $\mathrm{T}_\infty := \sup_{f\in\mathcal{F}_1} |\mathbb{X}_0(f)|$.

Next, we show that if $\tilde{H}_0$ is false, then $\hat{\mathrm{T}}_{\sup} \to \infty$ w.p.a.1. So assume that $\tilde{H}_0$ is false. Then, for all $\theta \in \mathbb{R}^{\dim(X)}$, there exists $(x, v) \in \mathrm{supp}(X) \times [-x'\theta, x'\theta]$ for which $\Pr(Y_0 - x'\theta \leq v|X = x) \neq \Pr(-(Y_0 - x'\theta) \leq v|X = x)$.[18] In particular, $\mathrm{cdf}_{U_\dagger|X} \neq \mathrm{cdf}_{-U_\dagger|X}$ on the interval $[-X'\theta_\dagger, X'\theta_\dagger]$, where $U_\dagger := Y_0 - X'\theta_\dagger$ and $\theta_\dagger$ is the "pseudo-true value" of the parameter of interest, i.e., the probability limit of $\hat{\theta}$ when $\tilde{H}_0$ is false, cf. the discussion after (A.2). But this means that, conditional on X, the symmetrically censored version of $U_\dagger$, defined in the same manner as V in (2.3) but with θ_0 (resp. U_0) replaced by $\theta_\dagger$ (resp. $U_\dagger$), cannot be distributed symmetrically about the origin. In other words, $\sup_{f\in\mathcal{F}_1} |\mathbb{P}_{U_\dagger,X} g_f^{\theta_\dagger} - \mathbb{P}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\theta_\dagger}| > 0$ because $\mathcal{F}_1$ separates probability measures. Note that under the alternative, $\hat{\mathrm{T}}_{\sup} = \sup_{f\in\mathcal{F}_1} n^{1/2}|\hat{\mathbb{P}}_{U_\dagger,X} g_f^{\hat{\theta}} - \hat{\mathbb{P}}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\hat{\theta}}|$. Let us examine how $\sup_{f\in\mathcal{F}_1} |\hat{\mathbb{P}}_{U_\dagger,X} g_f^{\hat{\theta}} - \hat{\mathbb{P}}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\hat{\theta}}|$ behaves under the alternative. Begin by writing

$$\sup_{f\in\mathcal{F}_1} |\hat{\mathbb{P}}_{U_\dagger,X} g_f^{\hat{\theta}} - \hat{\mathbb{P}}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\hat{\theta}}| = \sup_{f\in\mathcal{F}_1} |\mathbb{P}_{U_\dagger,X} g_f^{\theta_\dagger} - \mathbb{P}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\theta_\dagger}| + \hat{d},$$

where $\hat{d} := \sup_{f\in\mathcal{F}_1} |\hat{\mathbb{P}}_{U_\dagger,X} g_f^{\hat{\theta}} - \hat{\mathbb{P}}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\hat{\theta}}| - \sup_{f\in\mathcal{F}_1} |\mathbb{P}_{U_\dagger,X} g_f^{\theta_\dagger} - \mathbb{P}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\theta_\dagger}|$. By the reverse triangle inequality,

$$\begin{aligned}
|\hat{d}| &\leq \sup_{f\in\mathcal{F}_1} |(\hat{\mathbb{P}}_{U_\dagger,X} g_f^{\hat{\theta}} - \hat{\mathbb{P}}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\hat{\theta}}) - (\mathbb{P}_{U_\dagger,X} g_f^{\theta_\dagger} - \mathbb{P}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\theta_\dagger})| \\
&\leq \sup_{f\in\mathcal{F}_1} |\hat{\mathbb{P}}_{U_\dagger,X} g_f^{\hat{\theta}} - \hat{\mathbb{P}}_{U_\dagger,X} g_f^{\theta_\dagger}| + \sup_{f\in\mathcal{F}_1} |\hat{\mathbb{P}}_{U_\dagger,X} g_f^{\theta_\dagger} - \mathbb{P}_{U_\dagger,X} g_f^{\theta_\dagger}| \\
&\quad + \sup_{f\in\mathcal{F}_1} |\hat{\mathbb{P}}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\hat{\theta}} - \hat{\mathbb{P}}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\theta_\dagger}| + \sup_{f\in\mathcal{F}_1} |\hat{\mathbb{P}}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\theta_\dagger} - \mathbb{P}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\theta_\dagger}|.
\end{aligned}$$

Since $\hat{\theta}$ converges in probability to $\theta_\dagger$ under the alternative, Proposition A.2 implies that

$$\sup_{f\in\mathcal{F}_1} |\hat{\mathbb{P}}_{U_\dagger,X} g_f^{\hat{\theta}} - \hat{\mathbb{P}}_{U_\dagger,X} g_f^{\theta_\dagger}| = \sup_{f\in\mathcal{F}_1} |\hat{\mathbb{P}}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\hat{\theta}} - \hat{\mathbb{P}}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\theta_\dagger}| = o_{\mathrm{p}}(1).$$

Moreover, since $\mathcal{F}_1$ is $\mathbb{P}_{U_\dagger,X}$-Glivenko-Cantelli,

$$\sup_{f\in\mathcal{F}_1} |\hat{\mathbb{P}}_{U_\dagger,X} g_f^{\theta_\dagger} - \mathbb{P}_{U_\dagger,X} g_f^{\theta_\dagger}| = \sup_{f\in\mathcal{F}_1} |\hat{\mathbb{P}}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\theta_\dagger} - \mathbb{P}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\theta_\dagger}| = o_{\mathrm{p}}(1).$$

It follows that $\hat{d} = o_{\mathrm{p}}(1)$. Hence, $\hat{\mathrm{T}}_{\sup} = n^{1/2}(\sup_{f\in\mathcal{F}_1} |\mathbb{P}_{U_\dagger,X} g_f^{\theta_\dagger} - \mathbb{P}_{U_\dagger,X} g_{f^{\mathrm{r}}}^{\theta_\dagger}| + o_{\mathrm{p}}(1)) \to \infty$ w.p.a.1 when $\tilde{H}_0$ is false. □

Proof of Proposition A.1. Recall that our statistic is $\hat{\mathrm{T}}_{\max} := \hat{N}^{1/2}\hat{\mathrm{R}}_{\max}$ with $\hat{\mathrm{R}}_{\max}$ as defined in (3.3). Let $\hat{\mathcal{F}}_1 := \{\mathbb{1}_{(-\infty,s]\times(-\infty,t]} : (s,t) \in \hat{\mathscr{V}} \times \mathscr{X}\}$ be the collection of indicator functions from $\mathbb{R} \times \mathbb{R}^{\dim(X)} \to \{0, 1\}$. Then, we can write $\hat{\mathrm{R}}_{\max}$ as

$$\hat{\mathrm{R}}_{\max} = \frac{1}{\hat{N}/n} \max_{f\in\hat{\mathcal{F}}_1} \Big|\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{(0,\infty)}(X_j'\hat{\theta}) f(\hat{V}_j, X_j) - \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{(0,\infty)}(X_j'\hat{\theta}) f(-\hat{V}_j, X_j)\Big|.$$

[18] $\tilde{H}_0 \iff (\exists\theta_0 \in \mathbb{R}^{\dim(X)})\ (\forall(x,v) \in \mathrm{supp}(X) \times [-x'\theta_0, x'\theta_0])(\Pr(U_0 \leq v|X = x) = \Pr(-U_0 \leq v|X = x))$.

Next, we obtain a simplified expression for $\hat{V}$. Let $\Delta(X,\theta_1,\theta_2) := X'\theta_1 - X'\theta_2$ and

$$\begin{aligned} V(\theta) &:= \begin{cases} Y - X'\theta & \text{if } Y - X'\theta < X'\theta \\ X'\theta & \text{if } Y - X'\theta \geq X'\theta \end{cases} \\ &= \begin{cases} Y - X'\theta_0 - \Delta(X,\theta,\theta_0) & \text{if } Y - X'\theta_0 - \Delta(X,\theta,\theta_0) < \Delta(X,\theta,\theta_0) + X'\theta_0 \\ \Delta(X,\theta,\theta_0) + X'\theta_0 & \text{if } Y - X'\theta_0 - \Delta(X,\theta,\theta_0) \geq \Delta(X,\theta,\theta_0) + X'\theta_0, \end{cases} \end{aligned}$$

so that $V(\hat{\theta}) = \hat{V}$ and $V(\theta_0) = V$. Hence, since $U := Y - X'\theta_0$,

$$\begin{aligned} V(\theta) &= (U - \Delta(X,\theta,\theta_0))\mathbb{1}(U < 2\Delta(X,\theta,\theta_0) + X'\theta_0) \\ &\qquad + (\Delta(X,\theta,\theta_0) + X'\theta_0)\mathbb{1}(U \geq 2\Delta(X,\theta,\theta_0) + X'\theta_0) \\ &= U - \Delta(X,\theta,\theta_0) + (-U + 2\Delta(X,\theta,\theta_0) + X'\theta_0)\mathbb{1}(U \geq 2\Delta(X,\theta,\theta_0) + X'\theta_0). \end{aligned}$$

Therefore, since $U = \begin{cases} U_0 & \text{if } U_0 > -X'\theta_0 \\ -X'\theta_0 & \text{if } U_0 \leq -X'\theta_0 \end{cases}$, it follows that whenever $X'\theta_0 > 0$ we have (with $\Delta := \Delta(X,\theta,\theta_0)$ to save space)

$$\begin{aligned} V(\theta) &= \begin{cases} -X'\theta_0 - \Delta + 2(\Delta + X'\theta_0)\mathbb{1}(\Delta + X'\theta_0 \leq 0) & \text{if } U_0 \in (-\infty, -X'\theta_0] \\ U_0 - \Delta & \text{if } U_0 \in (-X'\theta_0, X'\theta_0] \cap (-\infty, 2\Delta + X'\theta_0) \\ \Delta + X'\theta_0 & \text{if } U_0 \in (-X'\theta_0, X'\theta_0] \cap [2\Delta + X'\theta_0, \infty) \\ U_0 - \Delta & \text{if } U_0 \in (X'\theta_0, \infty) \cap (-\infty, 2\Delta + X'\theta_0) \\ \Delta + X'\theta_0 & \text{if } U_0 \in (X'\theta_0, \infty) \cap [2\Delta + X'\theta_0, \infty) \end{cases} \\ &= \begin{cases} -X'\theta_0 - \Delta + 2(\Delta + X'\theta_0)\mathbb{1}(\Delta + X'\theta_0 \leq 0) & \text{if } U_0 \in (-\infty, -X'\theta_0] \\ U_0 - \Delta & \text{if } U_0 \in (-X'\theta_0, \infty) \cap (-\infty, 2\Delta + X'\theta_0) \\ \Delta + X'\theta_0 & \text{if } U_0 \in (-X'\theta_0, \infty) \cap [2\Delta + X'\theta_0, \infty) \end{cases} \\ &= \begin{cases} -X'\theta_0 - \Delta + 2(\Delta + X'\theta_0)\mathbb{1}(\Delta + X'\theta_0 \leq 0) & \text{if } U_0 \in (-\infty, -X'\theta_0] \\ U_0 - \Delta & \text{if } U_0 \in (-X'\theta_0, 2\Delta + X'\theta_0) \\ \Delta + X'\theta_0 & \text{if } U_0 \in (-X'\theta_0, \infty) \cap [2\Delta + X'\theta_0, \infty). \end{cases} \end{aligned}$$

Now, assume that

Assumption A.1. $\mathbb{P}_X(X'\theta_0 > \eta_0) = 1$ *for some* $\eta_0 > 0$*, and* $\text{supp}(X)$ *is bounded, i.e.,* $\sup_{x \in \text{supp}(X)} \|x\| \leq M$ *for some* $M > 0$*.*

Then, $\{x \in \text{supp}(X) : x'\theta_0 > \eta_0\} \subset \{x \in \text{supp}(X) : \Delta(x,\theta,\theta_0) + x'\theta_0 > \eta_0/2\}$ because, by (A.6), $x'\theta_0 > \eta_0 \implies x'\theta_0 + \Delta(x,\theta,\theta_0) > \eta_0 - \|x\|\|\theta - \theta_0\| \geq \eta_0 - M\|\theta - \theta_0\|$. Thus,

$\theta \in B(\theta_0, \eta_0/2M) \implies \Delta(X, \theta, \theta_0) + X'\theta_0 > \eta_0/2$ $\mathbb{P}_X$-a.s.. Hence, if $\theta \in B(\theta_0, \eta_0/2M)$, then

$$V(\theta) = \begin{cases} -X'\theta_0 - \Delta(X,\theta,\theta_0) & \text{if } U_0 \in (-\infty, -X'\theta_0] \\ U_0 - \Delta(X,\theta,\theta_0) & \text{if } U_0 \in (-X'\theta_0, 2\Delta(X,\theta,\theta_0) + X'\theta_0) \\ \Delta(X,\theta,\theta_0) + X'\theta_0 & \text{if } U_0 \in [2\Delta(X,\theta,\theta_0) + X'\theta_0, \infty). \end{cases} \qquad (\mathbb{P}_X\text{-a.s.})$$

Consequently,

$$\hat{\theta} \in B(\theta_0, \eta_0/2M) \implies n^{-1}\sum_{j=1}^{n} \mathbb{1}_{(0,\infty)}(X_j'\hat{\theta}) f(\hat{V}_j, X_j) \overset{\text{w.p.1}}{=} n^{-1}\sum_{j=1}^{n} \mathbb{1}_{(0,\infty)}(X_j'\hat{\theta}) f^{\hat{\theta}}(U_{0j}, X_j)$$
$$=: \hat{\mathbb{P}}_{U_0,X}\, g_f^{\hat{\theta}},$$

where

$$\begin{aligned} g_f^{\theta}(U_0, X) &:= \mathbb{1}_{(0,\infty)}(X'\theta) f^{\theta}(U_0, X) \\ f^{\theta}(U_0, X) &:= f(V(U_0, X, \theta), X) \\ V(U_0, X, \theta) &:= V(\theta), \end{aligned} \qquad \text{(A.1)}$$

so that $V(U_0, X, \theta_0) = V$. Therefore,

$$\hat{\theta} \in B(\theta_0, \eta_0/2M) \implies \hat{\mathrm{R}}_{\max} \overset{\text{w.p.1}}{=} \frac{1}{\hat{N}/n} \max_{f \in \hat{\mathcal{F}}_1} |\hat{\mathbb{P}}_{U_0,X}\, g_f^{\hat{\theta}} - \hat{\mathbb{P}}_{U_0,X}\, g_{f^{\mathrm{r}}}^{\hat{\theta}}|. \qquad \text{(A.2)}$$

Note that (A.2) holds whether H_0 is true or not, provided θ_0 is suitably redefined when H_0 is false. Indeed, since Powell's SCLS estimator is an M-estimator, under some regularity conditions it has a well defined probability limit even if H_0 is false. Hence, when H_0 is false, θ_0 can be defined to be the probability limit of $\hat{\theta}$, i.e., the pseudo-true value, denoted by $\theta_\dagger$. Consequently, if H_0 is false, (A.2) holds upon replacing θ_0 (resp. U_0) with $\theta_\dagger$ (resp. $U_\dagger := Y_0 - X'\theta_\dagger$) and restating assumptions involving θ_0 in terms of $\theta_\dagger$.

As in Lemma A.4 of CT, it can be shown that, w.p.a.1, $\hat{\mathcal{F}}_1$ is dense (in the $L_2(\mathbb{P}_{U_0,X})$-norm) in $\mathcal{F}_1 := \{\mathbb{1}_{(-\infty,s]\times(-\infty,t]} : (s,t) \in \mathbb{R} \times \mathbb{R}^{\dim(X)}\}$. This result holds, mutatis mutandis, even if H_0 is false, provided θ_0 is replaced by its pseudo-true value. Therefore, since $f \mapsto |\hat{\mathbb{P}}_{U_0,X}\, g_f^{\hat{\theta}} - \hat{\mathbb{P}}_{U_0,X}\, g_{f^{\mathrm{r}}}^{\hat{\theta}}|$ is continuous (in the $L_2(\mathbb{P}_{U_0,X})$-norm) on $\mathcal{F}_1$, it follows that

$$\hat{\mathrm{R}}_{\max} \overset{\text{w.p.a.1}}{=} \frac{1}{\hat{N}/n} \sup_{f \in \mathcal{F}_1} |\hat{\mathbb{P}}_{U_0,X}\, g_f^{\hat{\theta}} - \hat{\mathbb{P}}_{U_0,X}\, g_{f^{\mathrm{r}}}^{\hat{\theta}}|,$$

irrespective of whether H_0 is true or false.

Let $\mathcal{H} := \{x \mapsto \mathbb{1}_{(s,\infty)}(x'\theta_\dagger) : s \in \mathbb{R}, x \in \operatorname{supp}(X)\}$ and $F(\theta) := \Pr(X'\theta > 0)$. Then,

$$\begin{aligned}
|\hat{N}/n - \Pr(X'\theta_\dagger > 0)| &= |\int \mathbb{1}_{(0,\infty)}(x'\hat{\theta})\, \hat{\mathbb{P}}_X(dx) - \int \mathbb{1}_{(0,\infty)}(x'\theta_\dagger)\, \mathbb{P}_X(dx)| \\
&\leq |\int \mathbb{1}_{(-\Delta(x,\hat{\theta},\theta_\dagger),\infty)}(x'\theta_\dagger)\, \hat{\mathbb{P}}_X(dx) - \int \mathbb{1}_{(-\Delta(x,\hat{\theta},\theta_\dagger),\infty)}(x'\theta_\dagger)\, \mathbb{P}_X(dx)| \\
&\quad + |\int \mathbb{1}_{(-\Delta(x,\hat{\theta},\theta_\dagger),\infty)}(x'\theta_\dagger)\, \mathbb{P}_X(dx) - \int \mathbb{1}_{(0,\infty)}(x'\theta_\dagger)\, \mathbb{P}_X(dx)| \\
&\leq \sup_{h \in \mathcal{H}} |\int h(x'\theta_\dagger)\, (\hat{\mathbb{P}}_X(dx) - \mathbb{P}_X(dx))| + |F(\hat{\theta}) - F(\theta_\dagger)| \\
&= o_{\mathrm{p}}(1) + |F(\hat{\theta}) - F(\theta_\dagger)|,
\end{aligned}$$

because $\mathcal{H}$ is $\mathbb{P}_X$-Glivenko-Cantelli. Next, assume that

Assumption A.2. $\theta \mapsto \Pr(X'\theta > 0)$ *is continuous at* $\theta_\dagger$*.*

Hence, whether H_0 is true or not, $\hat{N}/n \xrightarrow{\mathrm{p}} 1$ because $\Pr(X'\theta_\dagger > 0) = 1$ under the version of Assumption A.1 with θ_0 replaced by $\theta_\dagger$. Therefore,

$$\begin{aligned}
\hat{\mathrm{T}}_{\max} &= \frac{1}{(\hat{N}/n)^{1/2}} \sup_{f \in \mathcal{F}_1} n^{1/2} |\hat{\mathbb{P}}_{U_0,X} g_f^{\hat{\theta}} - \hat{\mathbb{P}}_{U_0,X} g_{f^{\mathrm{r}}}^{\hat{\theta}}| \qquad \text{(w.p.a.1)} \\
&= (1 + o_{\mathrm{p}}(1)) \sup_{f \in \mathcal{F}_1} n^{1/2} |\hat{\mathbb{P}}_{U_0,X} g_f^{\hat{\theta}} - \hat{\mathbb{P}}_{U_0,X} g_{f^{\mathrm{r}}}^{\hat{\theta}}|,
\end{aligned}$$

irrespective of whether H_0 is true or false. The desired result follows. □

Proof of Proposition A.2. We will show that

$$\sup_{f \in \mathcal{F}_1} |\hat{\mathbb{P}}_{U_0,X} g_f^{\theta} - \hat{\mathbb{P}}_{U_0,X} g_f^{\theta_0} - D'(f,\theta_0)(\theta - \theta_0)| = o_{\mathrm{p}}(n^{-1/2}) + o(\|\theta - \theta_0\|);$$

the second result follows from the first upon replacing f by f^{r}. Note that since

$$\hat{\mathbb{P}}_{U_0,X} g_f^{\theta} - \hat{\mathbb{P}}_{U_0,X} g_f^{\theta_0} = \hat{\mathbb{P}}_{U_0,X}(g_f^{\theta} - g_f^{\theta_0}) = (\hat{\mathbb{P}}_{U_0,X} - \mathbb{P}_{U_0,X})(g_f^{\theta} - g_f^{\theta_0}) + \mathbb{P}_{U_0,X}(g_f^{\theta} - g_f^{\theta_0}),$$

to prove the first result it suffices to show that

$$\sup_{f \in \mathcal{F}_1} |n^{1/2}(\hat{\mathbb{P}}_{U_0,X} - \mathbb{P}_{U_0,X})(g_f^{\theta} - g_f^{\theta_0})| = o_{\mathrm{p}}(1) \tag{A.3}$$

$$\sup_{f \in \mathcal{F}_1} |\mathbb{P}_{U_0,X}(g_f^{\theta} - g_f^{\theta_0}) - D'(f,\theta_0)(\theta - \theta_0)| = o(\|\theta - \theta_0\|). \tag{A.4}$$

We begin with (A.3). Observe that

$$\begin{aligned}
\|g_f^{\theta} - g_f^{\theta_0}\|_{2,\mathbb{P}_{U_0,X}} &= \|\mathbb{1}_{(0,\infty)}(X'\theta) f^{\theta}(U_0,X) - \mathbb{1}_{(0,\infty)}(X'\theta_0) f^{\theta_0}(U_0,X)\|_{2,\mathbb{P}_{U_0,X}} \\
&\leq \|\mathbb{1}_{(0,\infty)}(X'\theta) f^{\theta}(U_0,X) - \mathbb{1}_{(0,\infty)}(X'\theta_0) f^{\theta}(U_0,X)\|_{2,\mathbb{P}_{U_0,X}} \\
&\quad + \|\mathbb{1}_{(0,\infty)}(X'\theta_0) f^{\theta}(U_0,X) - \mathbb{1}_{(0,\infty)}(X'\theta_0) f^{\theta_0}(U_0,X)\|_{2,\mathbb{P}_{U_0,X}} \\
&\leq \|\mathbb{1}_{(0,\infty)}(X'\theta) - \mathbb{1}_{(0,\infty)}(X'\theta_0)\|_{2,\mathbb{P}_X} + \|f^{\theta} - f^{\theta_0}\|_{2,\mathbb{P}_{U_0,X}}.
\end{aligned}$$

Now, since $|\mathbb{1}_{(a,\infty)} - \mathbb{1}_{(b,\infty)}| = \mathbb{1}_{(a\wedge b, a\vee b]}$,

$$\begin{aligned}|\mathbb{1}_{(0,\infty)}(X'\theta) - \mathbb{1}_{(0,\infty)}(X'\theta_0)| &= |\mathbb{1}_{(X'(\theta_0-\theta),\infty)}(X'\theta_0) - \mathbb{1}_{(0,\infty)}(X'\theta_0)| \\ &= \mathbb{1}_{(0\wedge X'(\theta_0-\theta), 0\vee X'(\theta_0-\theta)]}(X'\theta_0).\end{aligned}$$

By (A.6),

$$\begin{aligned}0 \wedge X'(\theta_0-\theta) &= (X'(\theta_0-\theta) - |X'(\theta_0-\theta)|)/2 \geq -\|X\|\|\theta_0-\theta\| \\ 0 \vee X'(\theta_0-\theta) &= (X'(\theta_0-\theta) + |X'(\theta_0-\theta)|)/2 \leq \|X\|\|\theta_0-\theta\|,\end{aligned}$$

implying that $(0\wedge X'(\theta_0-\theta), 0\vee X'(\theta_0-\theta)] \subset (-\|X\|\|\theta-\theta_0\|, \|X\|\|\theta-\theta_0\|]$. Hence,

$$\begin{aligned}\|\mathbb{1}_{(0,\infty)}(X'\theta) - \mathbb{1}_{(0,\infty)}(X'\theta_0)\|^2_{2,\mathbb{P}_X} &= \int \mathbb{1}_{(0\wedge x'(\theta_0-\theta), 0\vee x'(\theta_0-\theta)]}(x'\theta_0)\,\mathbb{P}_X(dx) \\ &\leq \int \mathbb{1}_{(-\|x\|\|\theta-\theta_0\|, \|x\|\|\theta-\theta_0\|]}(x'\theta_0)\,\mathbb{P}_X(dx).\end{aligned}$$

Assume that

Assumption A.3. *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,

$$\begin{aligned}\int \mathbb{1}_{(-\|x\|\|\theta-\theta_0\|, \|x\|\|\theta-\theta_0\|]}(x'\theta_0)\,\mathbb{P}_X(dx) &= \int \mathbb{1}_{(-q_1\|\theta-\theta_0\|, q_1\|\theta-\theta_0\|]}(q_2)\,\mathbb{P}_{Q_2,Q_1}(dq_2, dq_1) \\ &\leq \int \xi(q_1) \int \mathbb{1}_{(-q_1\|\theta-\theta_0\|, q_1\|\theta-\theta_0\|]}(q_2)\,dq_2\,\mathbb{P}_{Q_1}(dq_1) \\ &= 2\|\theta-\theta_0\| \int \xi(q_1)q_1\,\mathbb{P}_{Q_1}(dq_1). \qquad \text{(A.5)}\end{aligned}$$

Therefore,

$$\|\mathbb{1}_{(0,\infty)}(X'\theta) - \mathbb{1}_{(0,\infty)}(X'\theta_0)\|_{2,\mathbb{P}_X} \leq \text{const.}\,\|\theta-\theta_0\|^{1/2}.$$

Let $b(u_0, x, \theta, \theta_0) := V(u_0, x, \theta) - V(u_0, x, \theta_0)$ and $f \in \mathcal{F}_1$ so that $f = \mathbb{1}_{(-\infty,s]\times(-\infty,t]}$ for some $(s,t) \in \mathbb{R}\times\mathbb{R}^{\dim(X)}$. Hence, recalling that $f^\theta(u_0, x) = f(V(u_0, x, \theta_0) + b(u_0, x, \theta, \theta_0), x)$, we can write $f^\theta(u_0, x) = \mathbb{1}_{(-\infty, s-b(u_0,x,\theta,\theta_0)]}(V(u_0, x, \theta_0))\mathbb{1}_{(-\infty,t]}(x)$. Now,

$$((s - b(u_0, x, \tilde\theta, \theta_0)) \wedge s, (s - b(u_0, x, \tilde\theta, \theta_0)) \vee s] \subset (s - \|x\|\|\theta-\theta_0\|, s + \|x\|\|\theta-\theta_0\|]$$

because $\sup_{\tilde\theta\in\overline{B(\theta_0,\|\theta-\theta_0\|)}}|b(U_0,X,\tilde\theta,\theta_0)| \le \|X\|\,\|\theta-\theta_0\|$ by equation A.11 of Powell's paper (this result does not require the conditional symmetry of $U_0|X$). Hence, by (2.4),

$$\begin{aligned}
&\|f^\theta - f^{\theta_0}\|^2_{2,\mathbb{P}_{U_0,X}} \\
&\quad\le \int\int_{-x'\theta_0}^{x'\theta_0+2\Delta(x,\theta,\theta_0)} \mathbb{1}_{(s-\|x\|\|\theta-\theta_0\|,s+\|x\|\|\theta-\theta_0\|]}(u_0-\Delta(x,\theta,\theta_0))p_{U_0|X=x}(u_0)\,du_0\,\mathbb{P}_X(dx) \\
&\qquad+\int \mathbb{1}_{(s-\|x\|\|\theta-\theta_0\|,s+\|x\|\|\theta-\theta_0\|]}(-x'\theta_0-\Delta(x,\theta,\theta_0))\,\mathbb{P}_X(dx) \\
&\qquad+\int \mathbb{1}_{(s-\|x\|\|\theta-\theta_0\|,s+\|x\|\|\theta-\theta_0\|]}(x'\theta_0+\Delta(x,\theta,\theta_0))\,\mathbb{P}_X(dx) \\
&\quad\le \int\int \mathbb{1}_{(s-2\|x\|\|\theta-\theta_0\|,s+2\|x\|\|\theta-\theta_0\|]}(u_0)p_{U_0|X=x}(u_0)\,du_0\,\mathbb{P}_X(dx) \\
&\qquad+\int \mathbb{1}_{(s-2\|x\|\|\theta-\theta_0\|,s+2\|x\|\|\theta-\theta_0\|]}(-x'\theta_0)\,\mathbb{P}_X(dx)+\int \mathbb{1}_{(s-2\|x\|\|\theta-\theta_0\|,s+2\|x\|\|\theta-\theta_0\|]}(x'\theta_0)\,\mathbb{P}_X(dx) \\
&\quad=: a_1^\theta(s)+a_2^\theta(s)+a_3^\theta(s).
\end{aligned}$$

Assume that

Assumption A.4. *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,

$$\sup_{s\in\mathbb{R}} a_1^\theta(s) \le 4\|\theta-\theta_0\|\int\zeta(x)\|x\|\,\mathbb{P}_X(dx).$$

Similarly, following the argument leading to (A.5), it can be shown that

$$\sup_{s\in\mathbb{R}} a_2^\theta(s)\vee\sup_{s\in\mathbb{R}} a_3^\theta(s)\le \text{const.}\,\|\theta-\theta_0\|.$$

Therefore, we have shown that

$$\sup_{f\in\mathcal{F}_1}\|g_f^\theta-g_f^{\theta_0}\|_{2,\mathbb{P}_{U_0,X}}\le\text{const.}\,\|\theta-\theta_0\|^{1/2}.$$

Thus, the empirical process $\{n^{1/2}(\hat{\mathbb{P}}_{U_0,X}-\mathbb{P}_{U_0,X})(\mathbb{1}_{(0,\infty)}(X'\theta_0)f):f\in\mathcal{F}_1\}$ is asymptotically equicontinuous because $\mathcal{F}_1$ is $\mathbb{P}_{U_0,X}$-Donsker. Hence, the argument leading to (A.5) of CT goes through unchanged so that we obtain (A.3).

To verify (A.4), begin by observing that

$$\mathbb{P}_{U_0,X}g_f^\theta=\int_{\text{supp}(X)}\mathbb{1}_{(0,\infty)}(x'\theta)\int_{\mathbb{R}}f(V(u_0,x,\theta),x)p_{U_0|X=x}(u_0)\,du_0\,\mathbb{P}_X(dx).$$

Now,

$$\mathbb{1}_{(0,\infty)}(x'\theta) - \mathbb{1}_{(\eta_0,\infty)}(x'\theta_0) = \begin{cases} 0 & \text{if } x'\theta \leq 0 \cap x'\theta_0 \leq \eta_0 \text{ or } x'\theta > 0 \cap x'\theta_0 > \eta_0 \\ -1 & \text{if } x'\theta \leq 0 \text{ and } x'\theta_0 > \eta_0 \\ 1 & \text{if } x'\theta > 0 \text{ and } x'\theta_0 \leq \eta_0. \end{cases}$$

It can be shown, cf. Lemma A.1, that the sets $\{x \in \operatorname{supp}(X) : (x'\theta > 0) \cap (x'\theta_0 \leq \eta_0)\}$ and $\{x \in \operatorname{supp}(X) : (x'\theta \leq 0) \cap (x'\theta_0 > \eta_0)\}$ have zero $\mathbb{P}_X$-measure when θ is close enough to θ_0. Therefore, for θ in a small enough neighborhood of θ_0, we have

$$\begin{aligned}\mathbb{P}_{U_0,X} g_f^\theta &= \int_{\operatorname{supp}(X)} \mathbb{1}_{(\eta_0,\infty)}(x'\theta_0) \int_{\mathbb{R}} f(V(u_0,x,\theta),x) p_{U_0|X=x}(u_0)\, du_0\, \mathbb{P}_X(dx) \\ &= \int_{\operatorname{supp}(X)} \int_{\mathbb{R}} f(V(u_0,x,\theta),x) p_{U_0|X=x}(u_0)\, du_0\, \mathbb{P}_X(dx)\end{aligned}$$

because $\mathbb{P}_X(X'\theta_0 \in (-\infty,\eta_0]) = 0$ by assumption. Hence, with $r_1(x) := \mathbb{P}_{U_0|X=x}(U_0 \leq -x'\theta_0)$, $r_{2,\Delta}(x) := \mathbb{P}_{U_0|X=x}(U_0 \geq x'\theta_0 + 2\Delta(x,\theta,\theta_0))$, and using the definition of $V(u_0,x,\theta)$,

$$\begin{aligned}\mathbb{P}_{U_0,X} g_f^\theta = &\int_{\operatorname{supp}(X)} f(-x'\theta_0 - \Delta(x,\theta,\theta_0),x) r_1(x)\, \mathbb{P}_X(dx) \\ &+ \int_{\operatorname{supp}(X)} \int_{(-x'\theta_0, x'\theta_0 + 2\Delta(x,\theta,\theta_0))} f(u_0 - \Delta(x,\theta,\theta_0),x) p_{U_0|X=x}(u_0)\, du_0\, \mathbb{P}_X(dx) \\ &+ \int_{\operatorname{supp}(X)} f(x'\theta_0 + \Delta(x,\theta,\theta_0),x) r_{2,\Delta}(x)\, \mathbb{P}_X(dx).\end{aligned}$$

Since $u_0 \mapsto f(u_0 - \Delta(x,\theta,\theta_0),x)$ is bounded on $\mathbb{R}$, it is Riemann integrable with respect to the integrator $p_{U_0|X=x}$ on the interval $(-x'\theta_0, x'\theta_0 + 2\Delta(x,\theta,\theta_0))$. Consequently, employing the traditional notation for Riemann integrals, we have that

$$\begin{aligned}&\int_{(-x'\theta_0, x'\theta_0 + 2\Delta(x,\theta,\theta_0))} f(u_0 - \Delta(x,\theta,\theta_0),x) p_{U_0|X=x}(u_0)\, du_0 \\ &\qquad = \int_{-x'\theta_0}^{x'\theta_0 + 2\Delta(x,\theta,\theta_0)} f(u_0 - \Delta(x,\theta,\theta_0),x) p_{U_0|X=x}(u_0)\, du_0 \\ &\qquad = \int_{-x'\theta_0 - \Delta(x,\theta,\theta_0)}^{x'\theta_0 + \Delta(x,\theta,\theta_0)} f(u_0,x) p_{U_0|X=x}(u_0 + \Delta(x,\theta,\theta_0))\, du_0,\end{aligned}$$

where the last equality follows by a change of variables. Since Riemann integrals are orientable, i.e., $\int_a^b = -\int_b^a$ irrespective of the ordering of a and b, it follows that the decomposition

$$\int_{-x'\theta_0 - \Delta(x,\theta,\theta_0)}^{x'\theta_0 + \Delta(x,\theta,\theta_0)} = \int_{-x'\theta_0 - \Delta(x,\theta,\theta_0)}^{-x'\theta_0} + \int_{-x'\theta_0}^{x'\theta_0} + \int_{x'\theta_0}^{x'\theta_0 + \Delta(x,\theta,\theta_0)}$$

holds irrespective of whether $\Delta(x,\theta,\theta_0) > 0$ or $\Delta(x,\theta,\theta_0) \le 0$. Therefore,

$$\begin{aligned}
\mathbb{P}_{U_0,X} g_f^\theta = &\int_{\text{supp}(X)} f(-x'\theta_0 - \Delta(x,\theta,\theta_0), x) r_1(x)\, \mathbb{P}_X(dx) \\
&+ \int_{\text{supp}(X)} \int_{-x'\theta_0-\Delta(x,\theta,\theta_0)}^{-x'\theta_0} f(u_0,x) p_{U_0|X=x}(u_0 + \Delta(x,\theta,\theta_0))\, du_0\, \mathbb{P}_X(dx) \\
&+ \int_{\text{supp}(X)} \int_{-x'\theta_0}^{x'\theta_0} f(u_0,x) p_{U_0|X=x}(u_0 + \Delta(x,\theta,\theta_0))\, du_0\, \mathbb{P}_X(dx) \\
&+ \int_{\text{supp}(X)} \int_{x'\theta_0}^{x'\theta_0+\Delta(x,\theta,\theta_0)} f(u_0,x) p_{U_0|X=x}(u_0 + \Delta(x,\theta,\theta_0))\, du_0\, \mathbb{P}_X(dx) \\
&+ \int_{\text{supp}(X)} f(x'\theta_0 + \Delta(x,\theta,\theta_0), x) r_{2,\Delta}(x)\, \mathbb{P}_X(dx) \\
&=: T_1^\theta(f) + T_2^\theta(f) + T_3^\theta(f) + T_4^\theta(f) + T_5^\theta(f).
\end{aligned}$$

Hence, since $T_2^{\theta_0}(f) = T_4^{\theta_0}(f) = 0$,

$$\mathbb{P}_{U_0,X}(g_f^\theta - g_f^{\theta_0}) = [T_1^\theta(f) - T_1^{\theta_0}(f)] + T_2^\theta(f) + [T_3^\theta(f) - T_3^{\theta_0}(f)] + T_4^\theta(f) + [T_5^\theta(f) - T_5^{\theta_0}(f)].$$

Let $T_{2,a}^\theta(f) := \int \int_{-x'\theta_0-\Delta(x,\theta,\theta_0)}^{-x'\theta_0} f(u_0,x) p_{U_0|X=x}(u_0)\, du_0\, \mathbb{P}_X(dx)$ (note that $T_{2,a}^{\theta_0}(f) = 0$), and assume that[19]

Assumption A.5. *There exists a neighborhood of θ_0 on which $\theta \mapsto T_1^\theta(f)$ and $\theta \mapsto T_{2,a}^\theta(f)$ are twice differentiable in θ, uniformly in f, i.e., for θ in this neighborhood,*

$$\begin{aligned}
T_1^\theta(f) &= T_1^{\theta_0}(f) + (\nabla_\theta T_1^{\theta_0}(f))'(\theta - \theta_0) + O(\|\theta - \theta_0\|^2) \\
T_{2,a}^\theta(f) &= (\nabla_\theta T_{2,a}^{\theta_0}(f))'(\theta - \theta_0) + O(\|\theta - \theta_0\|^2),
\end{aligned}$$

where the $O(\|\theta - \theta_0\|^2)$ term does not depend upon $f \in \mathcal{F}_1$.

By this assumption, $T_1^\theta(f) - T_1^{\theta_0}(f) = (\nabla_\theta T_1^{\theta_0}(f))'(\theta - \theta_0) + O(\|\theta - \theta_0\|^2)$ holds uniformly in $f \in \mathcal{F}_1$. Next, assume that

Assumption A.6. *$u_0 \mapsto \mathbb{P}_{U_0|X=x}(U_0 \ge u_0)$ is twice differentiable with the second derivative uniformly bounded in $(t,x) \in \mathbb{R} \times \text{supp}(X)$ by some positive number K.*

Under this assumption,

$$|r_{2,\Delta}(x) - r_{2,0}(x) + 2p_{U_0|X=x}(x'\theta_0)\Delta(x,\theta,\theta_0)| \le K\Delta^2(x,\theta,\theta_0),$$

where $r_{2,0}(x) := \mathbb{P}_{U_0|X=x}(U_0 \ge x'\theta_0)$. Consequently, we can write

$$T_5^\theta(f) = T_{5,a}^\theta(f) - 2(T_{5,b}^\theta(f))'(\theta - \theta_0) + O(\|\theta - \theta_0\|^2),$$

[19]This "high level" differentiability assumption can be shown to hold under more primitive regularity conditions; cf. the remarks at the end of this appendix.

where the $O(\|\theta-\theta_0\|^2)$ term does not depend upon f and

$$T_{5,a}^{\theta}(f) := \int_{\mathrm{supp}(X)} f(x'\theta_0+\Delta(x,\theta,\theta_0),x) r_{2,0}(x)\,\mathbb{P}_X(dx)$$
$$T_{5,b}^{\theta}(f) := \int_{\mathrm{supp}(X)} f(x'\theta_0+\Delta(x,\theta,\theta_0),x) p_{U_0|X=x}(x'\theta_0) x\,\mathbb{P}_X(dx).$$

As with $T_1^{\theta}(f)$, Assumption A.5 implies that $\theta\mapsto T_{5,a}^{\theta}$ is twice differentiable in a neighborhood of θ_0, uniformly in f, so that $T_{5,a}^{\theta}(f) = T_{5,a}^{\theta_0}(f) + (\nabla_\theta T_{5,a}^{\theta_0}(f))'(\theta-\theta_0) + O(\|\theta-\theta_0\|^2)$ and the $O(\|\theta-\theta_0\|^2)$ term does not depend upon f. Similarly, $\theta\mapsto T_{5,b}^{\theta}$ is differentiable in a neighborhood of θ_0 with the derivative uniformly bounded in f. Hence, by an element by element mean-value expansion, $T_{5,b}^{\theta}(f) = T_{5,b}^{\theta_0}(f) + O(\|\theta-\theta_0\|)$ in some neighborhood of θ_0, where the $O(\|\theta-\theta_0\|)$ term does not depend upon f. It follows that, in a neighborhood of θ_0,

$$T_5^{\theta}(f) - T_5^{\theta_0}(f) = (\nabla_\theta T_5^{\theta_0}(f))'(\theta-\theta_0) + O(\|\theta-\theta_0\|^2),$$

where $\nabla_\theta T_5^{\theta_0}(f) := \nabla_\theta T_{5,a}^{\theta_0}(f) - 2\int_{\mathrm{supp}(X)} \mathbb{1}_{(\eta_0,\infty)}(x'\theta_0) f(x'\theta_0,x) p_{U_0|X=x}(x'\theta_0) x\,\mathbb{P}_X(dx)$ and the $O(\|\theta-\theta_0\|^2)$ term does not depend upon $f\in\mathcal{F}_1$.

Next, assume that

Assumption A.7. $(u_0,x)\mapsto \partial p_{U_0|X=x}(u_0)/\partial u_0$ *is uniformly bounded on* $\mathbb{R}\times\mathrm{supp}(X)$.

Then, $T_2^{\theta}(f) = T_{2,a}^{\theta}(f) + O(\|\theta-\theta_0\|^2)$, where and the $O(\|\theta-\theta_0\|^2)$ term does not depend upon $f\in\mathcal{F}_1$. Hence, by Assumption A.5 and the fact that $T_{2,a}^{\theta_0}(f)=0$,

$$T_2^{\theta}(f) = (\nabla_\theta T_{2,a}^{\theta_0}(f))'(\theta-\theta_0) + O(\|\theta-\theta_0\|^2),$$

where the $O(\|\theta-\theta_0\|^2)$ term does not depend upon $f\in\mathcal{F}_1$.

Letting $T_{4,a}^{\theta}(f) := \int\int_{x'\theta_0}^{x'\theta_0+\Delta(x,\theta,\theta_0)} f(u_0,x) p_{U_0|X=x}(u_0+\Delta(x,\theta,\theta_0))\,du_0\,\mathbb{P}_X(dx)$ and arguing similarly, we have that

$$T_4^{\theta}(f) = (\nabla_\theta T_{4,a}^{\theta_0}(f))'(\theta-\theta_0) + O(\|\theta-\theta_0\|^2),$$

uniformly in $f\in\mathcal{F}_1$.

Finally, assume that

Assumption A.8. $u_0\mapsto p_{U_0|X=x}(u_0)$ *is differentiable a.e. on* $\mathbb{R}$ *and the conditional second moment of* $\partial p_{U_0|X=x}(u_0)/\partial u_0$ *is uniformly bounded.*

Under this assumption, the square-root density $p_{U_0|X=x}^{1/2}$ is mean-square differentiable. Hence, as with (A.6) of CT, it can be shown that in a small enough neighborhood of θ_0,

$$T_3^{\theta}(f) - T_3^{\theta_0}(f) = (\nabla_\theta T_3^{\theta_0}(f))'(\theta-\theta_0) + o(\|\theta-\theta_0\|),$$

where $\nabla_\theta T_3^{\theta_0}(f) := \mathbb{E}[\mathbb{1}_{(-X'\theta_0,X'\theta_0)}(U_0)(\partial\log p_{U_0,X}(U_0,X)/\partial U_0) f(U_0,X) X]$ and the $o(\|\theta-\theta_0\|)$ term does not depend upon $f\in\mathcal{F}_1$.

Combining the expansions for $T_1^\theta(f) - T_5^\theta(f)$, we have that, in a neighborhood of θ_0,

$$\sup_{f\in\mathcal{F}_1} |\mathbb{P}_{U_0,X}(g_f^\theta - g_f^{\theta_0}) - D'(f,\theta_0)(\theta-\theta_0)| = o(\|\theta-\theta_0\|)$$

with $D(f,\theta_0) := \nabla_\theta T_1^{\theta_0}(f) + \nabla_\theta T_{2,a}^{\theta_0}(f) + \nabla_\theta T_3^{\theta_0}(f) + \nabla_\theta T_{4,a}^{\theta_0}(f) + \nabla_\theta T_5^{\theta_0}(f)$. Hence, (A.4) holds as well. The desired result follows. □

The following result may be of independent interest.

Lemma A.1. *Assume that* $\mathbb{P}_X(X'\theta_0 > \eta_0) = 1$ *for some* $\eta_0 > 0$, *and* $\operatorname{supp}(X)$ *is bounded, i.e.,* $\operatorname{supp}(X) \subset B(0_{\dim(X)\times 1}, M)$ *for some* $M > 0$. *Let* $S_\theta := \{x \in \operatorname{supp}(X) : (x'\theta > 0)\cap(x'\theta_0 \leq \eta_0)\}$ *and* $T_\theta := \{x \in \operatorname{supp}(X) : (x'\theta \leq 0)\cap(x'\theta_0 > \eta_0)\}$. *Then,* $\mathbb{P}_X(S_\theta) = 0$ *and* $\theta \in B(\theta_0, \eta_0/M) \implies \mathbb{P}_X(T_\theta) = 0$.

Proof of Lemma A.1. The results are obvious if $\theta = \theta_0$. So assume that $\theta \neq \theta_0$. The first result follows because $S_\theta \subset \{x \in \operatorname{supp}(X) : x'\theta_0 \leq \eta_0\}$ and $\mathbb{P}_X(X'\theta_0 > \eta_0) = 1$ by assumption. To show the second result, observe that

$$x \in T_\theta \implies 0 \geq x'\theta = x'\theta_0 + x'(\theta-\theta_0) \overset{\text{(A.6)}}{\geq} x'\theta_0 - \|x\|\|\theta-\theta_0\| > \eta_0 - \|x\|\|\theta-\theta_0\|,$$

because, by the Cauchy-Schwarz inequality,

$$|x'(\theta-\theta_0)| \leq \|x\|\|\theta-\theta_0\| \iff -\|x\|\|\theta-\theta_0\| \leq x'(\theta-\theta_0) \leq \|x\|\|\theta-\theta_0\|. \tag{A.6}$$

Therefore, since $\|x\| \leq M$,

$$T_\theta \subset \{x \in \operatorname{supp}(X) : M\|\theta-\theta_0\| > \eta_0\} = \begin{cases} \operatorname{supp}(X) & \text{if } \|\theta-\theta_0\| > \eta_0/M \\ \varnothing & \text{otherwise.} \end{cases}$$

Hence, $\mathbb{P}_X(T_\theta) \leq \mathbb{1}(\|\theta-\theta_0\| > \eta_0/M)$. Thus, $\theta \in B(\theta_0, \eta_0/M) \implies \mathbb{P}_X(T_\theta) = 0$. □

Proof of (3.4). Since $RU|X \overset{\text{d}}{=} -RU|X$, the density of $RU|X = x$ with respect to the Lebesgue measure on $\mathbb{R}$ and the counting measure on $\{-x'\theta_0, x'\theta_0\}$ is given by

$$\operatorname{pdf}_{RU|X=x}(u) := 0.5[\operatorname{pdf}_{U|X=x}(u) + \operatorname{pdf}_{U|X=x}(-u)], \qquad u \in \mathbb{R},$$

where $\operatorname{pdf}_{U|X=x}(u) := p_{U_0|X=x}(u)\mathbb{1}_{(-x'\theta_0,\infty)}(u) + \Pr(U_0 \leq -x'\theta_0|X=x)\mathbb{1}_{\{-x'\theta_0\}}(u)$ is the conditional density of U given $X = x$ (with respect to the Lebesgue measure on $\mathbb{R}$ and the counting measure on $-x'\theta_0$). Therefore,

$$\operatorname{pdf}_{RU|X=x}(u) = \begin{cases} 0.5p_{U_0|X=x}(-u) & \text{if } u \in (-\infty, -x'\theta_0) \\ 0.5[p_{U_0|X=x}(-u) + p_{U_0|X=x}(u)] & \text{if } u \in (-x'\theta_0, x'\theta_0) \\ 0.5p_{U_0|X=x}(u) & \text{if } u \in (x'\theta_0, \infty) \\ 0.5\Pr(U_0 \leq -x'\theta_0|X=x) & \text{if } u \in \{-x'\theta_0, x'\theta_0\}. \end{cases}$$

If $U_0|X \overset{d}{=} -U_0|X$, i.e., $p_{U_0|X=x}(u) = p_{U_0|X=x}(-u)$ for all $(u,x) \in \mathbb{R} \times \operatorname{supp}(X)$, then

$$\operatorname{pdf}_{RU|X=x}(u) = \begin{cases} 0.5 p_{U_0|X=x}(u) & \text{if } u \in (-\infty, -x'\theta_0) \cup (x'\theta_0, \infty) \\ p_{U_0|X=x}(u) & \text{if } u \in (-x'\theta_0, x'\theta_0) \\ 0.5 \Pr(U_0 \le -x'\theta_0 | X = x) & \text{if } u \in \{-x'\theta_0, x'\theta_0\}. \end{cases}$$

The desired result follows. □

Remarks. We now demonstrate that the linear expansions in Assumption A.5 can be shown to hold under more primitive regularity conditions. Throughout these remarks, $f \in \mathcal{F}_1$ so that $f = \mathbb{1}_{(-\infty,s]\times(-\infty,t]}$ for some $s \in \mathbb{R}$ and $t \in \mathbb{R}^{\dim(X)}$.

(i) We deal with $T^{\theta}_{2,a}(f)$ first because it is a little easier to handle than $T^{\theta}_{1}(f)$. Begin by observing that $T^{\theta}_{2,a}(f) = \int l^{\theta}(s,x) \mathbb{1}_{(-\infty,t]}(x)\, \mathbb{P}_X(dx)$ with

$$\begin{aligned} l^{\theta}(s,x) &:= \int_{-x'\theta}^{-x'\theta_0} \mathbb{1}_{(-\infty,s]}(u_0)\, p_{U_0|X=x}(u_0) \\ &= \int_{-x'\theta}^{s} \mathbb{1}_{(-\infty,s]}(u_0)\, p_{U_0|X=x}(u_0) + \int_{s}^{-x'\theta_0} \mathbb{1}_{(-\infty,s]}(u_0)\, p_{U_0|X=x}(u_0). \end{aligned}$$

Note that $l^{\theta_0}(s,x) = 0$, so that $T^{\theta_0}_{2,a}(f) = 0$, and

$$\begin{aligned} \int_{-x'\theta}^{s} \mathbb{1}_{(-\infty,s]}(u_0)\, p_{U_0|X=x}(u_0)\, du_0 &= \begin{cases} \int_{-x'\theta}^{s} \mathbb{1}_{(-\infty,s]}(u_0)\, p_{U_0|X=x}(u_0)\, du_0 & \text{if } -x'\theta < s \\ 0 & \text{if } -x'\theta = s \\ \int_{s}^{-x'\theta} \mathbb{1}_{(-\infty,s]}(u_0)\, p_{U_0|X=x}(u_0)\, du_0 & \text{if } -x'\theta > s \end{cases} \\ &= \begin{cases} \int_{-x'\theta}^{s} p_{U_0|X=x}(u_0)\, du_0 & \text{if } -x'\theta < s \\ 0 & \text{if } -x'\theta \ge s. \end{cases} \end{aligned}$$

Assume that there exists a set $A \subset \operatorname{supp}(X)$ such that $\mathbb{P}_X(A) = 1$ and for all $x \in A$ the function $u_0 \mapsto p_{U_0|X=x}(u_0)$ is continuous on $\mathbb{R}$. Then, it is easy to see that, for all $x \in A$, $m \mapsto \int_m^s \mathbb{1}_{(-\infty,s]}(u_0)\, p_{U_0|X=x}(u_0)\, du_0$ is differentiable on $\mathbb{R} \setminus \{s\}$ with

$$\frac{\partial}{\partial m} \int_m^s \mathbb{1}_{(-\infty,s]}(u_0)\, p_{U_0|X=x}(u_0)\, du_0 = \begin{cases} p_{U_0|X=x}(s) & \text{if } m < s \\ 0 & \text{if } m > s. \end{cases}$$

Hence, by the chain rule, for all $x \in A$, the map $\theta \mapsto l^{\theta}(s,x)$ is differentiable on $\mathbb{R}^{\dim(\theta_0)}$, except when $-x'\theta = s$, and

$$\frac{\partial}{\partial \theta} l^{\theta}(s,x) = \begin{cases} -x p_{U_0|X=x}(s) & \text{if } -x'\theta < s \\ 0 & \text{if } -x'\theta > s. \end{cases}$$

Assume that for all θ in a small enough neighborhood, say N_{θ_0}, of θ_0, the random variable $X'\theta$ is continuously distributed. Consequently, assuming that the operations of differentiation and integration can be interchanged and keeping in mind that $\mathbb{P}_X(-X'\theta = s) = 0$ for all $\theta \in N_{\theta_0}$,

$$\begin{aligned}
\theta \in N_{\theta_0} \implies \nabla_\theta T^\theta_{2,a}(f) &= \int_A [\frac{\partial}{\partial\theta} l^\theta(s,x)] \mathbb{1}_{(-\infty,t]}(x)\, \mathbb{P}_X(dx) \\
&= \int \mathbb{1}(-x'\theta < s)[-x p_{U_0|X=x}(s)] \mathbb{1}_{(-\infty,t]}(x)\, \mathbb{P}_X(dx) \\
&= -\int x \mathbb{1}_{(-\infty,s]}(-x'\theta) \mathbb{1}_{(-\infty,t]}(x) p_{U_0|X=x}(s)\, \mathbb{P}_X(dx) \\
&= -\int x f(-x'\theta, x) p_{U_0|X=x}(s)\, \mathbb{P}_X(dx) =: -Q(\theta, f).
\end{aligned}$$

Since the "bad" set of x's where $\theta \mapsto l^\theta(s,x)$ is not differentiable has $\mathbb{P}_X$-measure zero (even though it depends upon s), it does not affect the differentiability of $\theta \mapsto T^\theta_{2,a}(f)$.

Next, we show that under certain conditions, the map $\theta \mapsto Q(\theta, f)$ is uniformly (in f) Lipschitz at θ_0. Suppose that $\dim(\theta_0) = 2$ and both coordinates of θ_0 are positive. Assume that N_{θ_0} is small enough so that both coordinates of $\theta \in N_{\theta_0}$ are positive as well. Assume that $g(u_0) := \sup_{x \in \mathrm{supp}(X)} p_{U_0|X=x}(u_0)$ is well defined on $\mathbb{R}$ and that $\sup_{x\in\mathrm{supp}(X)} \|x\| < M$. Then,

$$\|Q(\theta, f) - Q(\theta_0, f)\| \le M g(s) \int h_s(\theta, x^{(1)})\, \mathrm{pdf}_{X^{(1)}}(x^{(1)})\, dx^{(1)},$$

where $h_s(\theta, x^{(1)}) := \int |\mathbb{1}_{(\infty,s]}(-x^{(1)}\theta^{(1)} - x^{(2)}\theta^{(2)}) - \mathbb{1}_{(\infty,s]}(-x^{(1)}\theta_0^{(1)} - x^{(2)}\theta_0^{(2)})|\, \mathrm{pdf}_{X^{(2)}|X^{(1)}=x^{(1)}}(x^{(2)})\, dx^{(2)}$. Since $|\mathbb{1}_{[a,\infty)} - \mathbb{1}_{[b,\infty)}| = \mathbb{1}_{[a\wedge b, a\vee b)}$,

$$\begin{aligned}
&|\mathbb{1}_{(\infty,s]}(-x^{(1)}\theta^{(1)} - x^{(2)}\theta^{(2)}) - \mathbb{1}_{(\infty,s]}(-x^{(1)}\theta_0^{(1)} - x^{(2)}\theta_0^{(2)})| \\
&\qquad = |\mathbb{1}_{[-\frac{s+x^{(1)}\theta^{(1)}}{\theta^{(2)}},\infty)}(x^{(2)}) - \mathbb{1}_{[-\frac{s+x^{(1)}\theta_0^{(1)}}{\theta_0^{(2)}},\infty)}(x^{(2)})| \\
&\qquad = \mathbb{1}_{[-\frac{s+x^{(1)}\theta^{(1)}}{\theta^{(2)}} \wedge -\frac{s+x^{(1)}\theta_0^{(1)}}{\theta_0^{(2)}}, -\frac{s+x^{(1)}\theta^{(1)}}{\theta^{(2)}} \vee -\frac{s+x^{(1)}\theta_0^{(1)}}{\theta_0^{(2)}})}(x^{(2)}).
\end{aligned}$$

Assume that $\mathrm{cdf}_{X^{(2)}|X^{(1)}}$ is Lipschitz, i.e., $|\mathrm{cdf}_{X^{(2)}|X^{(1)}=x^{(1)}}(a) - \mathrm{cdf}_{X^{(2)}|X^{(1)}=x^{(1)}}(b)| \le \zeta(x^{(1)})|a-b|$ with $\int \zeta(x^{(1)})\, \mathrm{pdf}_{X^{(1)}}(x^{(1)})\, dx^{(1)} < \infty$. Hence, since $a \vee b - a \wedge b = |a-b|$,

$$\begin{aligned}
h_s(\theta, x^{(1)}) &\le \int_{[-\frac{s+x^{(1)}\theta^{(1)}}{\theta^{(2)}} \wedge -\frac{s+x^{(1)}\theta_0^{(1)}}{\theta_0^{(2)}}, -\frac{s+x^{(1)}\theta^{(1)}}{\theta^{(2)}} \vee -\frac{s+x^{(1)}\theta_0^{(1)}}{\theta_0^{(2)}})} \mathrm{pdf}_{X^{(2)}|X^{(1)}=x^{(1)}}(x^{(2)})\, dx^{(2)} \\
&\le \zeta(x^{(1)}) |\frac{s + x^{(1)}\theta^{(1)}}{\theta^{(2)}} - \frac{s + x^{(1)}\theta_0^{(1)}}{\theta_0^{(2)}}| \\
&\le \|\zeta\|_{1,\mathbb{P}_{X^{(1)}}} \frac{|s| + |x^{(1)}|(\theta_0^{(1)} + \theta_0^{(2)})}{\theta^{(2)}\theta_0^{(2)}} \|\theta - \theta_0\|.
\end{aligned}$$

Therefore, assuming that $\sup_{u_0\in\mathbb{R}} |u_0 g(u_0)| + g(u_0) < \infty$ and $\inf\{\theta^{(2)} : (\theta_0^{(1)}, \theta^{(2)}) \in N_{\theta_0}\} > 0$,

$$\|Q(\theta,f) - Q(\theta_0,f)\| \leq M\frac{|sg(s)| + Mg(s)(\theta_0^{(1)} + \theta_0^{(2)})}{\theta^{(2)}\theta_0^{(2)}}\|\theta-\theta_0\| \leq K\|\theta-\theta_0\|$$

holds for all $\theta \in N_{\theta_0}$, where the constant K does not depend upon s or θ. Hence, we have that

$$\theta \in N_{\theta_0} \implies \sup_{f\in\mathcal{F}_1} \|Q(\theta,f) - Q(\theta_0,f)\| \leq K\|\theta-\theta_0\|.$$

Consequently, by a mean value expansion, there exists a $\lambda \in (0,1)$ such that

$$\begin{aligned}\theta \in N_{\theta_0} \implies T_{2,a}^{\theta}(f) &= (\nabla_\theta T_{2,a}^{\lambda\theta+(1-\lambda)\theta_0}(f))'(\theta-\theta_0)\\ &= (\nabla_\theta T_{2,a}^{\theta_0}(f))'(\theta-\theta_0) - (Q(\lambda\theta+(1-\lambda)\theta_0, f) - Q(\theta_0,f))'(\theta-\theta_0)\\ &= (\nabla_\theta T_{2,a}^{\theta_0}(f))'(\theta-\theta_0) + O(\|\theta-\theta_0\|^2),\end{aligned}$$

where the $O(\|\theta-\theta_0\|^2)$ term does not depend upon f. Hence, a uniform linear expansion for $T_{2,a}^{\theta}$ holds under certain regularity conditions.

(ii) Next, we look at $T_1^{\theta}(f)$. First, note that

$$T_1^{\theta}(f) = \int f(-x'\theta_0 - \Delta(x,\theta,\theta_0), x) r_1(x)\,\mathbb{P}_X(dx) = \int f(-x'\theta, x) r_1(x)\,\mathbb{P}_X(dx).$$

Next, suppose $\dim(X) = 2$, and both coordinates of θ_0 are positive. Assume that $\theta = (\theta^{(1)}, \theta^{(2)})$ is close enough to θ_0 so that both its coordinates are positive as well. Assume that $X^{(1)}$ and $X^{(2)}$ are continuously distributed. Then,

$$\begin{aligned}T_1^{\theta}(f) &= \int\int \mathbb{1}_{(\infty,s]}(-x^{(1)}\theta^{(1)} - x^{(2)}\theta^{(2)})\mathbb{1}_{(-\infty,t^{(1)}]}(x^{(1)})\mathbb{1}_{(-\infty,t^{(2)}]}(x^{(2)}) r_1(x^{(1)}, x^{(2)})\\ &\qquad\qquad \mathrm{pdf}_{X^{(2)}|X^{(1)}=x^{(1)}}(x^{(2)})\,\mathrm{pdf}_{X^{(1)}}(x^{(1)})\,dx^{(2)}\,dx^{(1)}\\ &= \int \mathbb{1}_{(-\infty,t^{(1)}]}(x^{(1)}) h_s(\theta, x^{(1)})\,\mathrm{pdf}_{X^{(1)}}(x^{(1)})\,dx^{(1)},\end{aligned}$$

where $h_s(\theta, x^{(1)}) := \int \mathbb{1}_{[-\frac{s+x^{(1)}\theta^{(1)}}{\theta^{(2)}}, t^{(2)}]}(x^{(2)}) r_1(x^{(1)}, x^{(2)})\ \mathrm{pdf}_{X^{(2)}|X^{(1)}=x^{(1)}}(x^{(2)})\,dx^{(2)}$. Note that

$$h_s(\theta, x^{(1)}) = \begin{cases}\int_{-\frac{s+x^{(1)}\theta^{(1)}}{\theta^{(2)}}}^{t^{(2)}} r_1(x^{(1)}, x^{(2)})\ \mathrm{pdf}_{X^{(2)}|X^{(1)}=x^{(1)}}(x^{(2)})\,dx^{(2)} & \text{if } -\frac{s+x^{(1)}\theta^{(1)}}{\theta^{(2)}} < t^{(2)}\\ 0 & \text{if } -\frac{s+x^{(1)}\theta^{(1)}}{\theta^{(2)}} \geq t^{(2)}.\end{cases}$$

Therefore, if $r_1(x^{(1)}, \cdot)$ and $\mathrm{pdf}_{X^{(2)}|X^{(1)}=x^{(1)}}(\cdot)$ are continuous for $\mathbb{P}_{X^{(1)}}$-a.a. $x^{(1)}$, then $\theta \mapsto h_s(\theta, x^{(1)})$ is differentiable in a neighborhood of θ_0 except when $-\frac{s+x^{(1)}\theta^{(1)}}{\theta^{(2)}} = t^{(2)}$. However, $\mathbb{P}_{X^{(1)}}(-\frac{s+X^{(1)}\theta^{(1)}}{\theta^{(2)}} = t^{(2)}) = 0$ because $X^{(1)}$ is continuously distributed. Therefore, following the argument justifying the linear expansion for $T_{2,a}^{\theta}$, we conclude that $\theta \mapsto T_1^{\theta}(f)$ is differentiable in a neighborhood of θ_0, with

$$\frac{\partial}{\partial\theta^{(1)}} T_1^{\theta}(f) = \int \frac{x^{(1)}}{\theta^{(2)}} \mathbb{1}_{(-\infty,t^{(1)}]}(x^{(1)}) \mathbb{1}_{(-\infty,t^{(2)}]}(-\frac{s + x^{(1)}\theta^{(1)}}{\theta^{(2)}})$$
$$r_1(x^{(1)}, -\frac{s + x^{(1)}\theta^{(1)}}{\theta^{(2)}}) \operatorname{pdf}_{X^{(2)}|X^{(1)}=x^{(1)}}(-\frac{s + x^{(1)}\theta^{(1)}}{\theta^{(2)}}) \operatorname{pdf}_{X^{(1)}}(x^{(1)})\, dx^{(1)}$$
$$\frac{\partial}{\partial\theta^{(2)}} T_1^{\theta}(f) = \int -\frac{s + x^{(1)}\theta^{(1)}}{(\theta^{(2)})^2} \mathbb{1}_{(-\infty,t^{(1)}]}(x^{(1)}) \mathbb{1}_{(-\infty,t^{(2)}]}(-\frac{s + x^{(1)}\theta^{(1)}}{\theta^{(2)}})$$
$$r_1(x^{(1)}, -\frac{s + x^{(1)}\theta^{(1)}}{\theta^{(2)}}) \operatorname{pdf}_{X^{(2)}|X^{(1)}=x^{(1)}}(-\frac{s + x^{(1)}\theta^{(1)}}{\theta^{(2)}}) \operatorname{pdf}_{X^{(1)}}(x^{(1)})\, dx^{(1)}.$$

Next, we show that under certain conditions the second derivative of $T_1^{\theta}(f)$ exists and is uniformly bounded in f. So let $g(x^{(1)}, x^{(2)}) := r_1(x^{(1)}, x^{(2)}) \operatorname{pdf}_{X^{(2)}|X^{(1)}=x^{(1)}}(x^{(2)})$. Then, since

$$\mathbb{1}_{(-\infty,t^{(1)}]}(x^{(1)}) \mathbb{1}_{(-\infty,t^{(2)}]}(-\frac{s + x^{(1)}\theta^{(1)}}{\theta^{(2)}}) = \mathbb{1}_{(-\infty,t^{(1)}]}(x^{(1)}) \mathbb{1}_{[-\frac{s+t^{(2)}\theta^{(2)}}{\theta^{(1)}},\infty)}(x^{(1)}) = \mathbb{1}_{[-\frac{s+t^{(2)}\theta^{(2)}}{\theta^{(1)}},t^{(1)}]}(x^{(1)}),$$

we can write

$$\frac{\partial}{\partial\theta^{(2)}} T_1^{\theta}(f) = \int -\frac{s + x^{(1)}\theta^{(1)}}{(\theta^{(2)})^2} \mathbb{1}_{[-\frac{s+t^{(2)}\theta^{(2)}}{\theta^{(1)}},t^{(1)}]}(x^{(1)}) g(x^{(1)}, -\frac{s + x^{(1)}\theta^{(1)}}{\theta^{(2)}}) \operatorname{pdf}_{X^{(1)}}(x^{(1)})\, dx^{(1)}$$
$$= \begin{cases} \int_{-\frac{s+t^{(2)}\theta^{(2)}}{\theta^{(1)}}}^{t^{(1)}} -\frac{s+x^{(1)}\theta^{(1)}}{(\theta^{(2)})^2} g(x^{(1)}, -\frac{s+x^{(1)}\theta^{(1)}}{\theta^{(2)}}) \operatorname{pdf}_{X^{(1)}}(x^{(1)})\, dx^{(1)} & \text{if } -\frac{s+t^{(2)}\theta^{(2)}}{\theta^{(1)}} < t^{(1)} \\ 0 & \text{if } -\frac{s+t^{(2)}\theta^{(2)}}{\theta^{(1)}} \geq t^{(1)}. \end{cases}$$

If $s = t^{(1)} = t^{(2)} = 0$, then $\frac{\partial}{\partial\theta^{(2)}} T_1^{\theta} = 0$ for all θ and, hence, its second derivative with respect to θ is also zero. Therefore, assume that at least one of $s, t^{(1)}, t^{(2)}$ is not zero. Then, assuming that the operations of differentiation and integration can be exchanged and that $(\theta^{(1)}, \theta^{(2)}) \mapsto g(x^{(1)}, -\frac{s+x^{(1)}\theta^{(1)}}{\theta^{(2)}})$ is differentiable for $\mathbb{P}_{X^{(1)}}$-a.a. $x^{(1)}$, the second derivative $\frac{\partial^2}{\partial(\theta^{(2)})^2} T_1^{\theta}(f)$ does not exist when $s + \theta^{(1)}t^{(1)} + \theta^{(2)}t^{(2)} = 0$. To rule out such cases, assume there exists a deleted neighborhood of θ_0, say D_{θ_0}, such that the elements of the set $\{1, \theta^{(1)}, \theta^{(2)} : (\theta^{(1)}, \theta^{(2)}) \in D_{\theta_0}\}$ are linearly independent. This means that given any $s, t^{(1)}, t^{(2)}$ not all zero, there does not exist $\theta \in D_{\theta_0}$ such that $s + \theta^{(1)}t^{(1)} + \theta^{(2)}t^{(2)} = 0$. Hence, if $\theta \in D_{\theta_0}$,

$$\frac{\partial^2}{\partial(\theta^{(2)})^2} T_1^{\theta}(f) = \mathbb{1}(-\frac{s + t^{(2)}\theta^{(2)}}{\theta^{(1)}} < t^{(1)})$$
$$\times \frac{\partial}{\partial\theta^{(2)}} \int_{-\frac{s+t^{(2)}\theta^{(2)}}{\theta^{(1)}}}^{t^{(1)}} -\frac{s + x^{(1)}\theta^{(1)}}{(\theta^{(2)})^2} g(x^{(1)}, -\frac{s + x^{(1)}\theta^{(1)}}{\theta^{(2)}}) \operatorname{pdf}_{X^{(1)}}(x^{(1)})\, dx^{(1)}.$$

Next, assuming that

$$\sup_{s\in\mathbb{R},\theta\in D_{\theta_0}} \left|\frac{\partial}{\partial\theta^{(2)}} \int_{-\frac{s+t^{(2)}\theta^{(2)}}{\theta^{(1)}}}^{t^{(1)}} -\frac{s + x^{(1)}\theta^{(1)}}{(\theta^{(2)})^2} g(x^{(1)}, -\frac{s + x^{(1)}\theta^{(1)}}{\theta^{(2)}}) \operatorname{pdf}_{X^{(1)}}(x^{(1)})\, dx^{(1)}\right| < \infty,$$

it follows that $\sup_{f\in\mathcal{F}_1,\theta\in D_{\theta_0}} |\frac{\partial^2}{\partial(\theta^{(2)})^2} T_1^\theta(f)| < \infty$. A similar argument can be used to show that the second derivative with respect to $\theta^{(1)}$, and the two cross-partial derivatives, are also uniformly bounded in $f \in \mathcal{F}_1$ and $\theta \in D_{\theta_0}$. Therefore, the uniform linear expansion for T_1^θ can be justified under certain regularity conditions.

TABLE 2. Censored regression: Rejection rates for $\hat{\mathrm{T}}_{\max}$ when H_0 is true.

Design	Description	Sample size	Nominal size 1% (.0044)	Nominal size 5% (.0097)	Nominal size 10% (.0134)
(a)	medium censoring, constant variance	50	.012	.054	.114
		100	.012	.052	.104
		200	.018	.064	.104
(b)	high censoring, constant variance	50	.02	.08	.136
		100	.006	.06	.108
		200	.018	.06	.116
(c)	medium censoring, increasing variance	50	.02	.064	.108
		100	.014	.052	.114
		200	.016	.072	.118
		400	.014	.054	.094
		600	.01	.058	.116
(d)	high censoring, increasing variance	50	.018	.086	.144
		100	.014	.07	.126
		200	.03	.076	.124
		400	.022	.062	.116
		600	.004	.04	.084
		800	.006	.062	.106

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

Table 3. Censored regression: Rejection rates for $\hat{\mathrm{T}}_{\max}$ when $\tilde{H}_0$ is false.

Design	Description	Sample size	Nominal size 1%	5%	10%
		50	.534	.766	.848
1	$\theta_0 = 1, \theta_1 = 1$, censoring = 3%	100	.942	.988	.996
	$\mu = 0, \sigma^2 = 1$, skewness = 0.56	200	1	1	1
		50	.524	.738	.842
2	$\theta_0 = 2, \theta_1 = 4.5$, censoring = 27%	100	.924	.974	.996
	$\mu = 0, \sigma^2 = 1$, skewness = −0.26	200	1	1	1
		50	.17	.404	.55
3	$\theta_0 = -0.5, \theta_1 = 1.25$, censoring = 48%	100	.578	.772	.862
	$\mu = -1, \sigma^2 = 1$, skewness = −0.25	200	.938	.988	.994
		50	.314	.578	.698
4	$\theta_0 = -0.5, \theta_1 = 3$, censoring = 49%	100	.762	.92	.954
	$\mu = -1, \sigma^2 = 1$, skewness = −0.53	200	.996	1	1
		50	.436	.662	.794
5	$\theta_0 = 0, \theta_1 = 1$, censoring = 50%	100	.878	.964	.976
	$\mu = -5, \sigma^2 = 1$, skewness = 0.94	200	1	1	1

The last three columns report the fraction of simulations for which $\hat{\mathrm{T}}_{\max} > c^*_{\alpha,B}$.

Table 4. Censored regression: Empirical power of $\hat{T}_{\text{Newey}}$ vs. $\hat{\mathrm{T}}_{\max}$.

Statistic	Sample size	Nominal size 1%	5%	10%
$\hat{T}_{\text{Newey}}$	50	.23	.308	.348
	100	.16	.208	.236
	200	.06	.106	.146
	500	.016	.052	.096
	1000	.016	.058	.106
$\hat{\mathrm{T}}_{\max}$	50	.3	.474	.54
	100	.316	.496	.592
	200	.328	.502	.606
	500	.386	.57	.682
	1000	.446	.666	.774

The last three columns report the fraction of simulations for which $\hat{T}_{\text{Newey}} > \mathrm{cdf}^{-1}_{\chi^2_2}(1-\alpha)$ and $\hat{\mathrm{T}}_{\max} > c^*_{\alpha,B}$.

Table 5. Truncated regression: Rejection rates for $\hat{\mathrm{T}}_{\max}$ when H_0 is true.

Design	Sample size	Nominal size 1% (.0044)	5% (.0097)	10% (.0134)
medium truncation, constant variance	50	.012	.054	.102
	100	.01	.04	.084
	200	.006	.032	.068
	300	.008	.052	.096

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

Table 6. Truncated regression: Rejection rates for $\hat{\mathrm{T}}_{\max}$ under the alternative.

Design		Sample size	Nominal size 1%	5%	10%
		50	.53	.772	.83
$\theta_0 = 2, \theta_1 = 4.5$	truncation = 27%	100	.942	.986	.998
$\mu = 0, \sigma^2 = 1$	skewness = -0.26	200	1	1	1

The last three columns report the fraction of simulations for which $\hat{\mathrm{T}}_{\max} > c^*_{\alpha,B}$.

Figure 1. QQ plots for the dataset in Lee's book: In plot (a), quantiles of $-\hat{V}$ are on the vertical axis and quantiles of $\hat{V}$ are on the horizontal axis. In plot (b), quantiles of standardized $\hat{V}$ are on the vertical axis and quantiles of $\mathrm{N}(0,1)$ are on the horizontal axis.

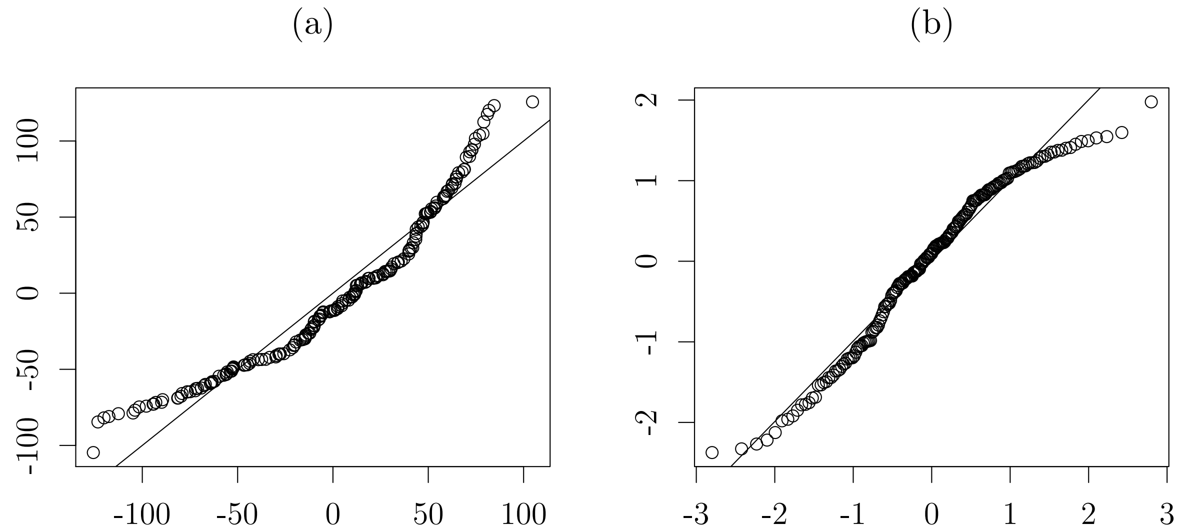
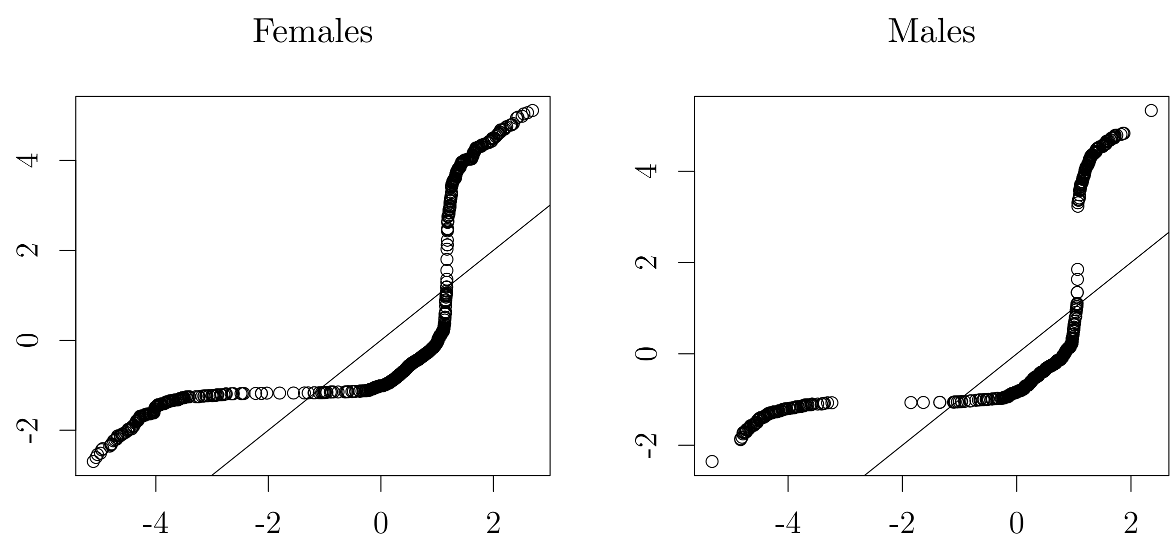


Figure 2. QQ plots for Jolliffe's dataset: In both plots, quantiles of $-\hat{V}$ are on the vertical axis and quantiles of $\hat{V}$ are on the horizontal axis.



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.

Amemiya, T. (1985): *Advanced econometrics*. Harvard University Press, Cambridge, Massachusetts.

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

Chay, K., and J. Powell (2001): "Semiparametric censored regression models," *Journal of Economic Perspectives*, 15, 29–42.

Chen, T., and G. Tripathi (2013): "Testing conditional symmetry without smoothing," *Journal of Nonparametric Statistics*, 25, 273–313.

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.

Heckman, J. J. (1976): "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," *Annals of Economic and Social Measurement*, 5, 475–492.

Honoré, B., E. Kyriazidou, and C. Udry (1997): "Estimation of type 3 tobit models using symmetric trimming and pairwise comparisons," *Journal of Econometrics*, 76, 107–128.

Hyndman, R. J., and Q. Yao (2002): "Nonparametric estimation and symmetry tests for conditional density functions," *Journal of Nonparametric Statistics*, 14, 259–278.

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.

Jolliffe, D. (2002): "The gender wage gap in Bulgaria: A semiparametric estimation of discrimination," *Journal of Comparative Economics*, 30, 276–295.

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

Lee, M.-J. (1995): "Semi-parametric estimation of simultaneous equations with limited dependent variables: A case study of female labour supply," *Journal of Applied Econometrics*, 10, 187–200.

——— (1996): *Methods of moments and semiparametric econometrics for limited dependent variable models*. Springer-Verlag, New York.

Levitt, S. (1996): "How do senators vote? Disentangling the role of voter preferences, party affiliation, and senator ideology," *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," *Journal of Nonparametric Statistics*, 17, 697–715.

Newey, W. K. (1987): "Specification tests for distributional assumptions in the tobit model," *Journal of Econometrics*, 34, 125–145.

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.

Santos Silva, J. M. C. (2001): "Influence diagnostics and estimation algorithms for Powell's SCLS," *Journal of Business and Economic Statistics*, 19, 55–62.

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

Department of Economics, University of Waterloo, Ontario, Canada.
E-mail address: t66chen@uwaterloo.ca

Faculty of Law, Economics and Finance, University of Luxembourg, Luxembourg.
E-mail address: gautam.tripathi@uni.lu