Economics Working Paper Series

2015 - 15

# An Improved Bootstrap Test for Restricted Stochastic Dominance

Thomas M. Lok and Rami V. Tabri

June 2015

# An Improved Bootstrap Test for Restricted Stochastic Dominance

Thomas M. Lok and Rami V. Tabri

School of Economics, The University of Sydney,

Sydney, New South Wales 2006, Australia

June 25, 2015

**Abstract**

This paper proposes a uniformly asymptotically valid method of testing for restricted stochastic dominance based on the bootstrap test of Linton et al. (2010). The method reformulates their bootstrap test statistics using a constrained estimator of the contact set that imposes the restrictions of the null hypothesis. As our simulation results show, this characteristic of our test makes it noticeably less conservative than the test of Linton et al. (2010) and improves its power against alternatives that have some non-violated inequalities.

JEL Classification: C12 (Hypothesis Testing); C14 (Semiparametric and Nonparametric Methods); I32 (Measurement and Analysis of Poverty)

Keywords: Empirical Likelihood; Constrained Estimation; Restricted Stochastic Dominance; Bootstrap Test.

# 1 Introduction

Stochastic dominance orderings of income distributions are fundamental in poverty and income studies. They can be used to determine whether poverty or social welfare is greater in one income distribution than in another for general classes of poverty indices and for ranges of possible poverty lines (e.g. Atkinson, 1987 and Foster and Shorrocks, 1988). These orderings can either be unrestricted or restricted, as to whether the comparison of the income distributions is carried out over the entire range of incomes or only over some *restricted* ranges of incomes. From a normative perspective, the unrestricted stochastic dominance orderings are deficient because they do not give equal ethical weight to all those who are below a survival poverty line. Whereas the rankings based on the restricted stochastic dominance orderings do not suffer from this deficiency[1].

In practice, population distributions are not in general observable, and so comparisons must be based on statistical tests that make use of distributions estimated from samples. Many tests that posit a null of unrestricted stochastic dominance of a given order appeared over the last two decades (e.g. McFadden, 1989, Barrett and Donald, 2003, Linton et al., 2005, Horváth et al., 2006, and Linton et al., 2010). All of them are applicable to testing for restricted stochastic dominance orderings, which is the empirically sensible course to follow. The reason being that there can be too little sample information from the tails of the distributions to be able to distinguish dominance curves statistically over the full range of incomes.

Linton et al. (2010) (LSW) propose a bootstrap method of testing for this ordering based on the estimation of the "contact set". The contact set is the set of incomes on which the dominance curves of the two distributions coincide. This paper proposes a uniformly asymptotically valid modification of the LSW test that uses a constrained estimator of the contact set. Specifically, the modification is to replace the contact set estimator in the LSW test procedure with the one based on the constrained empirical likelihood estimator of the restricted stochastic dominance curves. This approach reformulates their bootstrap test statistics using a contact set estimator that incorporates the statistical information from imposing the constraints of the null hypothesis.

[1] See Bourguignon and Fields (1997) for more on this point.

In contrast, the LSW contact set estimator ignores this statistical information because it's based on the sample analogue estimator of the restricted stochastic dominance curves. We report Monte Carlo simulation results that compare the modified LSW test and its unmodified counterpart. These results show the modified test has better Type I error properties, and substantially higher over all power.

Tests for restricted stochastic dominance are not new. Davidson and Duclos (2013) and Davidson (2009) propose asymptotic and bootstrap tests that posit instead a null of non-dominance. By contrast, our paper and the literature discussed earlier, have non-dominance as one of the configurations under the alternative. Therefore, these two approaches are not directly comparable, but they certainly do complement each other.

The rest of this paper is organized as follows. Section 2 presents the test problem, the model of the null hypothesis, and the constrained empirical likelihood estimator of Tabri (2015). Section 3 presents the main result of the paper, namely the uniform asymptotic validity of the modified LSW bootstrap test. Section 4 discusses the usefulness of the main result and Section5 reports the findings of Monte Carlo simulation experiments. Finally, Section 6 concludes and Section 7 collates the acknowledgements of the individuals and institutions who provided help during the research.

# 2 Background

Consider two populations, $A$ and $B$, with respective income distributions $F_A$ and $F_B$, and suppose that there is a joint CDF, $F$, whose marginal CDFs are $F_A$ and $F_B$. Accounting for statistical dependence between the incomes in the two populations is essential in many applications, such as the comparison of income distributions over time, or before and after an economic policy. Distribution $B$ is said to dominate distribution $A$, stochastically at order $s \in \mathbb{Z}_+$ and over the range

$[\underline{t}, \overline{t}] \subset \text{supp}\,(F_A) \cup \text{supp}\,(F_B)$, if

$$E_F\left[\frac{\left(t - X^B\right)^{s-1}}{(s-1)!} 1\left[X^B \leq t\right] - \frac{\left(t - X^A\right)^{s-1}}{(s-1)!} 1\left[X^A \leq t\right]\right] \leq 0 \quad \forall t \in [\underline{t}, \overline{t}], \tag{1}$$

where $\mathbf{X} = [X^A, X^B]$ is a random vector with CDF $F$, and supp $(F_K)$ is the support of $F_K,\ K = A, B$.

Let $P_0$ denote the "true" distribution of $\mathbf{X}$. Given $s \in \mathbb{Z}_+$ and $[\underline{t}, \overline{t}]$, we wish to test that $P_0$ satisfies the moment inequalities (1), where $P_0$ belongs to a large class of distributions $\mathcal{M}$, which we define below. The restrictions that define $\mathcal{M}$ ensures that the proposed modification of the LSW test is asymptotically valid, with uniformity. Let $\Delta(P_0)$ denote the contact set $\left\{t \in [\underline{t}, \overline{t}] : E_{P_0}[g(\mathbf{X}; t)] = 0\right\}$, where $g(\mathbf{X}; t)$ is the moment function in (1). The asymptotic behavior of the LSW test statistic depends on the form of $\Delta(P_0)$. Furthermore, the behavior of the proposed modification of this test depends on the covariances of the random variables $\{g(\mathbf{X}; t),\ t \in \Delta(P_0)\}$.

Let $P$ denote a generic value of the distribution of $\mathbf{X}$, and let $\mathcal{M}$ be some collection of $P$ that satisfies the following parameter space Assumption 2.1 for a given constant $c > 0$.

**Assumption 2.1.** *(i) Dependence: neither of the random variables $X^A$ and $X^B$ is a deterministic transformation of the other; (ii) Sampling: $\{\mathbf{X}_i\}_{i=1}^n$ is a random sample from $P$; and (iii) For every finite subset of $\Delta(P)$, denoted by $T$, the covariance matrix formed by the random variables $\{g(\mathbf{X}; t),\ t \in T\}$, denoted by $\Sigma_T(P)$, satisfies $\theta' \Sigma_T(P) \theta \geq c \quad \forall \theta \in \mathbb{R}^{|T|}$ such that $\|\theta\|_{\mathbb{R}^{|T|}} = 1$.*

The parameter space $\mathcal{M}_-$ for the model of the null hypothesis is the subset of $\mathcal{M}$ that satisfies (1). Part (i) of Assumption 2.1 allows for a rich dependence structure between the marginal random variables, which covers applications such as the ranking of pre- and post-policy income distributions. Part (iii) of Assumption 2.1 excludes distributions that become arbitrarily close to some distribution that puts probability 1 on a strict subspace of the sample space of income pairs.

**Remark.** The parameter space $\mathcal{M}$ is similar to the one in Tabri (2015), but differs from it in two important ways. Firstly, Tabri (2015)'s parameter space requires the continuity of the moment functions, which applies broadly to many robust orderings of poverty; however, this condition excludes the robust ranking of first-order stochastic dominance conditions from his applications because when $s = 1$ in (1) the moment functions are differences of indicator functions. For $s > 1,$ the moment functions are indeed continuous. Secondly, Tabri (2015)'s parameter space requires the invertibility of certain covariance matrices to develop inference based on the empirical likelihood-ratio statistic, which is not required in this paper's setup because the employed test statistic's distribution theory does not rely on these conditions.

Let $\delta_{\mathbf{X}_i}$ be the point-mass delta function at $\mathbf{X}_i,$ and let $\left\{\mathcal{T}_{N(n)} : n \geq 1\right\}$ be a given sequence of subsets of $[\underline{t}, \overline{t}]$ with $|\mathcal{T}_{N(n)}| = N(n)\ \forall n$ that converges to $[\underline{t}, \overline{t}]$ in the Hausdorff metric as $n \to +\infty$. LSW propose an estimator of $\Delta\left(P_0\right)$ based on the sample analogue estimator of the moments $E_{P_0}\left[g\left(\mathbf{X};\cdot\right)\right]$. Specifically, they estimate $\Delta\left(P_0\right)$ using

$$\hat{\Delta}_n = \left\{t \in [\underline{t}, \overline{t}] : \left|E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right]\right| \leq r_n\right\}, \quad \text{where} \quad \hat{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\mathbf{X}_i} \tag{2}$$

is the empirical distribution function (ECDF) of the random sample, and $\{r_n\}_{n\geq 1}$ is a suitably chosen null sequence of positive (possibly random) numbers that satisfies $\sqrt{n}r_n \to +\infty$ as $n \to +\infty$. The proposed contact set estimator replaces $E_{\hat{P}_n}\left[g\left(\mathbf{X};\cdot\right)\right]$ with $E_{\acute{P}_n}\left[g\left(\mathbf{X};\cdot\right)\right]$ in the definition of $\hat{\Delta}_n,$ where $\acute{P}_n = \sum_{i=1}^{n} \acute{p}_i \delta_{\mathbf{X}_i}$ with the probabilities $\acute{p}_1, \ldots, \acute{p}_n$ defined as the solution of the following optimization problem: $\max\limits_{p_1,\ldots,p_n} \sum\limits_{i=1}^{n} \log p_i$ subject to $p_i \geq 0\ i = 1, \ldots, n,$ $\sum_{i=1}^{n} p_i = 1,$ and

$$\sum_{i=1}^{n} p_i g\left(\mathbf{X}_i; t\right) \leq 0 \quad \forall t \in \mathcal{T}_{N(n)}. \tag{3}$$

The estimator $\acute{P}_n$ is the approximate constrained empirical likelihood estimator of $P_0,$ and we denote the contact set estimator based on it by $\acute{\Delta}_n$. The estimator $\hat{P}_n$ solves the above optimization problem, but without imposing the constraints (3); therefore, $E_{\hat{P}_n}\left[g\left(\mathbf{X};\cdot\right)\right]$ does not necessarily satisfy the restrictions of the null hypothesis. By contrast, from (3), the definition of $\acute{P}_n$ implies

$E_{\acute{P}_n}\left[g\left(\mathbf{X};\cdot\right)\right]$ approximately satisfies the constraints (1) but with the approximation disappearing asymptotically. Therefore, the estimator $\acute{\Delta}_n$ incorporates the statistical information from imposing the restrictions of the null hypothesis, whereas $\hat{\Delta}_n$ does not have this property. In consequence, we expect the modification of the LSW test this paper proposes to have better finite-sample properties than the LSW test.

# 3 Main Results

This section introduces the main results of the paper. In the setting of this paper, the Cramér von Mises type test statistic LSW use is given by $\hat{T}_n = n\int_{\underline{t}}^{\bar{t}}\left(\max\left\{E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right],0\right\}\right)^2 dt$. The LSW bootstrap test procedure follows these steps:

1. Using the data, compute $\hat{T}_n$ and $\hat{P}_n$.

2. Generate $B$ bootstrap samples each of size $n$, $\left\{\mathbf{X}^\star_{i,l}\right\}_{i=1}^n$ for $l=1,\ldots,B$, using resampling with replacement from $\hat{P}_n$. That is, draw $\mathbf{X}^\star_{i,l}$ randomly with replacement from $\{\mathbf{X}_i\}_{i=1}^n$ according to $\hat{P}_n$ for $i=1,\ldots,n$ and $l=1,\ldots,B$.

3. For each bootstrap sample, compute the bootstrap test statistic as follows:

$$\hat{T}^\star_{n,l} = \begin{cases} \int_{\underline{t}}^{\bar{t}}\left(\max\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[g\left(\mathbf{X}^\star_{i,l};t\right)-E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]\right],0\right\}\right)^2 dt, & \text{if } \hat{\Delta}_n=\emptyset, \\ \int_{\hat{\Delta}_n}\left(\max\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[g\left(\mathbf{X}^\star_{i,l};t\right)-E_{\hat{P}_n}\left[g\left(\mathbf{X};t\right)\right]\right],0\right\}\right)^2 dt, & \text{if } \hat{\Delta}_n\neq\emptyset, \end{cases}$$

   where $\hat{\Delta}_n$ is defined in (2).

4. Compute the approximate bootstrap p-value $\hat{\Upsilon}_B = \frac{1}{B}\sum_{l=1}^B 1\left[\hat{T}^\star_{n,l}\geq\hat{T}_n\right]$.

5. Reject $H_0$ if $\hat{\Upsilon}_B\leq\beta$, where $\beta\in(0,1/2)$ is a given nominal level.

The test procedure this paper proposes follows the steps of the LSW bootstrap test procedure, but with $\hat{\Delta}_n$ replaced by $\acute{\Delta}_n$ when computing the bootstrap test statistics in the third step above.

Let $\left\{\acute{T}^{\star}_{n,l}\right\}_{l=1}^{B}$ denote the bootstrap test statistics computed as above but with $\hat{\Delta}_n$ replaced by $\acute{\Delta}_n$, and let $\mathcal{A}_n$ denote the sigma-algebra generated by the random sample $\{\mathbf{X}_i\}_{i=1}^{n}$. The following result shows the bootstrap test statistics from the two procedures are asymptotically equivalent, uniformly in the model of the null hypothesis.

**Theorem 1.** *Suppose that $P_0 \in \mathcal{M}_-$. Then $\hat{T}^{\star}_{n,l} - \acute{T}^{\star}_{n,l} \xrightarrow{P} 0$ conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}_-$.*

*Proof.* See Appendix A. □

The next result is an immediate consequence of Theorem 1. It states the approximate bootstrap p-values from the two procedures are also uniformly asymptotically equivalent over $\mathcal{M}_-$.

**Corollary 1.** *Let $\acute{\Upsilon}_B = \frac{1}{B}\sum_{l=1}^{B} 1\left[\acute{T}^{\star}_{n,l} \geq \hat{T}_n\right]$. Then $\hat{\Upsilon}_B - \acute{\Upsilon}_B \xrightarrow{P} 0$ conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}_-$.*

*Proof.* See Appendix A. □

Since the LSW test is valid in the setting of the paper, Corollary 1 establishes the uniform asymptotic validity of the proposed modification of the LSW test.

# 4 Discussion

This section discusses the implications of Theorem 1 and Corollary 1 for testing the continuum of moment inequality restrictions (1), under the null. As already mentioned, these results show that the modification of the LSW test this paper proposes is asymptotically valid in a uniform sense. The important difference between the proposed test and the LSW one is that the former uses a restricted estimator of the contact set, whereas the latter does not. In finite-samples, this restricted estimator approximately imposes the restrictions of the null hypothesis (1) by imposing the restrictions in (3), with the approximation disappearing asymptotically. Accordingly, the proposed modification of the LSW test alters the bootstrap test statistics in a data-dependent way that incorporates the statistical information from imposing the restrictions of the null hypothesis.

The motivation and intuition behind using a restricted estimator in test procedures, in general, are well understood. Such procedures usually have better characteristics in comparison to tests that do not account for the information from imposing the restrictions of the null hypothesis in estimation. Under the null, the use of the restricted contact set estimator gives rise to a bootstrap distribution of the test statistic that is a more reliable estimator of the test statistic's sampling distribution. Under the alternative, constrained estimation of the contact set biases the bootstrap distribution of the test statistic in the direction of the null. In consequence, the test statistic computed from data would be more extreme on the basis of the approximate bootstrap p-value, in comparison to the setup that uses the unrestricted estimator of the contact set.

# 5 Monte Carlo Experiments

This section reports the results of Monte Carlo experiments that compares the performance of the LSW test with the one this paper proposes. The experimental setup is the same as the one in Section 5 of LSW. We find the modified test has noticeably reduced non-similarity on the boundary of the null hypothesis, and higher power against alternatives that have some non-violated inequalities (SNVI). Such alternatives have stochastic dominance conditions with some positive elements and some elements that are negative.

In each simulation experiment, the nominal level was fixed at 5%, $N(n) = \lfloor \sqrt{n} \rfloor + 1$ and $r_n(t) = \hat{\sigma}_t \sqrt{\frac{\log n}{n}}$ where $\hat{\sigma}_t^2$ is the sample analogue estimator of $E_{P_0}\left[g\left(\mathbf{X};t\right)\right]^2 - \left(E_{P_0}\left[g\left(\mathbf{X};t\right)\right]\right)^2$ where $t \in [\underline{t}, \overline{t}]$. Additionally, we set $\underline{t} = 0.05$ and $\overline{t} = 0.95$, and construct the grid as follows:

$$\mathcal{T}_{N(n)} = \left\{\underline{t} = t_1 < t_2 < \cdots < t_{N(n)} = \overline{t}\right\}, \text{ where } t_{i+1} = t_i + \frac{\left(\overline{t} - \underline{t}\right)}{\lfloor \sqrt{n} \rfloor}, \tag{4}$$

for $i = 1, \ldots, N(n) - 1$. The number of Monte Carlo replications was set to be 1000, and the number of bootstrap replications was 199.

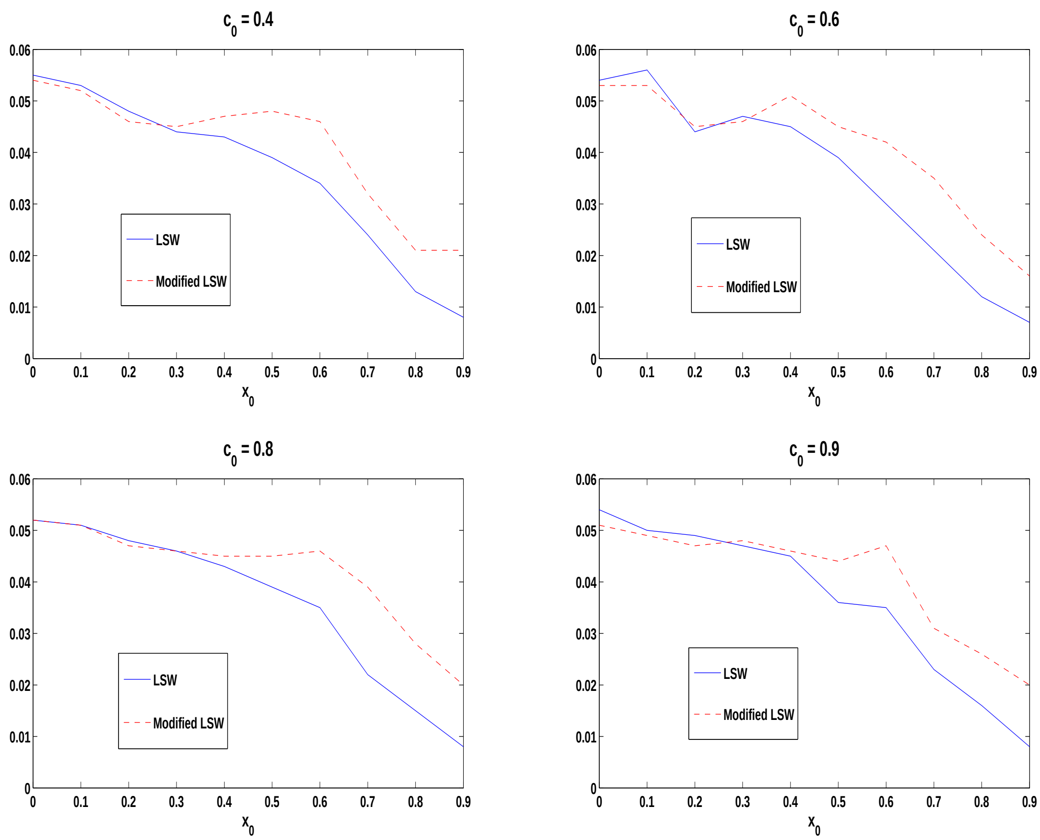


Figure 1: The empirical rejection probabilities under the null.

First we compare the type I error rate properties of the our test and LSW test. LSW use the following generating process under the null. Let $U_1$ and $U_2$ be $U(0,1)$ random variables. Then define $X^B = U_1$ and $X^A = c_0^{-1}(U_2 - a_0)1\left[0 < U_2 \leq x_0\right] + U_2 1\left[x_0 < U_2 < 1\right]$, where $c_0 = (x_0 - a_0)/x_0 \in (0,1)$ and $x_0 \in (0,1)$. In this setup, the inequalities (1) hold for each $s \in \mathbb{Z}_+$, and we examine the case $s = 1$. In the simulations, we took $x_0 \in \{0, 0.1, 0.2, \ldots, 0.9\}$ and $c_0 \in \{0.2, 0.4, 0.6, 0.8\}$. The sample size was fixed at 500. The case $x_0 = 0$ corresponds to the least favorable case. As $x_0$ gets larger, for a given $c_0 > 0$, the contact set gets smaller; therefore, the data-generating process (DGP) moves away from the least favorable case into the interior of the null.

The results are reported in Figure 1. For each value of $c_0$ we considered, the discrepancy between the performances of our method and the LSW test is not much for $x_0$ close to the least

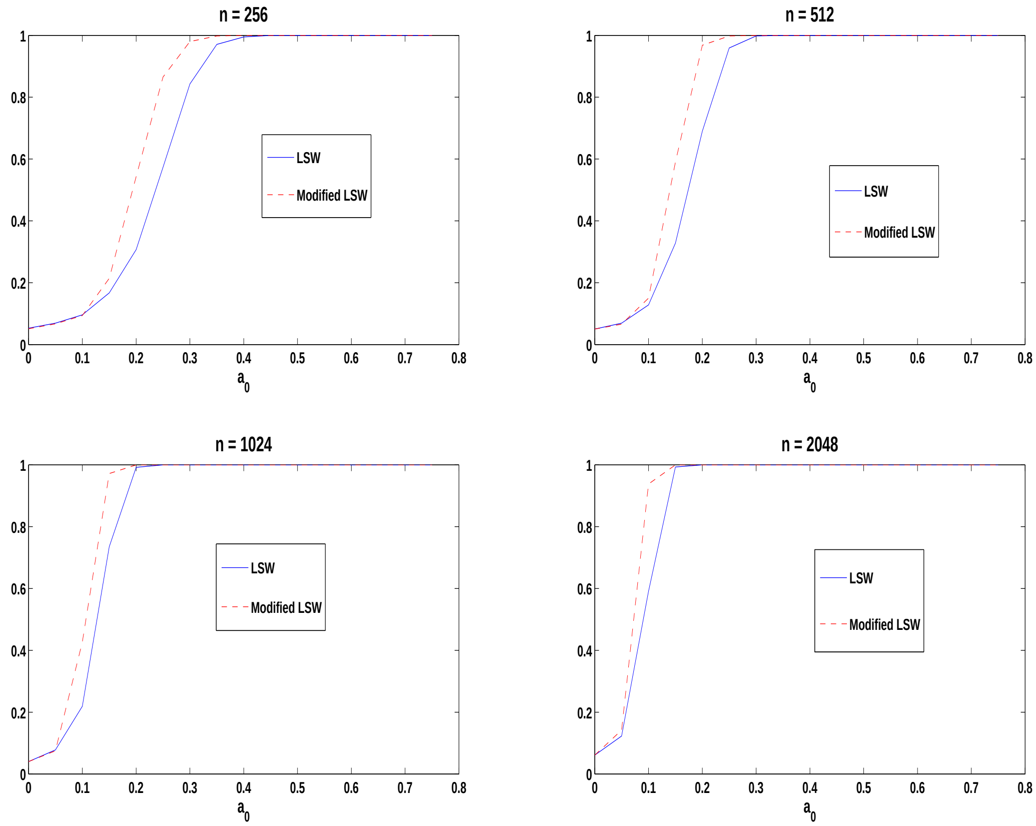


Figure 2: The empirical rejection probabilities under the alternative.

favorable case. However, as $x_0$ gets larger, our test shows rejection probabilities that are closer to the 5% nominal level than the ones based on the LSW test. These results suggest the bias of the LSW test is larger than the one this paper proposes.

Let us now focus on the power properties of the two methods. Consider the following configuration of DGPs from LSW. Set $X^A \sim U[0,1]$. Then define

$$X^B = (U - a_0 b_1)\, 1\left[a_0 b_1 \leq U \leq x_0\right] + (U + a_0 b_2)\, 1\left[x_0 < U \leq 1 - a_0 b_2\right] \tag{5}$$

for $a_0 \in (0,1)$, where $U \sim U[0,1]$. As $a_0$ becomes closer to zero, the distribution of $X^B$ becomes closer to the uniform distribution. The scale $a_0$ plays the role of the "distance" $P_0$ is from $H_0$. When $a_0$ is large, $P_0$ is farther from $H_0$, and when $a_0 = 0$, $X^A$ and $X^B$ have the same distribution which means $P_0$ belongs to the model of the null hypothesis under the least favorable configuration. We set $(b_1, b_2, x_0) = (0.1, 0.5, 0.15)$ and $a_0 \in \{0, 0.05, 0.1, 0.15, 0.2, \ldots, 0.75\}$. The configurations for which $a_0 \neq 0$ correspond to alternative DGPs for which there are some non-violated inequalities for the case for $s = 1$ in the moments (1). We considered the following sample sizes $n = 256, 512, 1024, 2048,$ and set $X^A$ and the uniform random variable in the definition of $X^B$ to be negatively correlated, with a correlation coefficient of -0.5.

The simulation results are reported in Figures 2. For each sample size and for $a_0$ sufficiently large, there is no difference between the two tests, which is expected since both tests are consistent. For $n = 1024, 2048,$ our test dominates the LSW test, and quite significantly when $a_0 = 0.1$ and $n = 2048$. This substantial improvement also holds when $a_0 = 0.1, 0.15$ and $n = 1024$. Both tests perform similarly when $a_0 = 0.05$. Overall, the simulation results show that our method performs better than the LSW test.

# 6 Conclusion

This paper proposes a new method of testing for restricted stochastic dominance. It is a modification of the Linton et al. (2010) test that incorporates the statistical information from imposing the restrictions of the null hypothesis in the estimation of the contact set. This modification alters the finite-sample distribution of the bootstrap test statistics in a data-dependent way. In comparison to the LSW test, the simulation study demonstrates that our test has noticeably reduced non-similarity on the boundary of the null and improved power against alternatives with some non-violated inequalities.

# 7 Acknowledgments

Rami V. Tabri thanks Drs. Peter Exterkate and Mervyn Silvapulle for helpful comments and discussions. Thomas M. Lok would like to thank Hayley Miles for her continued support and also Julian Lok for his helpful discussions and comments.

# References

A. Atkinson. On the Measurement of Poverty. *Econometrica*, 55:749–764, 1987.

G. F. Barrett and S. G. Donald. Consistent Tests for Stochastic Dominance. *Econometrica*, 71(1): 71–104, 2003.

François Bourguignon and Gary Fields. Discontinuous losses from poverty, generalized pa measures, and optimal transfers to the poor. *Journal of Public Economics*, 63(2):155 – 175, 1997. ISSN 0047-2727.

R. Davidson. Testing for Restricted Stochastic Dominance: Some Further Analysis. *Review of Economic Analysis*, 1:34–59, 2009.

R. Davidson and J-Y Duclos. Testing for Restricted Stochastic Dominance. *Econometric Reviews*, 32(1):84–125, 2013.

J. Foster and A. Shorrocks. Poverty Orderings. *Econometrica*, 56(1):173–177, 1988.

L. Horváth, P. Kokoszka, and R. Zitikis. Testing for Stochastic Dominance using the Weighted McFadden-type Statistic. *Journal of Econometrics*, 133:191–205, 2006.

O. Linton, E. Maasoumi, and Y-J. Whang. Consistent Testing for Stochastic Dominance under General Sampling Schemes. *Review of Economic Studies*, 72:735–765, 2005.

O. Linton, K. Song, and Y-J. Whang. An Improved Bootstrap Test for Stochastic Dominance. *Journal Of Econometrics*, 154:186–202, 2010.

D. McFadden. *Testing for Stochastic Dominance*. Studies in the Economics of Uncertainty in honor of Josef Hadar. Springer-Verlag, 1989.

Rami Victor Tabri. Empirical Likelihood for Robust Poverty Comparisons. Working Papers 2015-02, University of Sydney, School of Economics, May 2015.

This Appendix is not to be published. It will be made available on the web.

# Appendix

# to

# An Improved Bootstrap Test for Restricted Stochastic Dominance


*Thomas M. Lok and Rami V. Tabri*

*School of Economics, The University of Sydney,*

*Sydney, New South Wales, 2006, Australia. Email: rami.tabri@sydney.edu.au.*

# A Proofs of Main Results

Proof of Theorem 1:

*Proof.* The proof proceeds by the direct method. Let

$$\gamma_n^\star(t) = \left( \max\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ g\left(\mathbf{X}_{i,l}^\star; t\right) - E_{\hat{P}_n}\left[g\left(\mathbf{X}; t\right)\right]\right], 0 \right\} \right)^2, \tag{6}$$

then consider the following,

$$\left|\hat{T}_{n,l}^\star - \acute{T}_{n,l}^\star\right| = \begin{cases} \int_{[\underline{t},\overline{t}]-\acute{\Delta}_n} \gamma_n^\star(t)\ dt & \text{if } \acute{\Delta}_n \neq \emptyset, \hat{\Delta}_n = \emptyset \\ \int_{[\underline{t},\overline{t}]-\hat{\Delta}_n} \gamma_n^\star(t)\ dt & \text{if } \acute{\Delta}_n = \emptyset, \hat{\Delta}_n \neq \emptyset \\ \int_{\acute{\Delta}_n \ominus \hat{\Delta}_n} \gamma_n^\star(t)\ dt & \text{if } \acute{\Delta}_n \neq \emptyset, \hat{\Delta}_n \neq \emptyset \\ 0 & \text{if } \acute{\Delta}_n = \emptyset, \hat{\Delta}_n = \emptyset, \end{cases} \tag{7}$$

where $\ominus$ denotes the symmetric difference operator on sets. We have

$$\left|\hat{T}_{n,l}^\star - \acute{T}_{n,l}^\star\right| \leq \begin{cases} \left(\sup_{t\in[\underline{t},\overline{t}]} \gamma_n^\star(t)\right) \int_{[\underline{t},\overline{t}]-\acute{\Delta}_n} dt & \text{if } \acute{\Delta}_n \neq \emptyset, \hat{\Delta}_n = \emptyset \\ \left(\sup_{t\in[\underline{t},\overline{t}]} \gamma_n^\star(t)\right) \int_{[\underline{t},\overline{t}]-\hat{\Delta}_n} dt & \text{if } \acute{\Delta}_n = \emptyset, \hat{\Delta}_n \neq \emptyset \\ \left(\sup_{t\in[\underline{t},\overline{t}]} \gamma_n^\star(t)\right) \int_{\acute{\Delta}_n \ominus \hat{\Delta}_n} dt & \text{if } \acute{\Delta}_n \neq \emptyset, \hat{\Delta}_n \neq \emptyset, \\ 0 & \text{if } \acute{\Delta}_n = \emptyset, \hat{\Delta}_n = \emptyset. \end{cases} \tag{8}$$

To prove the result we need to prove that $\left(\sup_{t\in[\underline{t},\overline{t}]} \gamma_n^\star(t)\right)$ is $O_P(1)$ conditional on $\mathcal{A}_n$, uniformly in $\mathcal{M}_-$, and then use Lemma B.2 on the integrals in (8). Since the set of moment functions

$$\left\{\mathbf{x} \mapsto g(\mathbf{x}, t), t \in [\underline{t}, \overline{t}]\right\}$$

is uniform Donsker with respect to $\mathcal{M}_-$, Lemma A.2 of LSW implies that it is also bootstrap uniform Donsker. Therefore, applying Lemma A.1 (uniform continuous mapping theorem) of

LSW to $\left(\sup_{t\in[\underline{t},\overline{t}]}\gamma_n^{\star}(t)\right)$ yields the desired result.

Lemma B.3 shows that $\hat{\Delta}_n$ converges to $\Delta(P_0)$ in probability, uniformly in $\mathcal{M}_-$. Then, Lemma B.2 implies $\acute{\Delta}_n$ converges to $\Delta(P_0)$ in probability, uniformly in $\mathcal{M}_-$. Firstly, suppose that $\Delta(P_0)=\emptyset$. Then for large $n$, the bootstrap statistics $\hat{T}^{\star}_{n,l}$ and $\acute{T}^{\star}_{n,l}$ will be equal for large enough $n$ with probability tending to 1, uniformly in $\mathcal{M}_-$, which yields the desired result.

Now suppose that $\Delta(P_0)\neq\emptyset$. Then, for large $n$, we must have $\acute{\Delta}_n\neq\emptyset, \hat{\Delta}_n\neq\emptyset$ with probability tending to one, uniformly in $\mathcal{M}_-$. Applying Lemma B.2 to this case in (8) implies $\acute{\Delta}_n\ominus\hat{\Delta}_n$ converges in probability to the empty set, uniformly in $\mathcal{M}_-$. Therefore,

$$\left(\sup_{t\in[\underline{t},\overline{t}]}\gamma_n^{\star}(t)\right)\int_{\acute{\Delta}_n\ominus\hat{\Delta}_n}\xrightarrow{P}0 \tag{9}$$

conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}_-$. This concludes the proof. □

Proof of Corollary 1:

*Proof.* The proof proceeds by the direct method. Consider the following

$$\begin{aligned}\left|\hat{\Upsilon}_B-\acute{\Upsilon}_B\right| &= \left|\frac{1}{B}\sum_{l=1}^{B}1\left[\hat{T}^{\star}_{n,l}\geq\hat{T}_n\right]-\frac{1}{B}\sum_{l=1}^{B}1\left[\acute{T}^{\star}_{n,l}\geq\hat{T}_n\right]\right| \\ &= \left|\frac{1}{B}\sum_{l=1}^{B}\left(1\left[\hat{T}^{\star}_{n,l}\geq\hat{T}_n\right]-1\left[\acute{T}^{\star}_{n,l}\geq\hat{T}_n\right]\right)\right| \\ &\leq \frac{1}{B}\sum_{l=1}^{B}\left|1\left[\hat{T}^{\star}_{n,l}\geq\hat{T}_n\right]-1\left[\acute{T}^{\star}_{n,l}\geq\hat{T}_n\right]\right| \\ &= \frac{1}{B}\sum_{l=1}^{B}1\left[\hat{T}^{\star}_{n,l}\leq\hat{T}_n\leq\acute{T}^{\star}_{n,l}\quad \text{Xor}\quad \acute{T}^{\star}_{n,l}\leq\hat{T}_n\leq\hat{T}^{\star}_{n,l}\right] \\ &= 1-\frac{1}{B}\sum_{l=1}^{B}1\left[\hat{T}^{\star}_{n,l}\leq\hat{T}_n\leq\acute{T}^{\star}_{n,l}\quad \text{and}\quad \acute{T}^{\star}_{n,l}\leq\hat{T}_n\leq\hat{T}^{\star}_{n,l}\right]\end{aligned} \tag{10}$$

where Xor is the exclusive "or" operator.

The result of Theorem 1 implies

$$\frac{1}{B}\sum_{l=1}^{B} 1\left[\hat{T}^{\star}_{n,l} \leq \hat{T}_n \leq \acute{T}^{\star}_{n,l} \quad \text{and} \quad \acute{T}^{\star}_{n,l} \leq \hat{T}_n \leq \hat{T}^{\star}_{n,l}\right] \xrightarrow{P} 1 \tag{11}$$

conditional on $\mathcal{A}_n$, uniformly in $\mathcal{M}_-$. Therefore, the right side of (10) converges to zero in probability conditional on $\mathcal{A}_n$, uniformly in $\mathcal{M}_-$. This yields the desired result, and concludes the proof. □

# B Auxiliary Results

Let $w \in \mathbb{Z}_+ \cup \{+\infty\}$ , and define the Banach spaces, as indexed by $w$,

$$l^1_w = \left\{ a = (a_1, a_2, \ldots, a_w) \in \mathbb{R}^w : \sum_{j=1}^{w} |a_j| < +\infty \right\}, \tag{12}$$

normed by $\|a\|_{l^1_w} = \sum_{j=1}^{w} |a_j|$.

**Lemma B.1** (Asymptotic Bound for Lagrange Multipliers)**.**

*(i) Define the set of grid points at which the moment conditions are binding as*

$$\Delta(\acute{P}_n) = \left\{ t \in \mathcal{T}_{\mathcal{N}} \,:\, \sum_{i=1}^{n} p'_i g(X_i; t) = 0 \right\}$$

*with cardinality given by* $\omega_n = \left|\Delta(\acute{P}_n)\right|$*. For large* $n$ *and* $P_0 \in \mathcal{M}_-$*, we have* $\Delta(\acute{P}_n) \subset \Delta(P_0)$.

*(ii) Denote the vector of Lagrange multipliers on the constraints* (3) *by* $\boldsymbol{\mu}'$ *and the* $l^1_{\omega_n}$ *norm of the vector* $\boldsymbol{\mu}'$ *by* $||\boldsymbol{\mu}'||_{l^1_{\omega_n}}$*. Then* $||\boldsymbol{\mu}'||_{l^1_{\omega_n}} = o_P(1)$ *uniformly in* $\mathcal{M}_-$.

*Proof.*

(i) We show this result using proof by contrapositive, that is, we show that for large $n$,

$$t \notin \Delta(P_0) \implies t \notin \Delta(\acute{P}_n)$$

Consider $P_0 \in \mathcal{M}_-$ and any $t \in [\underline{t}, \overline{t}]$. From the first part of this lemma,

$$\sum_{i=1}^{n} p_i' g(X_i;t) \leq \frac{1}{n}\sum_{i=1}^{n} g(X_i;t) = \frac{1}{n}\sum_{i=1}^{n} g(X_i;t) - E_{P_0}\left[g(X;t)\right] + E_{P_0}\left[g(X;t)\right] \quad (13)$$

Now, consider $t \notin \Delta(P_0)$. As $P_0 \in \mathcal{M}_-$, this implies that $E_{P_0}\left[g(X;t)\right] < 0$. By the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i;t) - E_{P_0}\left[g(X;t)\right] = O_P(n^{-1/2})$$

uniformly in $\mathcal{M}_-$. Thus, for sufficiently large $n$, equation (13) simplifies to

$$\sum_{i=1}^{n} p_i' g(X_i;t) < 0$$

This shows that $t \notin \Delta(\acute{P}_n)$.

(ii) Recall that the cardinality of the set $\Delta(\acute{P}_n)$ is $\omega_n \leq N(n)$. By complementary slackness, for any $t \notin \Delta(\acute{P}_n)$, $\mu' g(X;t) = 0$. This allows the REL probabilities to be written as

$$p_i' = \frac{1}{n}\left(1 + \sum_{j=1}^{\omega_n} \mu_j' g(X_i;t_j)\right)^{-1} \quad (14)$$

For any choice of $t_j \in \Delta(\acute{P}_n)$, we have

$$\sum_{i=1}^{n} p_i' g(X_i;t_j) = \frac{1}{n}\sum_{i=1}^{n} \frac{g(X_i;t_j)}{1 + \sum_{j=1}^{\omega_n} \mu_j' g(X_i;t_j)} = 0 \quad (15)$$

To express the system of equations described by (15) in vectorised form, define the vector

$$\mathbf{g}_i = \left[g(X_i;t_1),\, g(X_i;t_2),\, \ldots,\, g(X_i;t_{\omega_n})\right]^T \quad (16)$$

Now, as all the elements of $\boldsymbol{\mu'}$ are non-negative, the $l^1_{\omega_n}$ norm is simply the sum of all elements of $\boldsymbol{\mu'}$, i.e. $||\boldsymbol{\mu'}||_{l^1_{\omega_n}} = \sum_{j=1}^{\omega_n} \mu'_j$. This means we can express the vector $\boldsymbol{\mu'}$ in the form

$$\boldsymbol{\mu'} = ||\boldsymbol{\mu'}||_{l^1_{\omega_n}} \boldsymbol{\theta}\,, \quad \boldsymbol{\theta} \in \mathbb{R}_+^{\omega_n}$$

Under this construction, the $j^{th}$ element of $\boldsymbol{\theta}$ is

$$\theta_j = \frac{\mu'_j}{\sum_{j=1}^{\omega_n} \mu'_j}$$

This implies that $\sum_{j=1}^{\omega_n} \theta_j = 1$. The system of equations defined by (15) for all $t \in \Delta(\acute{P}_n)$ can be written in the following form

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{g}_i}{1+(\boldsymbol{\mu'})^T\mathbf{g}_i} = 0 \implies \boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{g}_i}{1+(\boldsymbol{\mu'})^T\mathbf{g}_i}\right) = 0 \tag{17}$$

Define the quantity $Y_i = (\boldsymbol{\mu'})^T\mathbf{g}_i$. Using the manipulation $\frac{1}{1+Y_i} = 1 - \frac{Y_i}{1+Y_i}$ and the fact that $(\boldsymbol{\mu'})^T\mathbf{g}_i = \mathbf{g}_i^T\boldsymbol{\mu'}$ in equation (17) gives

$$\begin{aligned}
\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\left(1-\frac{\mathbf{g}_i^T\boldsymbol{\mu'}}{1+Y_i}\right)\right) &= 0 \\
\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right) &= \boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{g}_i\mathbf{g}_i^T\boldsymbol{\mu'}}{1+Y_i}\right) \\
\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right) &= \boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{g}_i\mathbf{g}_i^T||\boldsymbol{\mu'}||\boldsymbol{\theta}}{1+Y_i}\right) \\
\therefore \boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\right) &= ||\boldsymbol{\mu'}||_{l^1_{\omega_n}}\boldsymbol{\theta}^T\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{g}_i\mathbf{g}_i^T}{1+Y_i}\right)\boldsymbol{\theta}
\end{aligned} \tag{18}$$

We denote the sample analogue estimate of the covariance matrix of measurement functions over the set of all $t \in \Delta(\acute{P}_n)$ by

$$\hat{\Sigma}_{\Delta(\acute{P}_n)} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{g}_i\mathbf{g}_i^T$$

Define $Y_{max} = \max_i |Y_i|$. Note that

$$Y_{max} = \max_i |Y_i| = \max_i \sum_{j=1}^{\omega_n} \mu'_j |g(X_i; t_j)| \leq \sum_{j=1}^{\omega_n} \mu'_j = ||\boldsymbol{\mu'}||_{l^1_{\omega_n}} \tag{19}$$

where we have used the uniform boundedness of $g$. This follows from the compact connected support of the marginal distributions.

Now, consider

$$\begin{aligned}
||\boldsymbol{\mu'}||_{l^1_{\omega_n}} \left( \boldsymbol{\theta}^T \hat{\Sigma}_{\Delta(\acute{P}_n)} \boldsymbol{\theta} \right) &= ||\boldsymbol{\mu'}||_{l^1_{\omega_n}} \left( \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \mathbf{g}_i^T \right) \boldsymbol{\theta} \right) \\
&\leq ||\boldsymbol{\mu'}||_{l^1_{\omega_n}} \left( \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{g}_i \mathbf{g}_i^T}{1 + Y_i} \right) \boldsymbol{\theta} \right) (1 + Y_{max}) \\
&\leq ||\boldsymbol{\mu'}||_{l^1_{\omega_n}} \left( \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{g}_i \mathbf{g}_i^T}{1 + Y_i} \right) \boldsymbol{\theta} \right) (1 + ||\boldsymbol{\mu'}||_{l^1_{\omega_n}}) \\
\therefore ||\boldsymbol{\mu'}||_{l^1_{\omega_n}} \left( \boldsymbol{\theta}^T \hat{\Sigma}_{\Delta(\acute{P}_n)} \boldsymbol{\theta} \right) &\leq \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \right) (1 + ||\boldsymbol{\mu'}||_{l^1_{\omega_n}}) \qquad (20)
\end{aligned}$$

where the last line results from substituting the expression given in (18). Rearranging (20) gives

$$||\boldsymbol{\mu'}||_{l^1_{\omega_n}} \left[ \boldsymbol{\theta}^T \hat{\Sigma}_{\omega_n} \boldsymbol{\theta} - \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \right) \right] \leq \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \right) \tag{21}$$

We consider the components of (21) to find the required asymptotic bound on $||\boldsymbol{\mu'}||$. From part (ii) of this lemma, for large n we have $\Delta(\acute{P}_n) \subset \Delta(P_0)$. This means for large $n$, we have that for all $t \in \Delta(\acute{P}_n)$, $E_{P_0}[g(X;t_j)] = 0$. As a result,

$$\begin{aligned}
\boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \right) &= \sum_{j=1}^{\omega_n} \theta_j \left( \frac{1}{n} \sum_{i=1}^{n} g(X_i;t_j) - E_{P_0}[g(X;t_j)] \right) \\
\left| \boldsymbol{\theta}^T \left( \frac{1}{n} \sum_{i=1}^{n} \mathbf{g}_i \right) \right| &\leq \sum_{j=1}^{\omega_n} \theta_j \left| \frac{1}{n} \sum_{i=1}^{n} g(X_i;t_j) - E_{P_0}[g(X;t_j)] \right| \\
&\leq \max_j \left| \frac{1}{n} \sum_{i=1}^{n} g(X_i;t_j) - E_{P_0}[g(X;t_j)] \right| \left( \sum_{j=1}^{\omega_n} \theta_j \right) \\
&\leq \sup_{t \in \mathcal{T}} \left| \frac{1}{n} \sum_{i=1}^{n} g(X_i;t_j) - E_{P_0}[g(X;t_j)] \right| \qquad (22)
\end{aligned}$$

The last line follows from the fact that $\sum_{j=1}^{\omega_n} \theta_j = 1$ by construction. The upper bound given by equation (22) is $o_P(1)$ uniformly in $\mathcal{M}_-$. This follows from the functions being of Vapnik-Chervonenkis class. The moment functions $g$ belonging to a uniformly bounded Vapnik-Chervonenkis class of functions ensures that class of functions is also uniformly Glivenko-Cantelli.

Now, for sufficiently large $n$, part (ii) of this lemma tells us that $\Delta(\acute{P}_n) \subset \Delta(P_0)$. From assumption 2.1(iii), for the finite subset $\Delta(\acute{P}_n) \subset \Delta(P_0)$ the covariance matrix of measurement functions satisfies $\boldsymbol{\theta}^T \Sigma_{\Delta(\acute{P}_n)} \boldsymbol{\theta} \geq c > 0$. Using this result and the bound from equation (22), we can rewrite (21) as

$$||\boldsymbol{\mu}'||_{l^1_{\omega_n}} \leq \frac{o_P(1)}{c + o_P(1)} \qquad (23)$$

As this holds for all $P \in \mathcal{M}_-$, equation (23) shows that $||\boldsymbol{\mu}'||_{l^1_{\omega_n}} = o_P(1)$ uniformly in $\mathcal{M}_-$.

□

**Lemma B.2.** *Suppose that* $P_0 \in \mathcal{M}_-$. *Then Prob* $\left[\hat{\Delta}_n = \acute{\Delta}_n\right] \to 1$ *as* $n \to +\infty$, *uniformly in* $\mathcal{M}_-$.

*Proof.* First define the following quantities

$$\begin{aligned}\hat{\psi}(t) &= \frac{1}{n}\sum_{i=1}^{n} g(X_i;t) \\ \psi'(t) &= \sum_{i=1}^{n} p_i' g(X_i;t)\end{aligned} \tag{24}$$

The two contact sets can now be expressed in the following form

$$\begin{aligned}\hat{\Delta}_n &= \left\{t \in [\underline{t}, \overline{t}] \,:\, \left|\hat{\psi}(t)\right| \le r_n\right\} \\ \acute{\Delta}_n &= \left\{t \in [\underline{t}, \overline{t}] \,:\, |\psi'(t)| \le r_n\right\}\end{aligned} \tag{25}$$

Now, consider the following

$$\begin{aligned}\left|\hat{\psi}(t) - \psi'(t)\right| &= \left|\sum_{i=1}^{n} \frac{1}{n} g(X_i;t) - \sum_{i=1}^{n} p_i' g(X_i;t)\right| \\ &\le \sum_{i=1}^{n} \left|\left(\frac{1}{n} - p_i'\right) g(X_i;t)\right| \\ &\le \sum_{i=1}^{n} \left|\frac{1}{n}\left(1 - \frac{1}{1+\sum_{j=1}^{N} \mu_j' g(X_i;t_j)}\right)\right| \\ &= \sum_{i=1}^{n} \left|\frac{1}{n} \cdot \frac{\sum_{j=1}^{N} \mu_j' g(X_i;t_j)}{1+\sum_{j=1}^{N} \mu_j' g(X_i;t_j)}\right| \\ &= \sum_{i=1}^{n} \left|p_i' \sum_{j=1}^{N} \mu_j' g(X_i;t_j)\right| \\ &\le \sum_{i=1}^{n} p_i' \left|\sum_{j=1}^{N} \mu_j'\right| \\ &\le \sum_{j=1}^{N} |\mu_j'| \\ &= ||\boldsymbol{\mu'}||_{l^1_{\omega_n}}\end{aligned} \tag{26}$$

Lemma B.1(ii) gives the required rate of convergence of the difference between the behaviour of $\hat{\psi}(t)$ and $\psi'(t)$ in the form of $||\boldsymbol{\mu'}||$. Firstly, consider $t \in \hat{\Delta}_n$. For $P_0 \in \mathcal{M}_-$, this implies that

$|\hat{\psi}(t)| \leq r_n$. We then have

$$
\begin{aligned}
|\psi'(t)| &= |\psi'(t) - \hat{\psi}(t) + \hat{\psi}(t)| \\
&\leq |\psi'(t) - \hat{\psi}(t)| + |\hat{\psi}(t)| \\
&= ||\boldsymbol{\mu'}||_{l^1_{\omega_n}} + r_n \\
&= r_n, \quad \text{as } n \to +\infty
\end{aligned}
$$

The last equality follows as by construction, $r_n$ has a much slower rate of convergence than that of $||\boldsymbol{\mu'}||$. Hence, $|\psi'(t)| \leq r_n$ and so $t \in \acute{\Delta}_n$. This shows that $\hat{\Delta}_n \subset \acute{\Delta}_n$.

Next, consider $t \in \acute{\Delta}_n$. For $P_0 \in \mathcal{M}_-$, this implies that $|\psi'(t)| \leq r_n$. We then have

$$
\begin{aligned}
|\hat{\psi}(t)| &= |\hat{\psi}(t) - \psi'(t) + \psi'(t)| \\
&\leq |\psi'(t) - \hat{\psi}(t)| + |\psi'(t)| \\
&= ||\boldsymbol{\mu'}||_{l^1_{\omega_n}} + r_n \\
&= r_n, \quad \text{as } n \to +\infty
\end{aligned}
$$

Hence, $|\hat{\psi}(t)| \leq r_n$ and so $t \in \hat{\Delta}_n$. This shows that $\acute{\Delta}_n \subset \hat{\Delta}_n$. Combining the last two results completes the proof. $\square$

**Lemma B.3.** *Suppose that $P_0 \in \mathcal{M}_-$. Then Prob $\left[\hat{\Delta}_n = \Delta\left(P_0\right)\right] \to 1$ as $n \to +\infty$, uniformly in $\mathcal{M}_-$.*

*Proof.* First, we prove the Prob $\left[\hat{\Delta}_n \subset \Delta\left(P_0\right)\right] \to 1$ as $n \to +\infty$, uniformly in $\mathcal{M}_-$. The proof proceeds by contraposition. Consider $P_0 \in \mathcal{M}_-$ and any $t \in [\underline{t}, \overline{t}]$, we show that for large $n$, the probability of

$$t \notin \Delta(P_0) \implies t \notin \hat{\Delta}_n$$

tends to 1, uniformly in $\mathcal{M}_-$. We have

$$\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i;t)\right| = \left|\frac{1}{n}\sum_{i=1}^{n} g(X_i;t) - E_{P_0}\left[g(X;t)\right] + E_{P_0}\left[g(X;t)\right]\right|. \tag{27}$$

Since the set of moment functions $\left\{\mathbf{x} \mapsto g(\mathbf{x},t), t \in [\underline{t}, \overline{t}]\right\}$ is uniform Donsker with respect to $\mathcal{M}_-$, we have $\frac{1}{n}\sum_{i=1}^{n} g(X_i;t) - E_{P_0}\left[g(X;t)\right] = o_P(1)$ uniformly in $\mathcal{M}_-$, at the $\sqrt{n}$-rate. Therefore,

$$\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i;t)\right| = \left|O_P(\sqrt{n}) + E_{P_0}\left[g(X;t)\right]\right| \quad \text{uniformly in} \quad \mathcal{M}_-. \tag{28}$$

Since $r_n = o_P(1)$ uniformly in $\mathcal{M}_-$, slower than the $\sqrt{n}$-rate, the comparison of

$$\left|O_P(\sqrt{n}) + E_{P_0}\left[g(X;t)\right]\right|$$

and $r_n$ is asymptotically equivalent to the comparison of $\left|E_{P_0}\left[g(X;t)\right]\right|$ and $r_n$, which implies that $\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i;t)\right| > r_n$ as $n \to +\infty$, uniformly in $\mathcal{M}_-$. Therefore, the probability of $t \notin \hat{\Delta}_n$ tends to unity, uniformly in $\mathcal{M}_-$.

Now we prove the reverse uniform asymptotic set inclusion. That is, we show that for large $n$, the probability of

$$t \in \Delta(P_0) \implies t \in \hat{\Delta}_n$$

tends to unity, uniformly in $\mathcal{M}_-$. We have

$$\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i;t)\right| = \left|\frac{1}{n}\sum_{i=1}^{n} g(X_i;t) - E_{P_0}\left[g(X;t)\right]\right|. \tag{29}$$

By the same arguments used to prove the first part, $\frac{1}{n}\sum_{i=1}^{n} g(X_i;t) - E_{P_0}\left[g(X;t)\right] = o_P(1)$ uniformly in $\mathcal{M}_-$, at the $\sqrt{n}$-rate. Therefore, $\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i;t) - E_{P_0}\left[g(X;t)\right]\right| = o_P(1)$ uniformly in $\mathcal{M}_-$, at the $\sqrt{n}$-rate. So that $\left|\frac{1}{n}\sum_{i=1}^{n} g(X_i;t) - E_{P_0}\left[g(X;t)\right]\right| \leq r_n$ as $n \to +\infty$, uniformly in $\mathcal{M}_-$. Therefore, the probability of $t \in \hat{\Delta}_n$ tends to unity, uniformly in $\mathcal{M}_-$. This concludes the proof. $\square$