THE UNIVERSITY OF SYDNEY

Economics Working Paper Series

2015 - 2

Empirical Likelihood for Robust Poverty Comparisons

Rami Victor Tabri

May 2015

Empirical Likelihood for Robust Poverty Comparisons

Rami V. Tabri

School of Economics, The University of Sydney,

Sydney, New South Wales 2006, Australia

Version: Article

May 4, 2015

Abstract

Robust rankings of poverty are ones that do not rely on a single poverty measure with a single poverty line. Mathematically, such robust rankings of two populations specifies a continuum of unconditional moment inequality constraints. If these constraints could be imposed in estimation then a statistical test can be performed using an empirical likelihood-ratio (ELR) test, which is a nonparametric version of the likelihood-ratio test in parametric inference. While these constraints cannot be imposed exactly, we show that these can be imposed approximately with the approximation disappearing asymptotically. We then propose a bootstrap test procedure that implements the resulting approximate ELR test. The paper derives the asymptotic properties of this test, presents Monte Carlo experiments that show improved power compared to existing tests such as that of Linton et al. (2010), and provides an empirical illustration to Canadian income distribution data. More generally, the bootstrap test procedure provides a uniformly asymptotically valid nonparametric test of a continuum of unconditional moment inequality constraints. The proofs exploit the fact that the constrained optimization problem is a concave semi-infinite programming optimization problem.

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

Empirical Likelihood; Robust Poverty Comparison; Continuum of Moment Inequality Constraints; Bootstrap.

1 Introduction

The comparison of income distributions in studies of poverty is an important component in the economist's toolbox. The simplest rankings are based on a single poverty measure with a fixed poverty line, such as the proportion of households with incomes below $5000. Such rankings are very limited as they are based on a single measure and a single poverty line: they may produce contradictory conclusions at two different yet equally reasonable poverty lines or poverty measures.

Recently, attention has focused on robust one-way poverty comparisons in the sense that the ranking between the income distributions is unanimous across multiple poverty measures or a set of poverty lines. Specifically, a poverty-line ranking orders the distributions using a single poverty measure over a range of poverty lines, rather than a single poverty line. And a poverty-measure ranking orders the distributions using a single poverty line over a range of poverty measures where the poverty measures are in a pre-specified family that satisfies certain ethically desirable criteria (axioms). These two robust rankings are conceptually distinct but interconnected[1].

Since population income distributions are not observable in practice, a statistical test is employed to rank the distributions using sample data on incomes. As they are cumulative distribution functions (CDFs) of income, income distributions can be treated as CDFs of random variables for the purposes of estimation and tests. In this respect, a poverty measure with a fixed poverty line is an unconditional moment of a given income distribution: it indicates the extent of poverty associated with the distribution under consideration. Thus, ranking two income distributions using a single poverty measure with fixed poverty lines is characterized by an inequality restriction on these moments. Extrapolating from this case, a robust ranking corresponds to an infinite number of inequality restrictions on certain moments of the distributions. For this reason, a statistical test for such a ranking entails testing for an infinite number of moment inequality restrictions.

This paper proposes a nonparametric bootstrap test for the null hypothesis that a given robust one-way poverty comparison holds between two income distributions. The proposed test procedure uses the method of empirical likelihood (Owen, 1988; Qin and Lawless, 1994; Imbens et al., 1998; Kitamura, 2001). It is a nonparametric likelihood-based procedure that produces data-determined shapes for the distributions, and it is particularly appropriate in our setting where there are many moment inequality constraints imposed in estimation. We test the null hypothesis using the empirical likelihood-ratio (ELR) test statistic, which is a nonparametric counterpart of the likelihood-ratio statistic in parametric inferences. The advantage is that this test statistic has the internal Studentization property and accounts for the correlation between the different moment

[1] See Zheng (2000) for more on this point.

inequalities, which leads to a less conservative test.

The test this paper proposes uses concave *semi-infinite programming* (SIP) methods (Hettich and Kortanek, 1993; López and Still, 2007; Shapiro, 2009) to extend Canay (2010)'s empirical likelihood bootstrap, which is for a finite number of moment inequality restrictions, to the setting of the paper, where there is an *infinite* number of unconditional moment inequality constraints. To establish the uniform validity of our testing procedure, we derive the uniform asymptotic null distribution of the ELR statistic. This result extends the asymptotic distribution theory of this statistic for a finite number of inequality constraints (e.g. El Barmi, 1996) to the infinite case. The general theory is presented in such a way that it can be applied to any setting where the null hypothesis has infinitely many unconditional moment inequality constraints. Therefore, this result is of independent interest and significantly advances the current literature on constrained statistical inference. Another attractive feature of this paper is that the general infinite number of constraints case is motivated by the need to solve an important problem in poverty and income studies.

We also analyze the finite sample properties of our test in Monte Carlo simulation experiments using models for second and third orders of restricted stochastic dominance. We compare its performance with the bootstrap tests of Linton et al. (2010) (LSW) and Andrews and Shi (2010) (AS). LSW propose a bootstrap test for traditional unrestricted stochastic dominance under the null using one-sided integral-type test statistic that is a functional of the sample analogue estimator of the moments. Their method also applies to restricted stochastic dominance, which presents a point of comparison with the current work. AS also propose a bootstrap test, but for models defined by many (possibly an infinite number of) conditional moment inequalities and/or equalities. The framework of AS covers the models in this paper. In the paper's setup, the test statistics AS propose reduce to a one-sided Kolmogorov-Smirnov test statistic that is a functional of the Studentized sample analogue estimator of the moments. In contrast to the ELR test statistic, neither of the test statistics AS and LSW propose accounts for the correlation across the moment inequality constraints.

In the simulation results for these two models, all the of the procedures are found to control test level well in moderate to large sample sizes. The LSW test is found to be conservative relative to the AS and proposed tests, whereas the AS and proposed tests behave similarly. The test this paper proposes outperforms the LSW and AS tests in terms of power against alternatives with some non-violated inequalities[2], with the AS test outperforming the LSW test. Furthermore, the proposed test's power is substantially higher than the AS and LSW tests against such DGPs that are closer to the null. Finally, all of the tests have similar power properties against alternatives with population moment functions that have a continuum of binding moment inequalities.

This paper contributes to the literature on inference for robust poverty comparisons. The literature has focused almost exclusively on tests for traditional unrestricted stochastic dominance orderings. Some examples include McFadden (1989), Barrett and Donald (2003), Linton et al. (2005), Horváth et al. (2006), and LSW. From a normative perspective, such rankings are deficient because they do not give equal ethical weight to all those who are below their respective poverty lines, whereas the rankings based on the *restricted* stochastic dominance conditions do not suffer from this deficiency[3]. We contribute to this literature by considering tests for restricted stochastic dominance under the null, and more broadly, tests for other robust rankings.

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. Their approach is convenient since a rejection of this null entails acceptance of the only other possibility which is restrict stochastic dominance; however, their alternative hypothesis is a strong form of restricted stochastic dominance, which implies the null hypothesis posited in the paper. Thus, the proposed test procedure complements the aforementioned ones.

This paper also contributes to the growing literature on inference for models defined by an infinite number of unconditional moment inequality restrictions. There are several papers on inference for conditional moment inequalities, which can be treated as an infinite number of unconditional moment inequalities; see, for example, Andrews and Shi (2013, 2014) and Chernozhukov et al.

[2]The terminology for describing such alternatives is borrowed from Andrews (2011); it refers to alternatives with some positive moments and some moments that are negative and moderately small.

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

(2013). In contrast, the infinite number of inequalities in the paper are not generated from a conditional moment inequality model. The papers closest to the present work are AS and LSW. AS extends Andrews and Shi (2013) to cover models defined by infinitely many conditional moment inequalities and/or equalities. Their extension also covers the case of models defined by an infinite number of unconditional moment inequalities.

The approach of this paper differs from that of LSW and AS in two important ways. Firstly, the paper proposes the use of a likelihood-based test statistic, while AS and LSW propose one-sided test statistics based on the sample analogue estimator of the population moments. Secondly, the paper's approach to bootstrapping formulates the bootstrap data-generating process (DGP) using a constrained estimator of the underlying distributions. Whereas LSW and AS propose the use of the empirical CDF of the data as the bootstrap DGP, which does not incorporate the statistical information from imposing the constraints that define the null hypothesis. Therefore, the main technical contribution of this paper is to introduce a new method of testing that applies to cases in which an infinite number of unconditional moment inequalities defines the null hypothesis.

The rest of this paper is organized as follows. Section 2 describes robust one-way poverty comparisons within the framework of a moment inequality model, presents examples, and introduces the model of the null hypothesis. Section 3 defines the ELR statistic, provides its uniform asymptotic null distribution, and specifies a computational algorithm for computing it in practice. Section 4 describes the proposed empirical likelihood bootstrap test procedure, establishes its uniform asymptotic validity, and its consistency against all fixed alternatives. Section 5 provides the Monte Carlo simulation results. Section 6 illustrates the proposed method using data from the Canadian Family Expenditures survey for the year 1986. Section 7 concludes, and Section 8 collates the acknowledgements of the individuals and institutions who provided help during the research. All proofs are relegated to the Appendix.

2 Robust One-Way Poverty Comparisons

2.1 Setup

The models described in this paper are of the following general form:

$$E_{P_0}\left[g\left(\mathbf{X};t\right)\right] \leq 0 \quad \forall t \in [\underline{t},\overline{t}], \tag{2.1}$$

where $[\underline{t},\overline{t}] \subset \mathbb{R}$ is a predesignated compact interval, the observations $\{\mathbf{X}_i\}_{i=1}^n$ are a bivariate random sample on two populations A and B with typical element $\mathbf{X} = [X^A, X^B]$, P_0 is the unknown true distribution of $\mathbf{X}_i$ with respect to the measurable space $(\mathcal{X}, \mathcal{A})$, where $\mathcal{X} = [0,\overline{s}] \times [0,\overline{s}]$ is the sample space of jointly observable incomes from the two populations with $\overline{s} \in (0,+\infty)$ is known, and $\mathcal{A}$ is the Borel sigma algebra on $\mathcal{X}$. Furthermore, $g\left(\mathbf{x};t\right) = h\left(x^B;t\right) - h\left(x^A;t\right)$ where h is a known moment function that is weakly monotonic in its first argument for each t.

The object of interest is P_0, which is partially identified. We are interested in testing that P_0 satisfies the moment inequalities (2.1) under the null hypothesis. The next section characterizes robust one-way poverty comparisons within the framework of the moment inequality model (2.1), and presents a couple of examples.

2.2 Examples

The robust one-way comparisons of two populations are either poverty-line rankings or poverty-measure rankings. The former carries out the one-way comparison over a predesignated set of poverty measures with given poverty lines, whereas the latter fixes a poverty measure and compares the distributions over a given set of poverty lines. A poverty measure has the general form $\int h\left(x;z,\gamma\right)\, dL(x)$ where z is a poverty line[4], x is income, $L(x)$ is an income distribution, and γ is poverty aversion parameter that indexes the poverty measure within a pre-specified family of such measures. The function $h\left(x;z,\gamma\right) \geq 0$ is the poverty contribution to total poverty of someone with

[4] A poverty line in a population is the threshold below which one is considered to be poor.

income x, with $h(x; z, \gamma) = 0$ whenever $x > z$. Furthermore, $h(x; z, \gamma)$ is weakly monotonic in x for each z and γ.

For poverty-measure rankings, z^A, z^B are given and let $\Gamma \subset \mathbb{R}$ denote a set of poverty aversion parameters. Then, we say A has more poverty than B according to the poverty measures defined by the moment functions in $\{\pi(\cdot\,; \cdot, \gamma), \gamma \in \Gamma\}$ if

$$E_{P_0}\left[h\left(X^B; z^B, \gamma\right) - h\left(X^A; z^A, \gamma\right)\right] \leq 0 \quad \forall \gamma \in \Gamma. \tag{2.2}$$

For a poverty-measure ranking, γ is fixed and let $\mathcal{Z} \subset \mathbb{R}_+$ be a given set of poverty lines. Then, we say A has more poverty than B according to the poverty measures defined by the moment functions in $\{h(\cdot\,; z, \gamma), z \in \mathcal{Z}\}$ if

$$E_{P_0}\left[h\left(X^B; z, \gamma\right) - h\left(X^A; z, \gamma\right)\right] \leq 0 \quad \forall z \in \mathcal{Z}. \tag{2.3}$$

We now present two examples of robust one-way poverty comparisons.

Example 1 (The First Clark, Hemming and Ulf Family). Clark et al. (1981) proposed the following family of poverty measures: $\frac{1}{\gamma}\int_0^z \left[1 - \left(\frac{x}{z}\right)^\gamma\right] dL(x)$ where $\gamma \geq 1$. Income distributions can be ordered using this family across γ for given poverty lines, or across poverty lines for a given γ. In the former setting, let $1 \leq \underline{\gamma} < \overline{\gamma} < +\infty$, then the moment functions are

$$\mathbf{x} \mapsto \gamma^{-1}\left[1 - \left(\frac{x^B}{z^B}\right)^\gamma\right] 1\left[x^B \leq z^B\right] - \gamma^{-1}\left[1 - \left(\frac{x^A}{z^A}\right)^\gamma\right] 1\left[x^A \leq z^A\right] \quad \gamma \in [\underline{\gamma}, \overline{\gamma}],$$

where z^K, $K = A, B$ are given poverty lines. In the latter setting, let $0 < \underline{z} < \overline{z} \leq \overline{s}$, then the moment functions, with γ given, are

$$\mathbf{x} \mapsto \gamma^{-1}\left[1 - \left(\frac{x^B}{z}\right)^\gamma\right] 1\left[x^B \leq z\right] - \gamma^{-1}\left[1 - \left(\frac{x^A}{z}\right)^\gamma\right] 1\left[x^A \leq z\right] \quad z \in [\underline{z}, \overline{z}].$$

Example 2 (The Foster, Greer and Thorbecke Family). Foster et al. (1984) proposed the following family of poverty measures: $\frac{1}{\gamma}\int_0^z \left(\frac{z-x}{z}\right)^{\gamma} dL(x)$ where $\gamma \geq 0$. As in the previous example, income distributions can be ordered using this family either across a set of poverty aversion parameters or across a set of poverty lines. In the former setting, let $0 \leq \underline{\gamma} < \overline{\gamma} < +\infty$ and z^A, z^B are given. Then, the moment functions are

$$\mathbf{x} \mapsto \left(\frac{z^B - x^B}{z^B}\right)^{\gamma} 1\left[x^B \leq z^B\right] - \left(\frac{z^A - x^A}{z^A}\right)^{\gamma} 1\left[x^A \leq z^A\right] \quad \gamma \in [\underline{\gamma}, \overline{\gamma}].$$

For the poverty-line rankings, γ is fixed and let $z \in [\underline{z}, \overline{z}]$. Then, the moment functions are

$$\mathbf{x} \mapsto \left(\frac{z - x^B}{z}\right)^{\gamma} 1\left[x^B \leq z\right] - \left(\frac{z - x^A}{z}\right)^{\gamma} 1\left[x^A \leq z\right] \quad z \in [\underline{z}, \overline{z}].$$

For $\gamma = 0$, this poverty-line ranking fixes the *headcount ratio* as the poverty measure, and for $\gamma = 1$ it fixes the *per capita income gap*. Foster and Shorrocks (1988) called these poverty-line rankings "poverty orderings", and proved that they are in a one-to-one correspondence with the rankings based on stochastic dominance conditions. Specifically, they showed for $\gamma = 0, \underline{z} = 0$ and $\overline{z} = \overline{s}$, these moment functions correspond to the ones that define first-order stochastic dominance conditions. And more generally, for γ equal to a positive integer, this poverty-line ranking is equivalent to the ranking based on the $(\gamma + 1)$-order stochastic dominance conditions.

The next section introduces regularity conditions on the moment functions that covers a broad range of robust one-way poverty comparisons.

2.3 Conditions on Moment Functions

Without loss of generality, we represent the moment functions in (2.2) and (2.3) by the set of functions $\left\{\mathbf{x} \mapsto g(\mathbf{x}; t), t \in [\underline{t}, \overline{t}]\right\}$. In this notation, the index parameter t can either be a poverty measure in a pre-specified family, or a poverty line, so that the index parameter over which the comparison is *not* being conducted is suppressed for notational simplicity.

The set $\{\mathbf{x} \mapsto g(\mathbf{x};t),\ t \in [\underline{t}, \overline{t}]\}$ satisfies the following assumption:

Assumption 2.1. *(i) For each* $\mathbf{x} \in \mathcal{X}$, $t \mapsto g(\mathbf{x};t)$ *is a Lipschitz continuous function on* $[\underline{t}, \overline{t}]$ *with known Lipschitz constant,* L; *(ii)* $-1 \leq g \leq 1$, $\forall(\mathbf{x}, t) \in \mathcal{X} \times [\underline{t}, \overline{t}]$; *(iii) Pointwise measurable; (iv) Vapnik-Červonenkis class, and (v)* $\nexists t' \in [\underline{t}, \overline{t}]$ *such that* $g(\mathbf{x}; t') = 0\ \forall \mathbf{x} \in \mathcal{X}$.

Many sets of moment functions that define poverty measures satisfy the conditions in Assumptions 2.1. Condition (i) is standard in the concave SIP optimization literature. It also excludes moment functions that depend discontinuously on the parameter that indexes them, such as the ones that define the first-order stochastic dominance conditions. Condition (v) excludes moment functions that vanish uniformly over the sample space. To understand the impact of this condition, consider the moment functions for the traditional unrestricted s-th order stochastic dominance conditions:

$$\mathbf{x} \mapsto \frac{\left(t - x^B\right)^{s-1}}{(s-1)!} 1\left[x^B \leq t\right] - \frac{\left(t - x^A\right)^{s-1}}{(s-1)!} 1\left[x^A \leq t\right] \quad t \in [0, \overline{s}]. \tag{2.4}$$

Condition (v) excludes the moment function corresponding to $t = 0$ since it is equal to zero for every $\mathbf{x} \in \mathcal{X}$. Furthermore, the compactess of the interval $[\underline{t}, \overline{t}]$ forces $\underline{t} > 0$. Therefore, these conditions exclude the traditional unrestricted stochastic dominance conditions from our analysis. As already mentioned in Section 1, such rankings are deficient from a normative perspective; for this reason, the paper does not focus on them.

Conditions (i) and (v) justify the existence of Lagrange multiplier variables (via a Slater condition), whose large sample properties establish the asymptotic behavior of the ELR test statistic. The value of the bounds in condition (ii) are not important for the validity of the proposed method and are made for simplicity; all we require is that the moment functions are uniformly bounded with known bounds. Conditions (iii) and (iv) are important for developing the large sample properties of the ELR test statistic using empirical process theory. The pointwise measurability of this set is to ensure the measurability of the quantities we are interested in; see Appendix B for their formal definitions and a discussion of how to verify them in practice.

The bootstrap test procedure this paper proposes applies to both types of robust one-way poverty comparisons. It is uniformly asymptotically valid under the model of the null hypothesis, and we present this model in the next section.

2.4 Null Parameter Space

The bootstrap test procedure uses the ELR test statistic in testing whether P_0 satisfies (2.1) under the null hypothesis. The asymptotic behavior of this test statistic depends on the form of the contact set

$$\Delta\left(P_0\right) = \left\{t \in \left[\underline{t}, \overline{t}\right] : E_{P_0}\left[g\left(\mathbf{X}; t\right)\right] = 0\right\}, \tag{2.5}$$

and on the properties of the covariances of the random variables $\left\{g\left(\mathbf{X}; t\right), \ t \in \Delta\left(P_0\right)\right\}$. The setup allows for a continuum of binding moment inequalities, and the Lipschitz continuity of the moment functions excludes the case in which a countable number are binding.[5]. In general, $\Delta\left(P_0\right) = \Delta_d\left(P_0\right) \cup \Delta_c\left(P_0\right)$, where $\Delta_d\left(P_0\right)$ is the set of isolated points in $\Delta\left(P_0\right)$, and $\Delta_c\left(P_0\right)$ is union of the connected parts of $\Delta\left(P_0\right)$.

If $\Delta\left(P_0\right) = \Delta_d\left(P_0\right) = \left\{t_1^b, t_2^b, \ldots, t_m^b\right\}$ where $m \in \mathbb{Z}_+$, then we denote by $\Sigma_m\left(P_0\right)$ the covariance matrix formed by the random variables $\left\{g\left(\mathbf{X}; t_j^b\right), \ j = 1, \ldots, m\right\}$, which is given by

$$\begin{pmatrix} E_{P_0}\left[g\left(\mathbf{X}; t_1^b\right)\right]^2 & E_{P_0}\left[g\left(\mathbf{X}; t_1^b\right) g\left(\mathbf{X}; t_2^b\right)\right] & \cdots & E_{P_0}\left[g\left(\mathbf{X}; t_1^b\right) g\left(\mathbf{X}; t_m^b\right)\right] \\ E_{P_0}\left[g\left(\mathbf{X}; t_2^b\right) g\left(\mathbf{X}; t_1^b\right)\right] & E_{P_0}\left[g\left(\mathbf{X}; t_2^b\right) g\left(\mathbf{X}; t_2^b\right)\right] & \cdots & E_{P_0}\left[g\left(\mathbf{X}; t_2^b\right) g\left(\mathbf{X}; t_m^b\right)\right] \\ \vdots & \vdots & \ddots & \vdots \\ E_{P_0}\left[g\left(\mathbf{X}; t_m^b\right) g\left(\mathbf{X}; t_1^b\right)\right] & E_{P_0}\left[g\left(\mathbf{X}; t_m^b\right) g\left(\mathbf{X}; t_2^b\right)\right] & \cdots & E_{P_0}\left[g\left(\mathbf{X}; t_m^b\right)\right]^2 \end{pmatrix}.$$

If $\Delta_c\left(P_0\right) \neq \emptyset$, then by the Lipschitz continuity of the moment functions, it is sufficient to consider

[5] A countable number of binding moments on a compact interval means that the moment functions have to oscillate wildly, and hence, violate Lipschitz continuity.

the covariance matrix of the variables $\left\{g\left(\mathbf{X};t\right),\, t\in\dot{\Delta\left(P_0\right)}\right\}$, where

$$\dot{\Delta\left(P_0\right)}=\Delta_d\left(P_0\right)\cup\left(\Delta_c\left(P_0\right)\cap\mathbb{Q}\right) \tag{2.6}$$

and $\mathbb{Q}$ is the set of rational numbers. As $\dot{\Delta\left(P_0\right)}=\left\{t_1^b,t_2^b,\ldots\right\}$ is a countable set, it gives rise to the infinite covariance matrix $\Sigma_\infty\left(P_0\right)$.

We denote the cardinality of $\dot{\Delta\left(P_0\right)}$ by $w=\left|\dot{\Delta\left(P_0\right)}\right|$. If $\dot{\Delta\left(P_0\right)}$ is countable, then we set $w=\infty$. Additionally, define the vector spaces

$$l_w^\infty=\left\{a=\left(a_1,a_2,\ldots,a_w\right)\in\mathbb{R}^w:\sup_j\left|a_j\right|<+\infty\right\}\quad\text{and} \tag{2.7}$$

$$l_w^1=\left\{a=\left(a_1,a_2,\ldots,a_w\right)\in\mathbb{R}^w:\sum_{j=1}^{w}\left|a_j\right|<+\infty\right\} \tag{2.8}$$

with respective norms $\|a\|_{l_w^1}=\sum_{j=1}^w\left|a_j\right|$, and $\|a\|_{l_w^\infty}=\sup_j\left|a_j\right|$.

Let P denote a generic value of P_0. Next we define the null parameter space for P_0.

Definition 2.1. *Let $\mathcal{M}$ be some collection of P that satisfies the following conditions for a given constant $c>0$.*

(i) Dependence: neither of the random variables X^A and X^B is a deterministic transformation of the other,

(ii) Sampling: $\left\{\mathbf{X}_i\right\}_{i=1}^n$ is a random sample from P,

(iii) Injectivity: $\theta'\Sigma_w\left(P\right)\theta\geq c\quad\forall\theta\in l_w^1$ such that $\|\theta\|_{l_w^1}=1$,

(iv) Surjectivity: the covariance operator, $\Sigma_w\left(P\right):l_w^1\to l_w^\infty$, is surjective,

(v) The set of moment functions $\left\{\mathbf{x}\mapsto g(\mathbf{x};t),t\in[\underline{t},\overline{t}]\right\}$ satisfies Assumption 2.1,

(vi) $E_P\left[g\left(\mathbf{X};t\right)\right]\leq 0\quad\forall t\in[\underline{t},\overline{t}]$.

Condition (i) of Definition 2.1 allows for an arbitrary dependence structure between the marginal random variables except that one cannot be a deterministic transformation of the other. This flexibility in the dependence configuration allows for structures that could occur when dealing with situations involving the comparison of pre- and post-tax income distributions, or the distributions of separate incomes of married couples, for instance.

The large sample behavior of the ELR statistic depends on the properties of the covariance matrix $\Sigma_w(P_0)$. Since it is possible under H_0 for $w = \infty$ i.e. an infinite covariance matrix, these properties are best described as conditions on the operator $\Sigma_w(P_0) : l_w^1 \to l_w^\infty$. Condition (iii) of Definition 2.1 implies the null space of $\Sigma_w(P_0)$ contains only the zero vector $\mathbf{0}_w \in l_w^1$. That is, $\text{Ker}(\Sigma_w(P_0)) = \{\theta \in l_w^1 : \Sigma_w(P_0)\theta = \mathbf{0}_w\} = \{\mathbf{0}_w\}$. Therefore, the inverse operator $\Sigma_w^{-1}(P_0) : \text{Range}(\Sigma_w(P_0)) \to l_w^1$ exists, and $\text{Range}(\Sigma_w(P_0)) = l_w^\infty$ for finite w. Condition (iv) of Definition 2.1 ensures that $\text{Range}(\Sigma_\infty(P_0)) = l_\infty^\infty$ holds as well. Since $\sup_{i,j}\Sigma_{w,i,j}(P_0) \le 1$ is a consequence part (ii) of Assumption 2.1, Conditions (iii) and (iv) of Definition 2.1 imply that $\Sigma_w(P_0)$ and $\Sigma_w^{-1}(P_0)$ are bounded in the operator norms

$$\|\Sigma_w(P_0)\| = \sup_{\left\{\theta \in l_w^1; \|\theta\|_{l_w^1} = 1\right\}} \|\Sigma_w(P_0)\theta\|_{l_w^\infty} \quad \text{and} \tag{2.9}$$

$$\left\|\Sigma_w^{-1}(P_0)\right\| = \sup_{\left\{\theta \in l_w^\infty; \|\theta\|_{l_w^\infty} = 1\right\}} \left\|\Sigma_w^{-1}(P_0)\theta\right\|_{l_w^1}. \tag{2.10}$$

That is, $\|\Sigma_w(P_0)\| \le \sup_{i,j}\Sigma_{w,i,j}(P_0) \le 1$, and the boundedness of $\Sigma_w^{-1}(P_0)$ follows from the Bounded Inverse Theorem[6]. As can be seen in (2.9) and (2.10), the operators norms depend on the vector spaces l_w^1 and l_w^∞; however, for ease of exposition, we suppress the dependence of the operator norms on these spaces.

Now we introduce further notation. The n-fold product probability measures, P^n, defined on the product measurable space $(\mathcal{X}^n, \mathcal{A}^n)$ is used to compute the probabilities of events in $\mathcal{A}^n$. To keep the notation simple when describing the probability of events $A_n \in \mathcal{A}^n$, we adopt the convention that $\text{Prob}_P[A_n]$ is the probability of the event A_n with respect to the joint distribution of the bi-

[6]See Theorem 4.12-2 of Kreyszig (1989) for a formal statement.

variate random sample, P^n. Suppose $\{U_i\}_{i=1}^{+\infty}$ is an i.i.d sequence of random variables with respect to the product measure P^∞ for every $P \in \mathcal{M}$. Then we say that $\{U_i\}_{i=1}^{+\infty} = O_P(1)$ uniformly over $\mathcal{M}$ if for any $\epsilon > 0$, there exists $B > 0$ and N_0 such that $\sup_{P\in\mathcal{M}} \text{Prob}_P[|U_n| > B] < \epsilon$ for all $n > N_0$. Similarly, we say that $\{U_i\}_{i=1}^{+\infty} = o_P(1)$ uniformly over $\mathcal{M}$ if for any $\epsilon > 0$, $\sup_{P\in\mathcal{M}} \text{Prob}_P[|U_n| > \epsilon] \to 0$ as $n \to +\infty$.

3 Empirical Likelihood

This section introduces (i) the unrestricted and restricted empirical likelihood estimators of P_0, (ii) the uniform asymptotic distribution of the ELR test statistic, and (iii) a computational algorithm for approximating the ELR test statistic.

3.1 A Concave Semi-Infinite Programming Estimator

Because of the continuity of the moment functions, it is enough to impose the moment conditions on $[\underline{t}, \overline{t}] \cap \mathbb{Q}$, where $\mathbb{Q}$ is the set of rational numbers[7]. The restricted empirical log-likelihood problem is

$$l^r = \max_{p_1,\ldots,p_n} \left\{ \sum_{i=1}^n \log(p_i);\ p_i \geq 0,\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n p_i g(\mathbf{X}_i; t) \leq 0 \quad \forall t \in [\underline{t}, \overline{t}] \cap \mathbb{Q}, \right\}, \quad (3.1)$$

where p_i denotes the probability mass placed on $\mathbf{X}_i$ by a discrete distribution supported on $\{\mathbf{X}_i\}_{i=1}^n$. The optimization problem (3.1) is a (random) concave semi-infinite program since there is a finite number of choice variables, and an infinite number of constraints. It has a *unique* solution for realizations of $\{\mathbf{X}_i\}_{i=1}^n$ that yield a nonempty constraint set[8]; for $P_0 \in \mathcal{M}$, this constraint set is nonempty for a large enough sample size[9]. The unrestricted empirical log-likelihood problem, l^{ur}, is similar to l^r except that the moment inequality conditions $\sum_{i=1}^n p_i g(\mathbf{X}_i; t) \leq 0 \quad \forall t \in [\underline{t}, \overline{t}] \cap \mathbb{Q}$,

[7] This is because $[\underline{t}, \overline{t}] \cap \mathbb{Q}$ is dense in $[\underline{t}, \overline{t}]$.

[8] This result follows by a standard application of Weierstrass' Theorem to the problem (3.1). It is stated as Proposition C.1 in Appendix C for ease of exposition.

[9] This large sample property holds uniformly in $\mathcal{M}$, and it is stated as Lemma C.1 in Appendix C.

are not imposed. The solution in this case is simply $\hat{p}_i = 1/n$, and then $l^{ur} = -n\log(n)$. The ELR statistic is defined as

$$\tilde{\mathcal{E}}_n = 2\left(l^{ur} - l^r\right). \tag{3.2}$$

Hence, large values of this statistic suggest the restriction is not supported by the data.

Let $\tilde{\mathbf{p}}$ denote the solution of (3.1). Its characterization it in terms of Lagrange multipliers requires formulating a dual of the problem (3.1). To do this, we need to specify a pair of vector spaces, one which serves as the space into which the moment functions are embedded, and the second its dual space[10]. By Assumption 2.1, we can embed $\left\{\mathbf{x} \mapsto g(\mathbf{x};t),\, t \in [\underline{t}, \overline{t}] \cap \mathbb{Q}\right\}$ into the sequence space l_∞^∞. Its dual space is $ba\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)$: the Banach space of bounded finitely additive scalar-valued signed measures on the power set of $[\underline{t}, \overline{t}] \cap \mathbb{Q}$, endowed with the total variational norm[11]. An element $\nu \in ba\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)$ is a bounded linear functional on $\mathbb{R}^\infty$ of the form $\sum_{j=1}^{\infty} \nu_j w_j$, where $w \in \mathbb{R}^\infty$ and $\nu_j = \nu\left(\{t_j\}\right)$ for each $t_j \in [\underline{t}, \overline{t}] \cap \mathbb{Q}$

The Lagrangian associated with problem (3.1) is

$$L(\mathbf{p}, \mu, \lambda) = \sum_{i=1}^{n} \log\left(p_i\right) + \left(1 - \sum_{i=1}^{n} p_i\right)\lambda - n \sum_{j=1}^{+\infty} \sum_{i=1}^{n} p_i g\left(\mathbf{X}_i; t_j\right)\mu_j, \tag{3.3}$$

where $(\mathbf{p}, \mu, \lambda) \in \mathbb{R}_+^n \times ba\left(2^{\mathbb{N}}\right) \times \mathbb{R}$. As in the case with finitely many constraints, a saddle point of the Lagrangian yields the desired characterization of the probabilities in terms of the Lagrange multipliers. Such a characterization follows from the Karush-Kuhn-Tucker (KKT) conditions under a constraint qualification. Because there is an infinite number of inequality constraints, the Strong Slater Condition, introduced by Mordukhovich and Nghia (2013), is the appropriate con-

[10]Given a vector space Y, its dual space is by definition the set of all bounded linear functionals on Y.

[11]The total variation norm of a signed measure is defined through its Hahn-Jordan decomposition; see Theorem 3.4 of Folland (1999).

straint qualification. In the setting of the paper, it is an event in $\mathcal{A}^n$ given by

$$\mathcal{S}_n = \left\{ \exists\, p_i > 0 \; i = 1, \ldots, n \quad \text{such that} \quad \sum_{i=1}^{n} p_i = 1, \sup_{t \in [\underline{t}, \overline{t}] \cap \mathbb{Q}} \sum_{i=1}^{n} p_i g\left(\mathbf{X}_i; t\right) < 0 \right\}. \tag{3.4}$$

Denote the active at $\tilde{\mathbf{p}}$ constraints by: $\Delta\left(\tilde{\mathbf{p}}\right) = \left\{ t \in [\underline{t}, \overline{t}] \cap \mathbb{Q} : \sum_{i=1}^{n} \tilde{p}_i g\left(\mathbf{X}_i; t\right) = 0 \right\}$, and let $ba\left(2^{[\underline{t}, \overline{t}] \cap \mathbb{Q}}\right)_+ = \left\{ \mu \in ba\left(2^{[\underline{t}, \overline{t}] \cap \mathbb{Q}}\right) : \mu_j \geq 0 \, \forall j \in \mathbb{N} \right\}$. Then, on the event $\mathcal{S}_n$, for some $\tilde{\lambda} \in \mathbb{R}$ and $\tilde{\mu} \in ba\left(2^{[\underline{t}, \overline{t}] \cap \mathbb{Q}}\right)_+$, $\tilde{\mathbf{p}}$ satisfies the following KKT conditions

$$\tilde{p}_i > 0, \quad \frac{1}{\tilde{p}_i} - \tilde{\lambda} - n \sum_{j=1}^{+\infty} g\left(\mathbf{X}_i; t_j\right) \tilde{\mu}_j = 0, \quad i = 1 \ldots, n; \tag{3.5}$$

$$\sum_{i=1}^{n} \tilde{p}_i g\left(\mathbf{X}_i; t\right) \leq 0 \quad \forall t \in [\underline{t}, \overline{t}] \cap \mathbb{Q}; \quad \text{supp}(\tilde{\mu}) \subset \Delta\left(\tilde{\mathbf{p}}\right), \quad \sum_{i=1}^{n} \tilde{p}_i = 1, \tag{3.6}$$

where supp$(\tilde{\mu})$ is the support of $\tilde{\mu}$.

Multiplying both sides of the i-th equation in (3.5) by $\tilde{p}_i$ and summing over i yields $\tilde{\lambda} = n$ by complementary slackness and the constraint $\sum_{i=1}^{n} \tilde{p}_i = 1$; hence, the probabilities are given by

$$\tilde{p}_i = \frac{1}{n} \left(1 + \sum_{j=1}^{+\infty} \tilde{\mu}_j g\left(\mathbf{X}_i; t_j\right) \right)^{-1}, \quad i = 1, \ldots, n. \tag{3.7}$$

This characterization of the $\tilde{p}_i$ in (3.7) only occurs on the event $\mathcal{S}_n$. The next result shows the likelihood of this event tends to unity with uniformity, under the null hypothesis.

Proposition 3.1. *Suppose* $P_0 \in \mathcal{M}$. *Then,* $\sup_{P \in \mathcal{M}}$ *Prob*$_P\left[\mathcal{S}_n\right] \to 1$ *as* $n \to +\infty$.

Proof. See Appendix C.2. □

Consequently, under the null and for large enough n, upon substituting in the probabilities (3.7) into the expression for the ELR statistic (3.2) and re-arranging, results in the following expression for the test statistic

$$\tilde{\mathcal{E}}_n = \max_{\mu \in ba\left(2^{[\underline{t}, \overline{t}] \cap \mathbb{Q}}\right)_+} 2 \sum_{i=1}^{n} \log \left(1 + \sum_{j=1}^{+\infty} g\left(\mathbf{X}_i; t\right) \mu_j \right). \tag{3.8}$$

Therefore, the asymptotic behavior of $\tilde{\mathcal{E}}_n$ under the null depends on the asymptotic behavior of $\tilde{\mu}$. The restriction on μ in (3.8) significantly affects the asymptotic null distribution of $\tilde{\mathcal{E}}_n$, and results in a non-pitoval limit distribution, as the next section shows.

3.2 Uniform Asymptotic Null Distribution

This section presents the uniform asymptotic distribution theory for the ELR statistic (3.2) when $P_0 \in \mathcal{M}$. This result relies on the uniform convergence of the random function

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X};t\right) - E_{P_0}\left[g\left(\mathbf{X};t\right)\right]\right)\ t \in [\underline{t},\overline{t}], \tag{3.9}$$

which follows from Assumption 2.1, as it implies the set moment functions is uniformly Donsker and pre-Gaussian with respect to the probability measures in $\mathcal{M}$; we formalize this result as Lemma B.1 in the Appendix. The random function (3.9) is uniformly weakly convergent to a zero-mean Gaussian process, $G(t)\, t \in [\underline{t},\overline{t}]$, with covariance kernel

$$E_{P_0}\left[g\left(\mathbf{X};u\right)g\left(\mathbf{X};v\right)\right] - E_{P_0}\left[g\left(\mathbf{X};u\right)\right]E_{P_0}\left[g\left(\mathbf{X};u\right)\right] \quad (u,v) \in [\underline{t},\overline{t}] \times [\underline{t},\overline{t}]. \tag{3.10}$$

Furthermore, this uniform weak convergence implies the random function, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g\left(\mathbf{X};t\right)\ t \in \Delta\left(P_0\right)$, is uniformly weakly convergent to a zero-mean Gaussian process, $G(t)\, t \in \Delta\left(P_0\right)$, with the covariance kernel $E_{P_0}\left[g\left(\mathbf{X};u\right)g\left(\mathbf{X};v\right)\right],\ (u,v) \in \Delta\left(P_0\right) \times \Delta\left(P_0\right)$. If $\Delta\left(P_0\right) = \{t_1,\ldots,t_w\}$ and $w \in \mathbb{Z}_+$, then the limiting Gaussian process is a multivariate normal random vector with dimension w given by $\mathbf{G}_w \sim \text{MVN}\left(\mathbf{0}_w, \Sigma_w\left(P_0\right)\right)$. If $\Delta\left(P_0\right) = \Delta_d\left(P_0\right) \cup \Delta_c\left(P_0\right)$, then the limiting Gaussian process is the extension of the discrete Gaussian process $\mathbf{G}_\infty \sim \text{MVN}\left(\mathbf{0}_\infty, \Sigma_\infty\left(P_0\right)\right)$ on $\Delta\left(\dot{P}_0\right)$.

To develop the uniform asymptotic distribution of $\tilde{\mathcal{E}}_n$, it is necessary to restrict $\mathcal{M}$ to certain submodels as follows.

Definition 3.1. *Given* $e_0 \in \mathbb{R}_+$, *let*

$$\mathcal{M}\left(e_0\right)=\left\{P \in \mathcal{M}: \Delta\left(P\right)=\emptyset\right\} \cup\left\{P \in \mathcal{M}: \Delta\left(P\right) \neq \emptyset \text{ and } \left\|\Sigma_w^{-1}\left(P\right)\right\| \leq e_0\right\}, \tag{3.11}$$

where $\|\cdot\|$ *is the operator norm (2.9).*

Let $\rightsquigarrow$ denote weak convergence, and let $l_{w,-}^{\infty}=\left\{\theta \in l_w^{\infty}: \theta_j \leq 0\ \forall j\right\}$. The following theorem provides the uniform asymptotic distribution of $\tilde{\mathcal{E}}_n$ under the null.

Theorem 3.1. *For every* $e_0 \in \mathbb{R}_+$, *we have*

$$\tilde{\mathcal{E}}_n \rightsquigarrow \begin{cases} 0, & \textit{if } w=0, \\ \min_{\mathbf{U} \in l_{w,-}^{\infty}}\left(\mathbf{G}_w-\mathbf{U}\right)^{\prime} \Sigma_w^{-1}\left(P_0\right)\left(\mathbf{G}_w-\mathbf{U}\right), & \textit{if } w \neq 0, \end{cases}$$

uniformly in $\mathcal{M}\left(e_0\right)$.

Proof. See Appendix C.3. □

The form of the contact set has a significant discontinuous effect on the shape of the ELR's asymptotic null distribution. Theorem 3.1 shows that the limit distribution of $\tilde{\mathcal{E}}_n$ when no inequality constraint binds is degenerate at zero, since $\tilde{\mathcal{E}}_n \xrightarrow{P} 0$ in this case. If only a finite number of constraints bind, then the ELR statistic converges in distribution under the null to the familiar Gaussian QLR statistic which has the chi-bar-square distribution. The last case is when the set of binding moments has isolated and connected parts, or only connected parts. The form of the ELR statistic in this case is a generalization of the Gaussian QLR statistic.

A pre-requisite for using the result in Theorem 3.1 is the ability to compute $\tilde{\mathcal{E}}_n$. Its computation is infeasible in practice since it is impossible to impose an infinite number of inequality constraints in numerical optimization routines. However, it is possible to approximate $\tilde{\mathcal{E}}_n$ using standard methods of approximation in the SIP literature, which is discussed in the next section.

3.3 Computational Algorithm: An Exchange Method

This section introduces a computational algorithm that approximates the solution of the SIP problem (3.1). Because the optimization problem associated with this estimator is a SIP, we adopt a numerical approach to SIP to compute it. Nowdays the numerical approach to SIP has become an active research area; for a review on SIP algorithms, see Hettich and Kortanek (1993) and Reemsten and Gorner (1998).

An important first point that we emphasize is that from a numerical point of view, SIP is more difficult than finite programming. The difficulty arises with the feasibility test of a candidate solution to the SIP, because checking the feasibility in this case is obviously equivalent to solving the global maximization problem: $\max_{t\in[\underline{t},\overline{t}]} \sum_{i=1}^{n} \tilde{p}_i g\left(\mathbf{X}_i, t\right)$, and to check whether for a global solution to this problem that the continuum of constraints do indeed hold. The paper uses an "exchange method" to compute an approximation to $\tilde{\mathbf{p}}$. It is one of the main algorithmic approaches to solving SIP problems. This computational algorithm can be seen as a compromise between pure discretization methods and methods based on local reduction.

The discretization method requires that we choose finite subsets $\{t_j\}_{j=1}^{N} = \mathcal{T}_N \subset [\underline{t},\overline{t}] \cap \mathbb{Q}$, and instead of solving the SIP problem (3.1), we solve the finite program

$$\max_{p_1,\ldots,p_n} \left\{ \sum_{i=1}^{n} \log\left(p_i\right);\ p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i g\left(\mathbf{X}_i; t_j\right) \leq 0 \quad j = 1, \ldots, N \right\}. \tag{3.12}$$

The sequence of sets $\{\mathcal{T}_N\}_N$ are such that the Hausdorff distance between $\mathcal{T}_N$ and $[\underline{t},\overline{t}] \cap \mathbb{Q}$ tends to zero as $N \to +\infty$. That is, dist $\left(\mathcal{T}_N, [\underline{t},\overline{t}] \cap \mathbb{Q}\right) \to 0$ as $N \to +\infty$, where

$$\text{dist}\left(\mathcal{T}_N, [\underline{t},\overline{t}] \cap \mathbb{Q}\right) = \sup_{t\in[\underline{t},\overline{t}]\cap\mathbb{Q}} \min_{\breve{t}\in\mathcal{T}_N} |\breve{t} - t|. \tag{3.13}$$

The distance dist $\left(\mathcal{T}_N, [\underline{t},\overline{t}] \cap \mathbb{Q}\right)$ is a measure for the mesh-size of the discretization. So that when the problem (3.1) is indeed discretizable, its (unique) solution, provided it exists, is the point of accumulation of the corresponding sequence of solutions for the problems (3.12); see Sections 3 and 6 of Shapiro (2009) for sufficient conditions on the discretizability of general SIP problems.

On the other hand, methods based on local reduction require the determination of all the local maxima of the problem

$$\max_{t\in[\underline{t},\overline{t}]}\sum_{i=1}^{n}p_i g\left(\mathbf{X}_i,t\right), \tag{3.14}$$

as they depend on the probabilities. This approach is computationally expensive and can be infeasible in practice for data sets of a moderate size, because the number of independent variables in these solutions grows one-to-one with the sample size.

Conceptually, the exchange method at the N-th step has a given grid $\mathcal{T}_N$ and a fixed small value $\alpha > 0$. Then, one proceeds as follows:

1. Compute a solution $\tilde{\mathbf{p}}_N$ of the discretized problem (3.12).

2. Compute the local maxima $t_{j,N}, j = 1,\ldots,k$ of the problem (3.14) when $\mathbf{p} = \tilde{\mathbf{p}}_N$, such that one of them, say $t_{1,N}$ is a global solution; that is,

$$\sum_{i=1}^{n}\tilde{p}_{i,N}g\left(\mathbf{X}_i,t_{1,N}\right)=\max_{t\in[\underline{t},\overline{t}]}\sum_{i=1}^{n}\tilde{p}_{i,N}g\left(\mathbf{X}_i,t\right).$$

3. Stop, if $\sum_{i=1}^{n}\tilde{p}_{i,N}g\left(\mathbf{X}_i,t_{1,N}\right) \leq \alpha$, with an approximate solution $\tilde{\mathbf{p}}_N$. Otherwise, update $\mathcal{T}_{N+1} = \mathcal{T}_N \bigcup \{t_{j,N}, j = 1,\ldots,k\}$.

Naturally, the numerical accuracy of this method depends on the number of grid points, N, and on the tolerance number α. In practice, both will depend on the sample size. Individually, N and α introduce a bias variance trade-off in the computation of $\sum_{i=1}^{n}\tilde{p}_i g\left(\mathbf{X}_i,t\right)$ for each $t \in [\underline{t},\overline{t}]$. Large values of N and small values of α increase the bias but reduce variance, whereas small values of N and large values of α decrease the bias but increase the variance.

Denote the solution of the exchange algorithm by $\acute{p}_1,\ldots,\acute{p}_n$, then the ELR statistic based on it is defined as

$$\acute{\mathcal{E}}_n = 2\left\{-n\log(n) - \sum_{i=1}^{n} \log\left(\acute{p}_i\right)\right\}. \tag{3.15}$$

Theorem 3.2. *Suppose* $P_0 \in \mathcal{M}$. *If* $\min\left\{N(n)^{-1}, \alpha(n)\right\} \to 0$ *as* $n \to +\infty$, *then*

1. $\acute{\mathcal{E}}_n = \tilde{\mathcal{E}}_n + o_P(1)$ *uniformly in* $\mathcal{M}$.
2. $\acute{\mathcal{E}}_n - \tilde{\mathcal{E}}_n = O_P\left(n^{-1/2}\min\left\{\frac{L}{N(n)}, \alpha(n)\right\}\right)$ *uniformly in* $\mathcal{M}$,

where L *is the Lipschitz constant presented in Assumption 2.1.*

Proof. See Appendix C.4. ☐

Theorem 3.2 shows that the ELR statistic arising from the exchange algorithm is uniformly asymptotically equivalent to the ELR statistic (3.2). Consequently, $\acute{\mathcal{E}}_n$ can be used to test the null hypothesis $P_0 \in \mathcal{M}$ in practice. A remarkable point regarding Theorem 3.2 is that the validity of the uniform asymptotic equivalence does not require any restrictions on the rates of $N(n)$ and $\alpha(n)$.

In general, the computation of fixed asymptotic critical values is infeasible because the asymptotic null distribution in Theorem 3.1 depends discontinuously on P_0 through the contact set (2.5). This feature of the ELR statistic motivates the use of the bootstrap, which is discussed in the next section.

4 Bootstrap Test Procedure

This section introduces the bootstrap ELR test for the null hypothesis $P_0 \in \mathcal{M}$. The testing procedure extends the approach of Canay (2010) to the setting of the paper.

The bootstrap DGP is the set of probabilities on the data points that is the solution of a modified version of the exchange algorithm. The modification replaces the finite program (3.12) in the exchange algorithm Section 3.3 with the following optimization problem:

$$\max_{p_1,\ldots,p_n}\left\{\sum_{i=1}^{n}\log\left(p_i\right);\, p_i\geq 0,\, \sum_{i=1}^{n}p_i=1,\, \sum_{i=1}^{n}p_i g\left(\mathbf{X}_i;t_j\right)\leq -\eta_n(t_j) \quad j=1,\ldots,N\right\}, \quad (4.1)$$

where the $\eta_n(t_j)\geq 0$ are (possibly random) numbers that satisfy the following properties

$$\text{Prob}_{P_0}\left[\lim_{n\to+\infty}\eta_n(t_j)=0 \text{ and } \liminf_{n\to+\infty}\eta_n(t_j)\left(n/\left(2\log\log n\right)\right)^{1/2}\geq\sigma_j\right]=1 \quad \text{and} \quad (4.2)$$

$$\sup_{t\in[\underline{t},\overline{t}]}\eta_n(t)\xrightarrow{P}0 \quad \text{uniformly in} \quad \mathcal{M}, \quad \text{where} \quad (4.3)$$

$$\sigma_j^2=E_{P_0}\left[g\left(\mathbf{X};t_j\right)\right]^2-\left(E_{P_0}\left[g\left(\mathbf{X};t_j\right)\right]\right)^2 \quad j=1,\ldots,N(n). \quad (4.4)$$

The sequence $\eta_n(t_j)$ provides a rule to determine whether the t_j-th moment is binding or slack. It is similar to the sequences in Andrews and Soares (2010), LSW, and Canay (2010).

Denote the bootstrap DGP by $(\overline{p}_1,\ldots,\overline{p}_n)$, and let $\{\mathbf{X}_i^\star\}_{i=1}^n$ be a random sample from it. Furthermore, denote by $(\overline{p}_1^\star,\ldots,\overline{p}_n^\star)$ the solution of the modified exchange method algorithm in which $\{\mathbf{X}_i^\star\}_{i=1}^n$ replaces the data. The bootstrap ELR statistic is defined as

$$\overline{\mathcal{E}}_n^\star=2\left\{-n\log(n)-\sum_{i=1}^{n}\log\left(\overline{p}_i^\star\right)\right\}. \quad (4.5)$$

Letting B_n be the number of bootstrap replications, the approximate bootstrap p-value is defined as

$$\Upsilon_{B_n}=\frac{1}{B_n}\sum_{j=1}^{B_n}1\left[\overline{\mathcal{E}}_{n,j}^\star\geq\acute{\mathcal{E}}_n\right], \quad (4.6)$$

where $\acute{\mathcal{E}}_n$ is given by (3.15). The bootstrap test rejects H_0 if $\Upsilon_{B_n}\leq\beta$, where $\beta\in(0,1/2)$ is a given nominal level.

4.1 Uniform Asymptotic Validity Results

The following theorem shows the asymptotic null distribution of the bootstrapped ELR statistic (4.5) is exactly the same asymptotic null distribution of the ELR statistic (3.2) provided in Theorem 3.1.

Theorem 4.1. *For each* n, *let* $\mathcal{A}_n$ *denote the sigma algebra generated by* $\{\mathbf{X}_i\}_{i=1}^n$. *For every* $e_0 \in \mathbb{R}_+$, *the modified bootstrap ELR statistic defined in (4.5) satisfies*

$$\overline{\mathcal{E}}_n^\star \rightsquigarrow \begin{cases} 0, & \textit{if } w = 0, \\ \min_{\mathbf{U} \in l_{w,-}^\infty} \left(\mathbf{G}_w - \mathbf{U}\right)' \Sigma_w^{-1}\left(P_0\right)\left(\mathbf{G}_w - \mathbf{U}\right), & \textit{if } w \neq 0 \end{cases}$$

conditional on $\mathcal{A}_n$ *in* P_0 *uniformly in* $\mathcal{M}(e_0)$.

Proof. See Appendix D.1. □

The result of Theorem 4.1 is uniform in $\mathcal{M}(e_0)$. Furthermore, it implies the following for the bootstrap ELR test.

Corollary 4.1. *Suppose the conditions of Theorem 4.1 hold. Additionally, let* Υ_{B_n} *be given by (4.6) and* $\beta \in (0, 1/2)$. *Then, for every* $e_0 \in \mathbb{R}_+$

$$\limsup_{n \to +\infty} \sup_{P \in \mathcal{M}(e_0)} \textit{Prob}_P\left[\Upsilon_{B_n} \leq \beta\right] \leq \beta. \tag{4.7}$$

Proof. See Appendix D.2. □

Corollary 4.1 shows the bootstrap ELR test has asymptotically correct size, uniformly in $\mathcal{M}(e_0)$.

4.2 Test Consistency

Next we consider the power of the proposed bootstrap test against all alternatives. The power of this test is shown to converge to 1 as $n \to +\infty$, which means the test is consistent against all alternatives.

Let $\Delta^+(P_0) = \left\{t \in [\underline{t}, \overline{t}] : E_{P_0}[g(\mathbf{X}, t)] \geq 0\right\}$. Under the alternative, the Lipschitz continuity of the moment functions implies $\Delta^+(P_0)$ has the cardinality of the continuum. As with the setup under H_0, $\Delta^+(P_0) = \Delta_d^+(P_0) \cup \Delta_c^+(P_0)$ where $\Delta_d^+(P_0)$ is the set of isolated points in $\Delta^+(P_0)$ and $\Delta_c^+(P_0)$ is the union of the connected parts of $\Delta^+(P_0)$. Furthermore, let $\dot{\Delta^+}(P_0) = \Delta_d^+(P_0) \cup (\Delta_c^+(P_0) \cap \mathbb{Q})$, and let $\Omega(P_0)$ denote the infinite matrix with typical element

$$\Omega_{t,t'}(P_0) = E_{P_0}[g(\mathbf{X}, t)\, g(\mathbf{X}, t')] \quad t, t' \in \dot{\Delta^+}(P_0).$$

The following assumption is needed to prove test consistency.

Assumption 4.1. *(i)* $\theta' \Omega(P_0) \theta > 0 \; \forall \theta \in l_\infty^1$ *with* $\|\theta\|_{l_\infty^1} = 1$. *(ii)* $\exists t' \in [\underline{t}, \overline{t}]$ *with* $E_{P_0}[g(\mathbf{X}, t')] > 0$, *and* $\{\mathbf{X_i}\}_{i=1}^n$ *is IID* P_0.

Next we have the result on test consistency.

Theorem 4.2. *Suppose* P_0 *satisfies Assumption 4.1, and that Assumption 2.1 holds. Additionally, let* $\min\{N(n)^{-1}, \alpha(n)\} \to 0$ *as* $n \to +\infty$. *Then, Prob*$_{P_0}[\Upsilon_{B_n} \leq \beta] \to 1$ *as* $n \to +\infty$, *where* $\beta \in (0, 1/2)$ *is a given nominal level and* Υ_{B_n} *is the approximate bootstrap p-value (4.6).*

Proof. See Appendix D.3. ☐

Theorem 4.2 shows that the proposed bootstrap procedure is consistent against all fixed alternatives.

5 Monte Carlo Simulations

The main purpose of this section is to evaluate the finite sample performance of the tools developed in previous sections. In each simulation experiment, the nominal level was fixed at 5%, $N(n) = n^{1/2}, \alpha(n) = n^{-1/2}$, and $\eta_n(t) = \hat{\sigma}_t \sqrt{\frac{2 \log n}{n}}$ where $\hat{\sigma}_t^2$ is the sample analogue estimator of σ_t^2.

Additionally, given the interval $[\underline{t}, \overline{t}]$, the grid was constructed as follows:

$$\mathcal{T}_{N(n)} = \left\{\underline{t} = t_1 < t_2 < \cdots < t_{N(n)} = \overline{t}\right\}, \text{ where } t_{i+1} = t_i + \frac{\left(\overline{t} - \underline{t}\right)}{N(n)}, \tag{5.1}$$

for $i = 1, \ldots, N(n) - 1$. We considered the following sample sizes $n = 256, 512, 1024, 2048$ with 1000 MC replications per experiment, and 199 bootstrap samples per MC replication. Finally, we also report the empirical rejection frequency of the bootstrap tests of LSW and AS; see Appendices H and I for the implementation of these tests in the setting of the current paper. Otherwise, the experiments were implemented using Matlab.

As described in Section 3.3, the exchange method uses finite programs which have known closed-forms for the first-order and second-order conditions. Supplying this derivative information to the Matlab optimization routine substantially speeds up the execution time of the exchange method algorithm, especially for the cases in which the sample size is large e.g. $n \geq 1024$. Overall, the Matlab code this paper uses executes the algorithm very rapidly on a desktop machine with 12 CPUs and 8 gigabytes of RAM.

5.1 Independent Uniform and Discontinuous DGPs

We considered the case of restricted second and third orders of SD[12] between the following statistically independent random variables. The CDF of X^A, $F_A(\cdot\,; a_0)$, depends on a parameter

[12] See Example 2 in Section 2.2 for the definition of restricted SD conditions.

$a_0 \in [0.5, 0.75]$, and it is given by

$$F_A(x; a_0) = \begin{cases} 1, & \text{if } x \geq 1 \\ x, & \text{if } 0.75 < x < 1 \\ mx + b, & \text{if } 0.5 < x \leq 0.75 \\ a_0, & \text{if } x = 0.5 \\ x, & \text{if } 0 \leq x < 0.5 \\ 0, & \text{if } x < 0, \end{cases}$$

where $m = 3 - 4a_0$ and $b = a_0 - m/2$. $F_A(\cdot\,; a_0)$ has a mass point at $x = 0.5$ with probability mass equal to $a_0 - 0.5$ when $a_0 > 0.5$. Finally, when $a_0 = 0.5$, $F_A(\cdot\,; a_0)$ is the CDF of $U[0, 1]$. We set $X^B \sim U[0, 1]$. Figure 1 depicts $F_A(x; a_0)$ for $a_0 \in \{0.75, 0.6\}$.

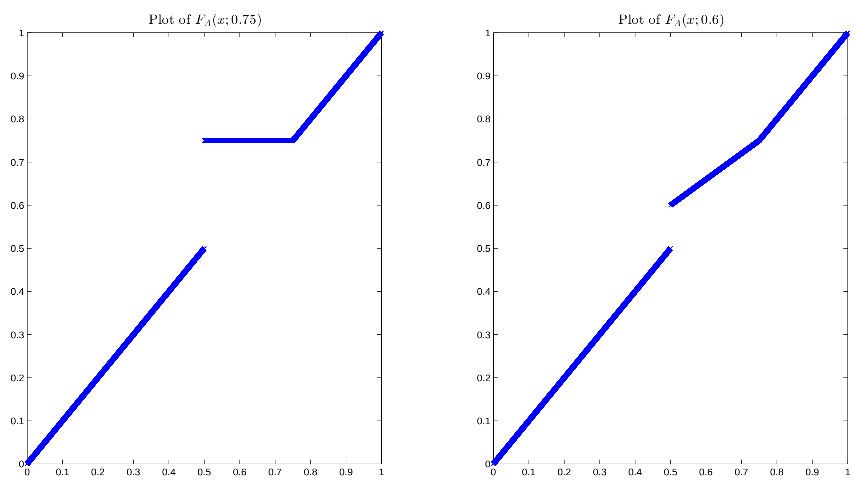

Figure 1: Plots of $F_A(\cdot\,; 0.75)$ (left) and $F_A(\cdot\,; 0.6)$ (right).

The motivation for using DGPs with mass points is that population income distributions can be discontinuous and/or continuous but non-differentiable at points in their supports. These properties of income distributions are salient as Zinde-Walsh (2008) shows in examples that they can be a result of policy and institutional effects.

5.1.1 Experiments Under H_0

Our choice of distributions for populations A and B described in the previous subsection are such that X^B dominates X^A at the second and third orders $\forall a_0 \in [0.5, 0.75]$, and therefore, this SD relationship holds for any choice of $[\underline{t}, \overline{t}]$. Figure 2 depicts the SD functions

$$t \mapsto E_P\left[\frac{\left(t - X^B\right)^{s-1}}{(s-1)!} 1\left[X^B \leq t\right] - \frac{\left(t - X^A\right)^{s-1}}{(s-1)!} 1\left[X^A \leq t\right]\right]$$

for $t \in [0.05, 0.95], a_0 = 0.75$, and $s = 2, 3$. The interesting feature of this set of DGPs is that the restricted second order SD function has a point of non-differentiability at $x = 0.5$ when $a_0 \in (0.5, 0.75]$. At $a_0 = 0.5$, $F_A = F_B$ which implies that the index set of binding population moments is $[\underline{t}, \overline{t}]$, where as it is equal to $[\underline{t}, 0.5) \cup [0.75, \overline{t}]$ for any $a_0 \in (0.5, 0.75]$. Therefore, as a_0 increases from 0.5 to 0.75, probability mass is progressively shifted away from the set $(0.5, 0.75]$ towards the set $\{0.5\}$.

In all of the experiments, we considered $a_0 \in \{0.5, 0.525, 0.55, \ldots, 0.75\}$, and set $[\underline{t}, \overline{t}] = [0.05, 0.95]$. The rejection frequencies for restricted second order SD (SSD) and third order SD (TSD) are presented in Figures 3 and 4 respectively.

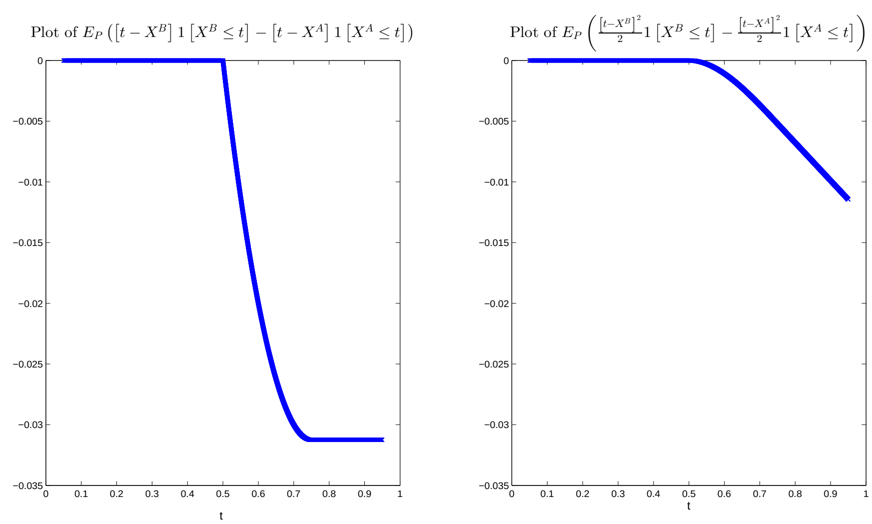

Figure 2: Plots of SSD function (left) and TSD function (right).

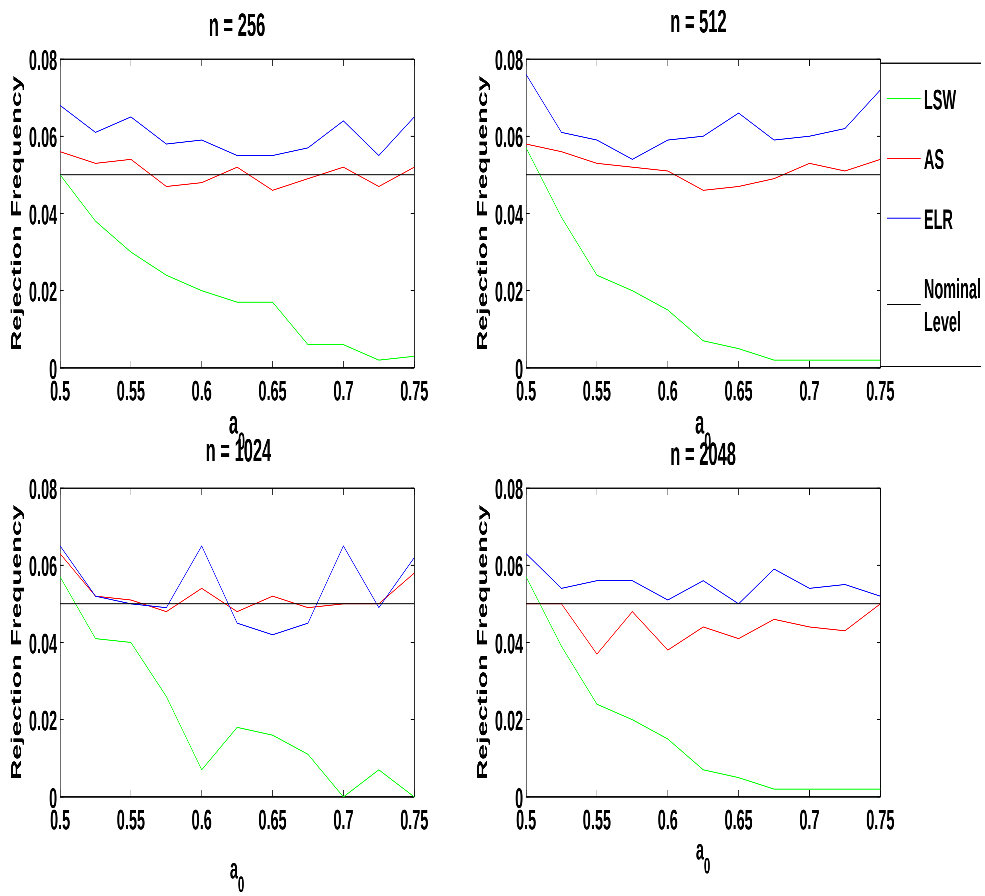

Figure 3: Plots of the rejection frequency for the test with the SSD function.

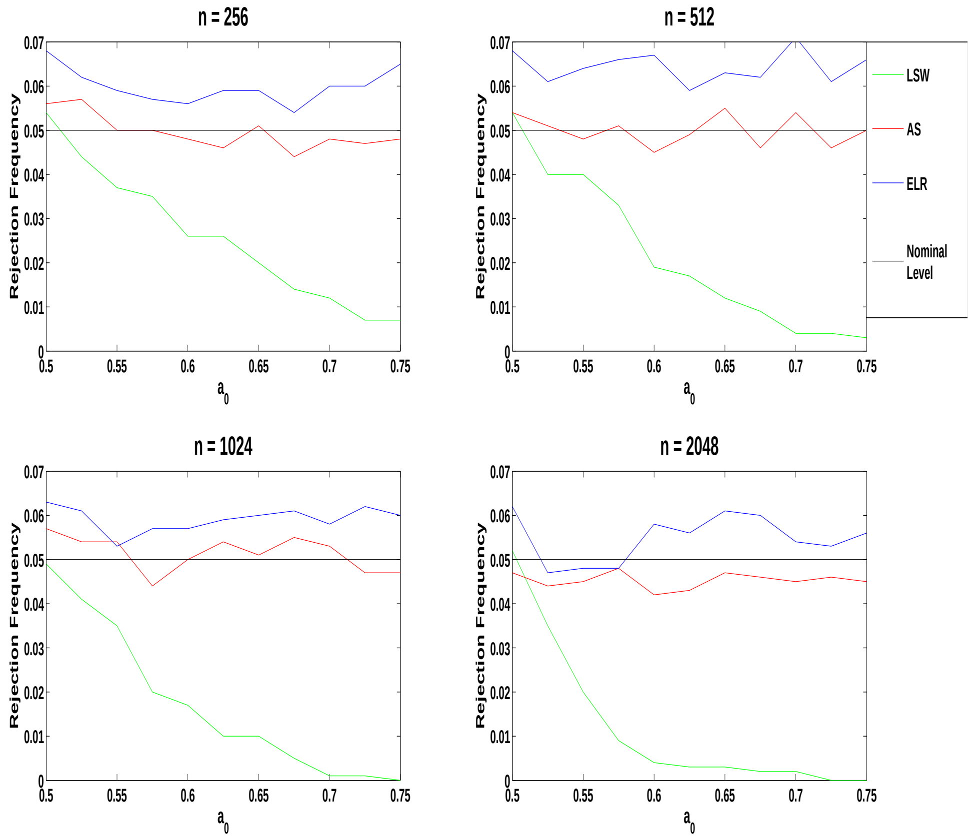

Figure 4: Plots of the rejection frequency for the test with the TSD function.

For $a_0 \neq 0.5$, the LSW test has null rejection probabilities that decrease to zero for the SSD and TSD DGPs. Note that these rejection probabilities decrease quite rapidly with the sample size for $a_0 \neq 0.5$. In contrast, the proposed test and that of AS have null rejection probabilities that are approximately within 0.02 of the 0.05 nominal level. For $a_0 = 0.5$, the DGP is in the least favorable case of the model of the null hypothesis, and therefore we expect the rejection probabilities for both tests to be close to the nominal level of 0.05.

The simulation results imply that the DGPs with $a_0 \neq 0.5$ do not belong to the subset of the boundary of H_0 in which the LSW test is asymptotically similar. The reason for the non-similarity of the LSW test at these DGPs is that the bootstrap p-value is too large when some of the moments in $E_{P_0}[g(\mathbf{X};t)]$ are negative and moderately small. Despite their test being consistent against all types of alternatives[13], a potential consequence of this non-similarity is that the LSW test might have relatively low power in finite samples against alternatives with some non-violated inequalities i.e. a population moment function $E_{P_0}\left[g\left(\mathbf{X};t\right)\right]$ (as a function of t) whose image contains positive and negative values with the latter being moderately small. Section 5.2 examines the finite sample power properties of the tests against such alternatives.

5.1.2 Experiments Under H_1

The DGPs in this section considers are as in the previous section, but with the roles of X^A and X^B reversed. In this case, $E_{P_0}\left[g\left(\mathbf{X};t\right)\right] \geq 0 \ \forall t \in [0.05, 0.95]$, and this population moment function equals zero on the interval $[0.05, 0.5]$ for all $a_0 \in (0.5, 0.75]$. Furthermore, as a_0 becomes larger, P_0 becomes farther from H_0.

The power curves of the tests are reported in Figure 5. The results show the tests behave similarly under these DGPs, as their this little difference between their empirical power functions.

[13] See Theorem 3 of LSW.

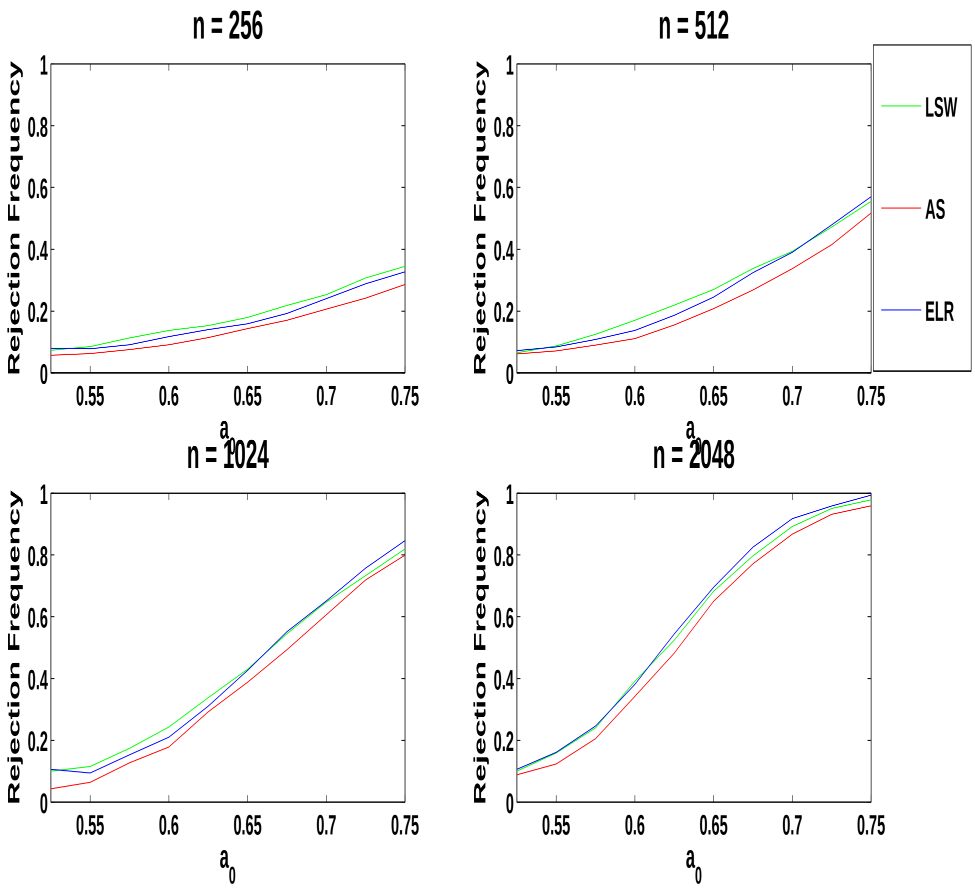

Figure 5: Plots of the rejection frequency for the test with X^A SSD X^B.

5.2 Experiments Under H_1 : Some Non-Violated Inequalities

Consider the following DGP from LSW. Set $X^A \sim U[0,1]$. Then, define

$$X^B = (U - a_0 b_1)\, 1\left[a_0 b_1 \leq U \leq x_0\right] + (U + a_0 b_2)\, 1\left[x_0 < U \leq 1 - a_0 b_2\right] \tag{5.2}$$

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 CDF of X^B is $F_B\left(x^B; a_0, b_1, b_2, x_0\right) = x^B + a_0 \delta\left(x^B\right),$

where

$$\delta\left(x^B\right)=\begin{cases}0, & \text{if } x^B\leq 0\\ b_1, & \text{if } 0<x^B\leq x_0-b_1\\ x_0, & \text{if } x_0-b_1<x^B\leq x_0+b_2\\ -b_2, & \text{if } x_0+b_2<x^B\leq 1\\ 0, & \text{if } x^B>1.\end{cases}$$

The scale a plays the role of the "distance" P_0 is from H_0. When a 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.

In the simulation experiments, we set

$$(b_1,b_2,x_0)=(0.1,0.5,0.15)\quad\text{and}\quad a_0\in\{0.05,0.1,0.15,0.2,\ldots,0.75\}.$$

These configurations correspond to alternative DGPs for which there are some non-violated inequalities in the restricted SSD function with $[0.05,0.95]$ as its domain of definition. This function is depicted in Figure 6 for $a_0=0.1$.

The power curves for the tests are reported in Figure 7. The proposed test dominates the AS and LSW tests since its power curve is greater than or equal to that of the other tests. For $a_0>0.5$, there is no difference between all three tests as the rejection probabilities are equal to unity. However, outside these configurations, the power of the AS and ELR tests is substantially higher than that of the LSW test for quite modest sample sizes. Furthermore, the power of the ELR test is at least as large as that of the AS test and strictly higher for DGP configurations closer the H_0. Finally, because the LSW test is consistent against all kinds of alternatives, we expect its finite sample power to improve with larger sample sizes.

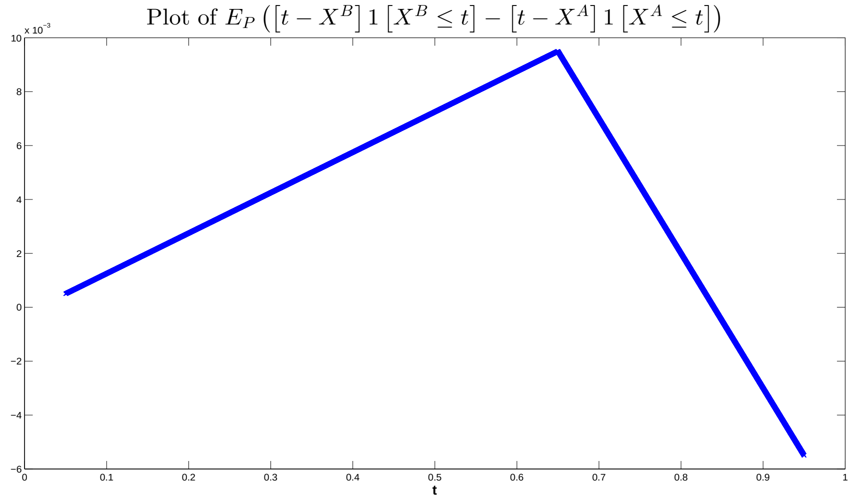

Figure 6: Plot of the restricted SSD function.

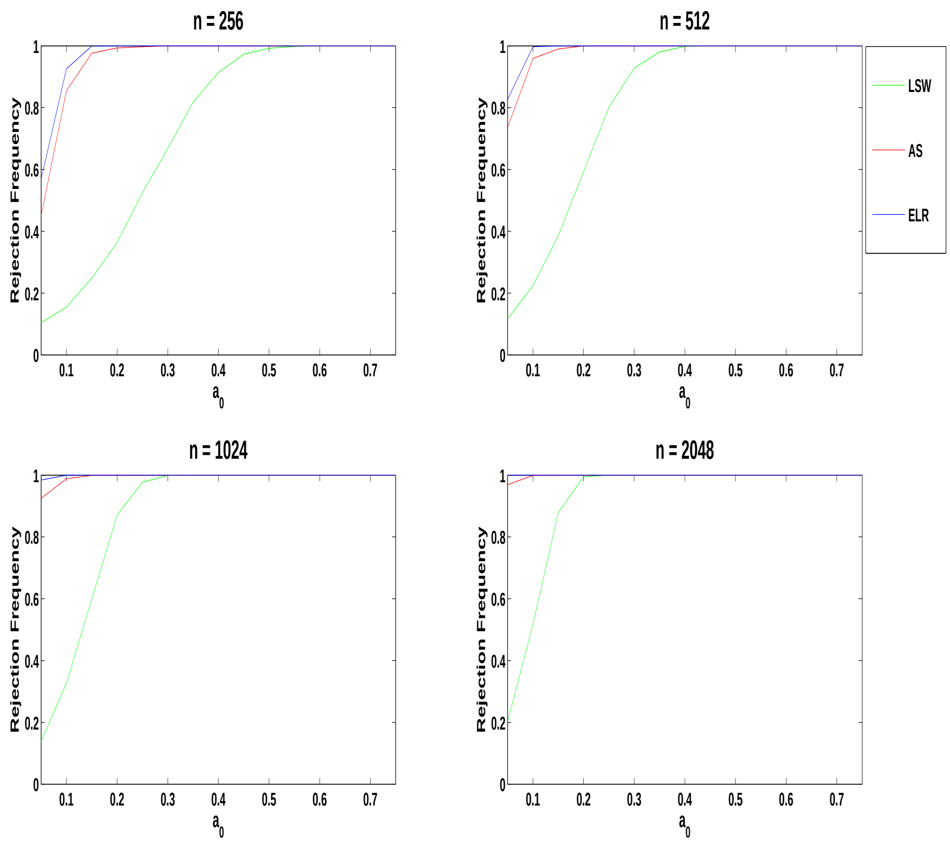

Figure 7: Plots of the rejection frequency for the test with the SSD function and SNVIs.

The simulation results in Figure 7 indicate that the rate of this improvement for their test is rather slow for DGPs which are close to H_0, e.g. $a_0 = 0.1$

6 Empirical Illustration

This section illustrates the proposed method in the context of an empirical example on policy evaluation. The data is from the Canadian Family Expenditures survey for the year 1986, which is used by Barrett and Donald (2003). This survey in a given year reports the incomes of households before and after a tax and transfer policy. We consider the comparison of the income distributions in the year 1986 before and after this policy using restricted SSD conditions. In Table 1 below we have supplied some basic descriptive statistics for these data.

Table 1: Descriptive Statistics

	Pre-Policy	Post-Policy
Sample Size	9,470	9,470
Mean	36,975	30,378
Std. Dev.	24,767	18,346
Median	32,658	27,337
Min	56.61	121.92
Max	206,670	180,390

The boxplots of the two income distributions are reported in Figure 8 along with a scatter plot of data. The scatter plot reveals a strong correlational dependence between the two distributions with a correlation coefficient of 0.982, which is expected between pre-policy and post-policy incomes. The boxplots suggest that the policy reduced income inequality as the post-policy incomes appear closer to one another.

The question we ask is whether the policy reduced poverty, and we proceed by testing the null hypothesis that the pre-policy distribution dominates its post-policy counterpart stochastically at the second order, over the interval $[\underline{t}, \overline{t}] = [1000, 10395]$. The upper boundary point of this interval is 40% of the median of the post-policy income data. Restricted SSD is a poverty-line ordering that compares the distributions using the *per capita income gap* poverty measure over a range of poverty lines. In this illustrative example, the range of poverty

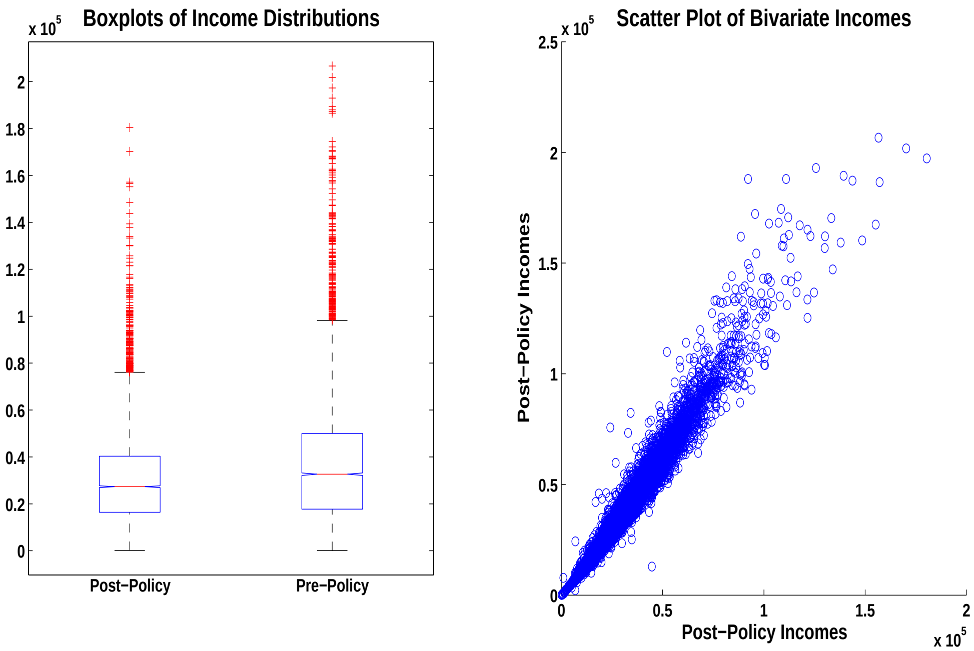

Figure 8: Descriptive plots of the income data.

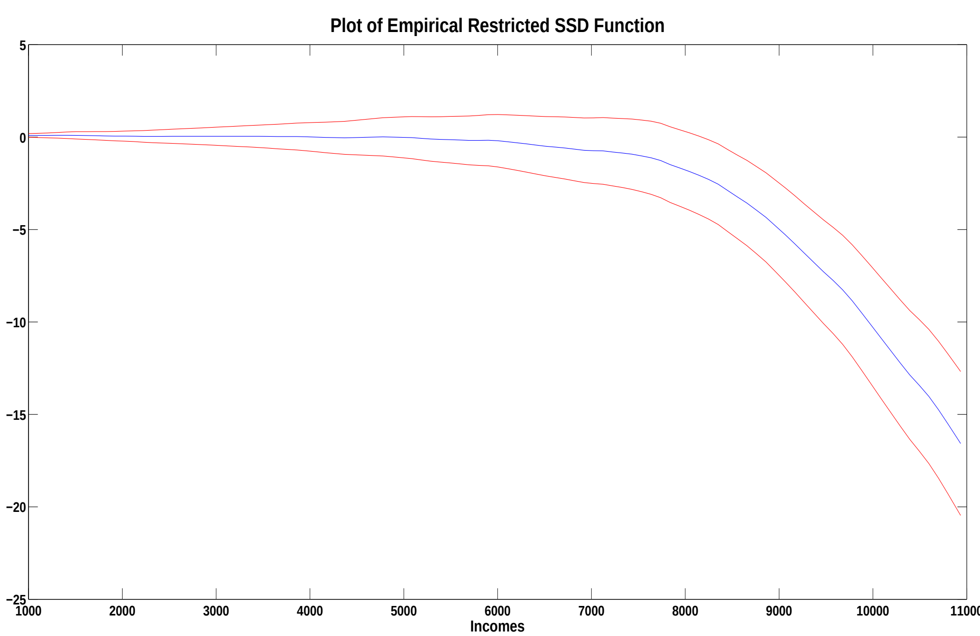

Figure 9: Plot of the empirical restricted SSD function.

lines is given by $[1000, 10395]$, which seem to be reasonable enough to encompass most of the plausible poverty lines for an adult equivalent.

In terms of this paper's notation, the pre-policy distribution corresponds to population B, and the post-policy one to population A. Figure 9 plots the empirical SSD function over this interval in blue along with the 95% (pointwise) confidence interval bands in red. Although it is not easily visible from the figure, the empirical SSD function is positive and no larger than 1 on the set $[1000, 4035.6) \cup (4635.2, 4853.2)$, and non-positive otherwise. Since $n = 9470$, the empirical SSD function is an accurate estimate of its population counterpart; hence, if the true joint distribution belongs to H_1, then it is likely that it corresponds to a configuration in which there are some non-violated inequalities. Therefore, as suggested by the simulation results in Section 5.2, one should use the proposed test over the AS and LSW tests to better detect such a DGP when it is close to H_0.

We set $\alpha(n), N(n), \eta_n(\cdot)$, and the grid's construction as in Section 5. Furthermore, 499 bootstrap samples were used. Table 2 reports the bootstrap p-values for the AS, LSW, and proposed tests.

Table 2: Output of Tests: Bootstrap P-Values

ELR	AS	LSW
0	0.1303	0.6834

The bootstrap p-values of the AS and LSW tests are greater than all of the conventional significance levels. Therefore, these tests does not reject the null hypothesis at all of the conventional significance levels. On the other hand, the bootstrap p-value of the test this paper proposes is 0. Hence, the proposed test rejects the null hypothesis at all conventional significance levels.

7 Conclusion

This paper proposes a new method of testing robust one-way poverty comparisons. Specifically, our bootstrap test has asymptotic sizes that are exactly correct in a uniform sense under regularity conditions. Our simulation study uses restricted stochastic dominance conditions of the third and second orders, and demonstrates that our method works better than the bootstrap tests of Linton et al. (2010) and Andrews and Shi (2010) for quite modest sample sizes in the case of alternative DGPs that have some non-violated inequalities. It should be noted that their tests also apply to first order stochastic dominance conditions, whereas ours does not.

While our setting has focused on matched data, the methods proposed in this paper can be easily extended to the setup of two independent random samples of incomes with natural modifications. The methods proposed in the paper can also be easily extended to multidimensional robust poverty comparisons. In that case, one uses classes of multidimensional poverty measures (e.g. Bourguignon and Chakravarty, 2003), and a concave SIP problem which has a multidimensional index parameter set. Furthermore, the conditions that define the model of the null hypothesis must be adjusted appropriately to reflect the multidimensional nature of the moment functions.

8 Acknowledgments

I started working on the statistical problem this paper addresses when I was a postdoctoral fellow at the Cowles Foundation for Research in Economics. Financial support from the Cowles Foundation is gratefully acknowledged, and I thank them for their hospitality. I also thank Donald Andrews, Yuichi Kitamura, Xiaohong Chen, Colin Cameron, Mervyn Silvapulle, Garry Barrett, Adrian Pagan, Peter Exterkate, Elie Tamer, and Thomas Lok for helpful discussions and comments. Finally, I thank Dr. Brennan Thompson for facilitating my use of the Orca cluster at the SHARCNET computing facility.

References

D. W. K Andrews. Similar-on-the-Boundary Tests for Moment Inequalities Exist, But Have Poor Power. Cowles Foundation Discussion Paper No. 1815, 2011.

D. W. K Andrews and X. Shi. Inference Based on (Possibly Infinitely) Many Moment Inequalities. Cowles Foundation Discussion Paper, Yale University, 2010.

D. W. K Andrews and X. Shi. Inference Based on Conditional Moment Inequalities. *Econometrica*, 82, 2013.

D. W. K. Andrews and X. Shi. Nonparametric inference based on conditional moment inequalities. *Journal of Econometrics*, 179:31–45, 2014.

D. W. K Andrews and G. Soares. Inference for Parameters Defined by Moment Inequalities using Generalized Moment Selection. *Econometrica*, 78(1):119–157, 2010.

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

François Bourguignon and Satya R. Chakravarty. The measurement of multidimensional poverty. *The Journal of Economic Inequality*, 1(1):25–49, 2003. ISSN 1569-1721. doi: 10.1023/A:1023913831342.

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. doi: http://dx.doi.org/10.1016/S0047-2727(96)01589-7. URL `http://www.sciencedirect.com/science/article/pii/S0047272796015897`.

I. A. Canay. EL Inference for Partially Identified Models:Large Deviations Optimality and Bootstrap Validity. *Journal of Econometrics*, 156(2):408–425, 2010.

V. Chernozhukov, S. Lee, and A. Rosen. Intersection Bounds: Estimation and Inference. *Econometrica*, 82, 2013.

Stephen Clark, Richard Hemming, and David Ulph. On indices for the measurement of poverty. *The Economic Journal*, 91(362):pp. 515–526, 1981. ISSN 00130133. URL `http://www.jstor.org/stable/2232600`.

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.

H. El Barmi. Empirical Likelihood Ratio Test for or Against a Set of Inequality Constraints. *Journal of Statistical Planning and Inference*, 55:191–204, 1996.

G. B. Folland. *Real Analysis: Modern Techniques and their Applications*. Wiley Inter-Science, 2nd edition, 1999.

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

J. Foster, E. J. Greer, and E. Thorbecke. A Class of Decomposable Poverty Measures. *Econometrica*, 52:761–766, 1984.

R. Hettich and K. O. Kortanek. Semi-Iinfinite Programming:Theory, Methods, and Applications. *SIAM Review*, 35(3):380–429, 1993.

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

G. Imbens, R. H. Spady, and P. Johnson. Information Theoretic Approaches to Inference in Moment Condition Models. *Econometrica*, 66(2):333–357, 1998.

Y. Kitamura. Asymptotic Optimality of Empirical likelihood for Testing Moment Restrictions. *Econometrica*, 69(6):300–325, 2001.

E. Kreyszig. *Introductory Functional Analysis with Applications*. John Wiley & Sons, 1989.

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.

Marco López and Georg Still. Semi-infinite programming. *European Journal of Operational Research*, 180(2):491 – 518, 2007. ISSN 0377-2217. doi: http://dx.doi.org/10.1016/j.ejor.2006.08.045. URL `http://www.sciencedirect.com/science/article/pii/S0377221706008782`.

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

B. Mordukhovich and T.T.A. Nghia. Constraint qualifications and optimality conditions for nonconvex semi-infinite and infinite programs. *Mathematical Programming*, 139 (1-2):271–300, 2013. ISSN 0025-5610. doi: 10.1007/s10107-013-0672-x. URL `http://dx.doi.org/10.1007/s10107-013-0672-x`.

A. Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75: 237–249, 1988.

Art Owen. *Empirical Likelihood*, volume 92 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, 2001.

J. Qin and J. Lawless. Empirical Likelihood and General Estimating Equations. *The Annals of Statistics*, 22(1):300–325, 1994.

R. Reemsten and S. Gorner. Numerical methods for semi-infinite programming: A survey. In R. Reemsten and J. Ruckmann, editors, *Semi-infinite Programming*. Kluwer Academic Publishers, 1998.

A. Shapiro. Semi-Infinite Programming, Duality, Discretization, and Optimality Conditions. *Optimization*, 8(2):133–161, 2009.

A. Sheehy and J. Wellner. Uniform Donsker Classes of Functions. *The Annals of Probability*, 20 (4):1983–2030, 1992.

G Still. Discretization in semi-infinite programming: the rate of convergence. *Mathematical programming*, 91(1):53–69, 2001.

A. W. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer, first edition, 1996.

Buhong Zheng. Poverty orderings. *Journal of Economic Surveys*, 14(4): 427–466, 2000. ISSN 1467-6419. doi: 10.1111/1467-6419.00117. URL `http://dx.doi.org/10.1111/1467-6419.00117`.

V. Zinde-Walsh. Kernel Estimation when Density may not Exist. *Econometric Theory*, 24(3): 696–725, 2008.

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

Appendix

to

Empirical Likelihood for Robust Poverty Comparisons

Rami V. Tabri

School of Economics, The University of Sydney,

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

A Outline

This Appendix provides proofs of the results in the text. It also introduces further discussions of Assumption 2.1, and of the finite-sample properties of the constrained and unconstrained estimators of the population moments. And it presents the bootstrap procedures of LSW and AS in the framework of this paper.

This Appendix is organized as follows. Section B provides a further discussion of Assumption 2.1. Section C presents the proofs of the results in Section 3 along with other intermediate results, and a further discussion of the finite-sample properties of the constrained and unconstrained estimators of the population moments. Section D presents the proofs of the results in Section 4. Section E contains the auxiliary technical results that are used in the proofs presented in Section C. Similarly, Section F contains the auxiliary technical results that are used in the proofs presented in Section D. Finally, Sections H and I present the bootstrap test procedures of LSW and AS respectively within the paper's setup.

B Further Discussion of Assumption 2.1

This section provides a further discussion of the conditions in Assumption 2.1. Given a set of moment functions, all of the conditions of Assumption 2.1 are easily verifiable in practice. The conditions that are less known to applied researchers are pointwise measurability, and the VC property. For this reason, this section focuses on these two properties.

Sets of moment functions that are continuous on a separable domain are pointwise measurable. The definition of this concept is the following:

Definition B.1. *A class $\mathcal{G}$ of measurable functions, $g : \mathcal{X} \to \mathbb{R}$ on the probability space $(\mathcal{X}, \mathcal{A}, P)$, is pointwise measurable if there exists a countable subset $\mathcal{G}' \subset \mathcal{G}$ such that, for every $g \in \mathcal{G}$ there exists a sequence $\{g'_m\} \in \mathcal{G}'$ with $g'_m(x) \to g(x)$, pointwise for each $x \in \mathcal{X}$.*

To show that a class of moment functions is a VC-class, one can use Theorem 2.6.7 of van der

Vaart and Wellner (1996) (VDW) which is a result on the entropy bounds for such classes. Specifically, if $\mathcal{G}$ is a VC-class, then for $r \geq 1$ and any probability measure Q on $(\mathcal{X}, \mathcal{A})$,

$$N(\epsilon, \mathcal{G}, L_r(Q)) \leq U\, V(\mathcal{G})(16e)^{V(\mathcal{G})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{G})-1)}, \tag{B.1}$$

for a universal constant U and $0 < \epsilon < 1$, where $V(\mathcal{G})$ as the VC-index of the set of subgraphs of functions in $\mathcal{G}$, and $N(\epsilon, \mathcal{G}, L_r(Q))$ is the covering number of $\mathcal{G}$ (i.e. the minimal number of ϵ balls in the $L_r(Q)$ norm needed to cover $\mathcal{G}$). Many classes of functions in practice satisfy this type of bound in their entropy numbers (see van der Vaart and Wellner, 1996, page 134).

The important point is that condition (B.1) implies

$$\int_0^{+\infty} \sqrt{\sup_Q N(\epsilon, \mathcal{G}, L_2(Q))}\, d\epsilon < +\infty, \tag{B.2}$$

where the supremum is taken over all finitely discrete probability measures Q on $(\mathcal{X}, \mathcal{A})$. Condition (B.2) is the uniform entropy bound. This condition along with suitable measurability requirements on a uniformly bounded set of functions $\mathcal{G}$ implies uniform weak convergence of its empirical process, where the uniformity holds over a predesignated set of probability measures (see van der Vaart and Wellner, 1996, Theorem 2.8.3).

An immediate consequence of Assumption 2.1 is that the set $\left\{\mathbf{x} \mapsto g(\mathbf{x}; t),\, t \in [\underline{t}, \overline{t}]\right\}$ is uniformly Donsker and pre-Gaussian with respect to the probability measures in $\mathcal{M}$. For future references, we formalize this result here.

Lemma B.1. *Let $\mathcal{M}$ be the set of probability measures in Definition 2.1. Then the class of moment functions, $\left\{\mathbf{x} \mapsto g(\mathbf{x}; t),\, t \in [\underline{t}, \overline{t}]\right\}$, is Donsker and pre-Gaussian uniformly in $\mathcal{M}$.*

Proof. Let $\mathcal{G} = \left\{\mathbf{x} \mapsto g(\mathbf{x}; t),\, t \in [\underline{t}, \overline{t}]\right\}$. Assumption 2.1 implies that the sets

$$\mathcal{G}_{\delta,P} = \left\{f - h : f, h \in \mathcal{G}, \|f - h\|_{P,2} < \delta\right\} \quad \text{and} \quad \mathcal{G}^2_{\infty} = \left\{(f-h)^2 : f, g \in \mathcal{G}\right\} \tag{B.3}$$

are pointwise measurable for every $\delta > 0$ and $P \in \mathcal{M}$, which is sufficient for them to be P-

measurable for every $\delta > 0$ and $P \in \mathcal{M}$. Therefore, we meet all the conditions in Theorem 2.8.3 of van der Vaart and Wellner (1996), which implies the desired result. $\square$

This intermediate result is the driving force behind the *uniform* asymptotic validity of the proposed test, which is essential for the asymptotic size of the proposed test to provide a good approximation to its finite sample counterpart.

C Proofs of Results in Section 3

This section provides the intermediate results mentioned in Section 3, and the proofs of main results in the same section.

C.1 Intermediate Results

Let $\mathcal{H}_n = \{p_i, i = 1, \ldots, n; \sum_{i=1}^n p_i = 1, p_i \geq 0, \forall i = 1, \ldots, n\}$, and denote the interior of this set by $\mathcal{H}_n^\circ$. Additionally, let $\mathcal{H}_n^0(\mathbf{X}) = \left\{\mathbf{p} \in \mathcal{H}_n : \sum_{i=1}^n p_i g\left(\mathbf{X}_i; t\right) \leq 0 \quad \forall t \in [\underline{t}, \overline{t}] \cap \mathbb{Q}, \right\}$.

Proposition C.1. *On the event* $\{\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset\}$, *the random set*

$$\arg\max\left\{\sum_{i=1}^n \log\left(p_i\right);\ p_i \geq 0,\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n p_i g\left(\mathbf{X}_i; t\right) \leq 0 \quad \forall t \in [\underline{t}, \overline{t}] \cap \mathbb{Q}\right\}$$

is nonempty and a singleton.

Proof. The proof proceeds by verifying the conditions of Weierstrass' Theorem. The objective function is strictly concave in the probabilities. The constraint set, $\mathcal{H}_n^0(\mathbf{X})$, is certainly bounded. It is the countable intersection of closed half-planes (which are convex), and since convexity and closedness are preserved under countable intersections, it is closed and convex. Thus, we are done whenever $\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ \neq \emptyset$. $\square$

Lemma C.1. *Let* $P \in \mathcal{M}$. *Then* $\sup_{P \in \mathcal{M}}$ *Prob*$_P\left[\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ = \emptyset\right] \to 0$ *as* $n \to +\infty$.

Proof. The proof proceeds by the direct method.

Note that $\{\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ = \emptyset\} = \{\forall \mathbf{p} \in \mathcal{H}_n^\circ, \max_{t\in\mathcal{T}} \sum_{i=1}^n p_i g(\mathbf{X}_i, t) > 0\}$. Therefore,

$$\{\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ = \emptyset\} \subset \left\{ \max_{t\in\mathcal{T}} \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_i, t) > 0 \right\}$$

which implies that $\mathrm{Prob}_P[\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ = \emptyset] \leq \mathrm{Prob}_P\left[\max_{t\in\mathcal{T}} \frac{1}{n}\sum_{i=1}^n g(\mathbf{X}_i, t) > 0\right]$. Now the right side of the above inequality is less than or equal to $\mathrm{Prob}_P\left[\max_{t\in\mathcal{T}} \left(\frac{1}{n}\sum_{i=1}^n g(\mathbf{X}_i, t)\right) > 0\right]$ since $P \in \mathcal{M}$ implies $\max_{t\in\mathcal{T}} \Psi(t) \leq 0$. Furthermore, we have

$$\mathrm{Prob}_P\left[\max_{t\in\mathcal{T}}\left[\frac{1}{n}\sum_{i=1}^n g(\mathbf{X}_i, t) - \Psi(t)\right] > 0\right] \leq \mathrm{Prob}_P\left[\max_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^n g(\mathbf{X}_i, t) - \Psi(t)\right| > 0\right].$$

To summarize, we have that

$$\mathrm{Prob}_P\left[\mathcal{H}_n^0(\mathbf{X}) \cap \mathcal{H}_n^\circ = \emptyset\right] \leq \mathrm{Prob}_P\left[\max_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i=1}^n g(\mathbf{X}_i, t) - \Psi(t)\right| > 0\right].$$

Now using the fact that the set of moment functions is uniform Glivenko-Cantelli because it is VC and uniformly bounded, we can conclude that the probabilities on the right side of the above inequality go to zero uniformly in P over $\mathcal{M}$. This concludes the proof. $\square$

C.2 Proof of Proposition 3.1

Proof. The proof proceeds by using the direct method.

$\mathrm{Prob}_P[\mathcal{S}_n] = 1 - \mathrm{Prob}_P[\forall \mathbf{p} \in \mathcal{H}_n^\circ, \max_{t\in\mathcal{T}} \sum_{i=1}^n p_i g(\mathbf{X}_i, t) \geq 0]$. On the complement of $\mathcal{S}_n$, if $\hat{\mathbf{p}} \notin \mathcal{H}_n^0(\mathbf{X})$, then using the same methods in the proof of Lemma C.1 we can show that $\mathrm{Prob}_P[\forall \mathbf{p} \in \mathcal{H}_n^\circ, \max_{t\in\mathcal{T}} \sum_{i=1}^n p_i g(\mathbf{X}_i, t) \geq 0]$ converges to zero uniformly over the elements in $\mathcal{M}$.

$\hat{\mathbf{p}} \in \mathcal{H}_n^0(\mathbf{X})$ cannot occur on the complement of $\mathcal{S}_n$. If it did, then we must have $\tilde{\mathbf{p}} = \hat{\mathbf{p}}$. Now consider the probabilities $\check{\mathbf{p}} = \frac{1}{n}\hat{\mathbf{p}} + (1 - \frac{1}{n})\mathbf{p}^{-}$, where

$$p_i^{-} = \begin{cases} 0, & \text{if } g\left(\mathbf{X}_i, t\right) > 0 \quad \forall t \in [\underline{t}, \overline{t}] \\ \frac{1}{|I_n^{-}|}, & \text{if } g\left(\mathbf{X}_i, t\right) < 0 \quad \forall t \in [\underline{t}, \overline{t}]. \end{cases}$$

and $I_n^{-} = \left\{ i \in \{1, \ldots, n\} : g\left(\mathbf{X}_i; t\right) < 0 \quad \forall t \in [\underline{t}, \overline{t}] \right\}$. Clearly $\check{\mathbf{p}} \in \mathcal{H}_n^{\circ}$ which cannot occur on the complement $\mathcal{S}_n$, and hence, yields a contradiction. Finally, note that by Lemma E.1 (in Appendix E), the probability of the event $\{I_n^{-} \neq \emptyset\}$ tends to 1 as $n \to +\infty$, uniformly over the elements in $\mathcal{M}$. □

C.3 Proof of Theorem 3.1

Proof. The proof proceeds by the direct method. First, consider the case $\Delta\left(P_0\right) = \emptyset$. In this case, for large enough n, $\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i; t\right) < 0 \, \forall t \in [\underline{t}, \overline{t}]$ with probability tending to 1, and hence, by part (i) of Property C.1, $\frac{1}{n}\sum_{i=1}^{n} \tilde{p}_i g\left(\mathbf{X}_i; t\right) < 0 \, \forall t \in [\underline{t}, \overline{t}]$ for large n with probability tending to 1, which implies $\tilde{\mathcal{E}}_n \xrightarrow{P} 0$.

Now consider the case $\Delta\left(P_0\right) = \Delta_d\left(P_0\right) \cup \Delta_c\left(P_0\right)$. This means we will focus on the case in which $\dot{\Delta}\left(P_0\right) = \left\{t_1^b, t_2^b, \ldots\right\}$ is countable. Recall that the set of Lagrange multiplier measure on the inequality constraints is given by the set $ba\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)$. Additionally, let

$$ba\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)_{-} = \left\{\mu \in ba\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right) : \mu_j \le 0 \quad \forall t_j \in [\underline{t}, \overline{t}] \cap \mathbb{Q}\right\}, \tag{C.1}$$

$$ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right) = \left\{\mu \in ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right) : \text{supp}(\mu) \subset \dot{\Delta}\left(P_0\right)\right\}, \quad \text{and} \tag{C.2}$$

$$ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)_{-} = \left\{\mu \in ba\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)_{-} : \text{supp}\left(\mu\right) \subset \dot{\Delta}\left(P_0\right)\right\}. \tag{C.3}$$

The ELR statistic (3.8) can be expressed as

$$\min_{\tau \in ba\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)_{-}} \max_{\mu(\tau) \in ba\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)} 2\left\{\sum_{i=1}^{n} \log\left(1 + \sum_{t \in [\underline{t},\overline{t}]\cap\mathbb{Q}} g\left(\mathbf{X}_i; t\right)\mu\left(\{t\}\right)\right) - n \sum_{t \in [\underline{t},\overline{t}]\cap\mathbb{Q}} \tau\left(\{t\}\right)\mu\left(\{t\}\right)\right\}.$$

By part 1 of Lemma E.4, $\tilde{\mathcal{E}}_n$ can be expressed as

$$\min_{\tau\in ba_0\left(2^{[\underline{t},\bar{t}]\cap\mathbb{Q}}\right)_-}\ \max_{\mu_b(\tau)\in ba_0\left(2^{[\underline{t},\bar{t}]\cap\mathbb{Q}}\right)} 2\left\{\sum_{i=1}^{n}\log\left(1+\sum_{t\in\Delta(\dot{P}_0)} g\left(\mathbf{X}_i;t\right)\mu\left(\{t\}\right)\right)-n\sum_{t\in\Delta(\dot{P}_0)}\tau\left(\{t\}\right)\mu\left(\{t\}\right)\right\}$$

for large n with probability tending to one. Fix $\tau\in ba_0\left(2^{[\underline{t},\bar{t}]\cap\mathbb{Q}}\right)_-$ such that $\sum_{t\in\Delta(\dot{P}_0)}\tau\left(\{t\}\right)=O_P\left(n^{-1/2}\right)$ uniformly in $\mathcal{M}$. Then the first order condition for $\tilde{\mu}_b\left(\tau\right)$ is

$$\frac{1}{n}\sum_{i=1}^{n}\frac{g\left(\mathbf{X}_i;t_k^b\right)}{1+\sum_{j=1}^{+\infty}\mu_{b,j}g\left(\mathbf{X}_i;t_j^b\right)}-\tau_{b,k}=0\quad\forall t_k^b\in\Delta\left(\dot{P}_0\right),\tag{C.4}$$

where $\mu_{b,j}=\mu\left(\left\{t_j^b\right\}\right)\ \forall j$, and $\tau_{b,j}=\tau\left(\left\{t_j^b\right\}\right)\ \forall j$. Let $\gamma_i=\sum_{j=1}^{\infty}g\left(\mathbf{X}_i;t_j^b\right)\tilde{\mu}_{b,j}\left(\tau\right)\ i=1,\ldots n$. Consider the following expansion of (C.4):

$$\begin{aligned}0=\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i;t_j^b\right)-\sum_{c=1}^{\infty}\left[\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i;t_j^b\right)g\left(\mathbf{X}_i;t_c^b\right)\right]\mu_{b,c}\\+\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i;t_j^b\right)\frac{\gamma_i^2}{1+\gamma_i}-\tau_{b,j},\end{aligned}\tag{C.5}$$

which is based on the equality: $\frac{1}{1+\gamma_i}=1-\gamma_i+\frac{\gamma_i^2}{1+\gamma_i}$. Re-arranging (C.5) as follows:

$$\hat{\Psi}(t_j^b)-\tau_{b,j}+r_{1,n}(t_j^b)=\sum_{c=1}^{\infty}\left[\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i;t_j^b\right)g\left(\mathbf{X}_i;t_c^b\right)\right]\mu_{b,c}\tag{C.6}$$

where $r_{1,n}(t_j^b)=\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i;t_j^b\right)\frac{\gamma_i^2}{1+\gamma_i}$ and $\hat{\Psi}(t_j^b)=\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i;t_j^b\right)$, yields the following infinite matrix problem:

$$\underline{\hat{\Psi}_b}-\underline{\tau_b}+\underline{r_{1,n}}=\hat{\Sigma}_\infty\,\underline{\mu_b},\quad\text{where}\quad\underline{\hat{\Psi}_b}=[\hat{\Psi}(t_1^b),\hat{\Psi}(t_2^b),\ldots]'\tag{C.7}$$

$$\underline{\mu_b}=[\mu_{b,1},\mu_{b,2},\ldots]',\quad\underline{\tau_b}=[\tau_{b,1},\tau_{b,2},\ldots]',\quad\underline{r_{1,n}}=[r_{1,n}(t_1^b),r_{1,n}(t_2^b),\ldots]',\tag{C.8}$$

and $\hat{\Sigma}_\infty$ is the sample infinite covariance matrix formed by the moment functions on $\Delta\left(\dot{P}_0\right)$.

For large n, the system (C.7) has a unique solution given by $\underline{\tilde{\mu}_b}\left(\underline{\tau_b}\right)=\hat{\Sigma}_\infty^{-1}\left[\underline{\hat{\Psi}_b}-\underline{\tau_b}+\underline{r_{1,n}}\right]$.

However, this solution does not uniquely define an element in $ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)$ because this set contains measures that are also *not* countably additive; that is, $\sum_{j=1}^{\infty}|\tilde{\mu}_{b,j}\left(\underline{\tau_b}\right)| < \|\tilde{\mu}_b\left(\underline{\tau_b}\right)\|_{TV} < +\infty$ is feasible, where $\sum_{j=1}^{\infty}|\tilde{\mu}_{b,j}\left(\underline{\tau_b}\right)| < +\infty$. Therefore, the set of solutions is given by

$$\Omega = \left\{\mu \in ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right) : \underline{\mu_b} = \hat{\Sigma}_{\infty}^{-1}\left[\underline{\hat{\Psi}_b} - \underline{\tau_b} + \underline{r_{1,n}}\right], \sum_{j=1}^{\infty}|\mu_{b,j}| \leq \|\mu_b\|_{TV}\right\}. \tag{C.9}$$

We circumvent this non-uniqueness by using the Moore-Penrose solution, $\arg\min_{\xi\in\Omega}\{\|\xi\|_{TV}\}$, which is given by $\xi_j^{\star} = \tilde{\mu}_{b,j}\left(\underline{\tau_b}\right) \quad \forall t_j \in \Delta\left(\dot{P}\right), \quad \text{and} \quad \|\xi^{\star}\|_{TV} = \sum_{j=1}^{\infty}|\tilde{\mu}_{b,j}\left(\underline{\tau_b}\right)|$.

Lemma E.4 and part (ii) of Assumption 2.1 implies for each $i = 1, \ldots n$

$$|\gamma_i| \leq \left|\sum_{j=1}^{\infty} g\left(\mathbf{X}_i; t_j\right)\tilde{\mu}_{b,j}\left(\underline{\tau_b}\right)\right| \leq \|\tilde{\mu}_b\left(\underline{\tau_b}\right)\|_{l^1} = O_P\left(n^{-1/2}\right) \text{ uniformly in } \mathcal{M}, \tag{C.10}$$

which in turn implies

$$\|\underline{r_{1,n}}\|_{l^{\infty}_{\infty}} \leq \max_{1\leq i\leq n}\gamma_i^2 = O_P\left(n^{-1}\right) \text{ uniformly in } \mathcal{M}. \tag{C.11}$$

Next, use $\log\left(1+\gamma_i\right) = \gamma_i - \gamma_i^2/2 + r_{2,i}$ where for some finite $C > 0$

$$\text{Prob}_P\left[|r_{2,i}| \leq C|\gamma_i|^3, 1 \leq i \leq n\right] \to 1 \quad n \to +\infty. \tag{C.12}$$

Now we can approximate the likelihood ratio and then use $\underline{\xi^{\star}} = \hat{\Sigma}_{\infty}^{-1}\left[\underline{\hat{\Psi}_b} - \underline{\tau_b} + \underline{r_{1,n}}\right]$.

$$\tilde{\mathcal{E}}_n = \min_{\tau\in ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)_-} 2\left\{\sum_{i=1}^{n}\log\left(1+\sum_{j=1}^{\infty} g\left(\mathbf{X}_i; t_j^b\right)\xi_j^{\star}\right) - n\sum_{j=1}^{\infty}\tau_{b,j}\xi_j^{\star}\right\} \tag{C.13}$$

$$= \min_{\tau\in ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)_-}\left\{2n\underline{\xi^{\star}}'\underline{\hat{\Psi}_b} - n\underline{\xi^{\star}}'\hat{\Sigma}_{\infty}\underline{\xi^{\star}} - 2n\underline{\xi^{\star}}'\underline{\tau} + 2\sum_{i=1}^{n} r_{2,i}\right\} \tag{C.14}$$

$$= \min_{\tau\in ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)_-}\left\{n\left[\underline{\hat{\Psi}_b} - \underline{\tau_b}\right]'\hat{\Sigma}_{\infty}^{-1}\left[\underline{\hat{\Psi}_b} - \underline{\tau_b}\right] - n\underline{r_{1,n}}'\hat{\Sigma}_{\infty}^{-1}\underline{r_{1,n}} + \sum_{i=1}^{n} r_{2,i}\right\} \tag{C.15}$$

$$= T_n + O_P\left(n^{-1/2}\right) \text{ uniformly in } \quad \mathcal{M}(e_0), \tag{C.16}$$

where

$$T_n = \min_{\tau \in ba_0\left(2^{[\underline{t},\bar{t}]\cap\mathbb{Q}}\right)_-} \left\{ n \left[\underline{\hat{\Psi}_b} - \underline{\tau_b}\right]' \hat{\Sigma}_\infty^{-1} \left[\underline{\hat{\Psi}_b} - \underline{\tau_b}\right] \right\} \tag{C.17}$$

$$\left| \sum_{i=1}^n r_{2,i} \right| \leq \sum_{i=1}^n |\gamma_i|^3 \leq n \max_{1\leq i\leq n} |\gamma_i|^3 \leq n O_P\left(n^{-3/2}\right) = O_P\left(n^{-1/2}\right) \text{ uniformly in } \mathcal{M}, \tag{C.18}$$

and for large enough n we have $\left\|\hat{\Sigma}_\infty^{-1}\right\| \leq e_0$ which implies

$$\left| n \underline{r_{1,n}}' \hat{\Sigma}_\infty^{-1} \underline{r_{1,n}} \right| \leq n \left\| \underline{r_{1,n}} \right\|_{l_\infty^\infty}^2 \left\| \hat{\Sigma}_\infty^{-1} \right\| \leq \left\| \underline{r_{1,n}} \right\|_{l_\infty^\infty}^2 e_0 = n\, O_P\left(n^{-2}\right) \text{ uniformly in } \mathcal{M}(e_0). \tag{C.19}$$

Since the asymptotic equivalence (C.16) is uniform in $\mathcal{M}(e_0)$, we can now focus on T_n to prove the weak convergence of $\tilde{\mathcal{E}}_n$ to the QLR statistic.

Next note that

$$\begin{aligned} T_n &= \min_{\tau \in ba_0\left(2^{[\underline{t},\bar{t}]\cap\mathbb{Q}}\right)_-} \left\{ \left[\sqrt{n}\underline{\hat{\Psi}_b} - \sqrt{n}\underline{\tau_b}\right]' \hat{\Sigma}_\infty^{-1} \left[\sqrt{n}\underline{\hat{\Psi}_b} - \sqrt{n}\underline{\tau_b}\right] \right\} \\ &= \min_{\underline{u_b} \in l_{\infty,-}^\infty} \left\{ \left[\sqrt{n}\underline{\hat{\Psi}_b} - \underline{u_b}\right]' \hat{\Sigma}_\infty^{-1} \left[\sqrt{n}\underline{\hat{\Psi}_b} - \underline{u_b}\right] \right\} \end{aligned} \tag{C.20}$$

where $\underline{u_b} = \sqrt{n}\underline{\tau_b}$. Then, define $\sqrt{n}\underline{\hat{\Psi}_b} \xrightarrow{P} \mathbf{G}_\infty \sim \text{MVN}\left(\mathbf{0}_\infty, \Sigma_\infty\right)$ on $\Delta\left(\dot{P}_0\right)$, so that $\hat{\Sigma}_\infty^{-1} \xrightarrow{P} \left(\Sigma_\infty\left(P_0\right)\right)^{-1}$ in the operator norm (2.10) uniformly over $\mathcal{M}(e_0)$, and

$$\min_{\underline{u_b} \in l_{\infty,-}^\infty} \left\{ \left[\sqrt{n}\underline{\hat{\Psi}_b} - \underline{u_b}\right]' \hat{\Sigma}_\infty^{-1} \left[\sqrt{n}\underline{\hat{\Psi}_b} - \underline{u_b}\right] \right\} \rightsquigarrow \min_{\underline{u_b} \in l_{\infty,-}^\infty} \left\{ \left[\mathbf{G}_\infty - \underline{u_b}\right]' \left(\Sigma_\infty\left(P_0\right)\right)^{-1} \left[\mathbf{G}_\infty - \underline{u_b}\right] \right\} \tag{C.21}$$

uniformly over $\mathcal{M}(e_0)$ by Proposition G.1.

Finally, the case $\Delta\left(P_0\right) = \Delta_d\left(P_0\right) = \left\{t_1^b, \ldots, t_w^b\right\}, w \in \mathbb{Z}_+$ follows similar steps as in the previous case while keeping track of the fact that we have a finite dimensional problem. The main differences are $ba_0\left(2^{[\underline{t},\bar{t}]\cap\mathbb{Q}}\right) = \left\{\omega \in ba_0\left(2^{[\underline{t},\bar{t}]\cap\mathbb{Q}}\right) : \text{supp}(\omega) \subset \Delta_d\left(P_0\right)\right\}$, and the matrix equation (C.7) is now finite dimensional, where the population covariance matrix $\Sigma_w\left(P_0\right)$ has a bounded inverse. So one uses the sequence spaces l_w^1 and l_w^∞ instead of l_∞^1 and l_∞^∞. $\square$

C.4 Proof of Theorem 3.2

Proof. The proof proceeds by the direct method. The ELR statistic (3.15) can be decomposed as follows:

$$\acute{\mathcal{E}}_n = \tilde{\mathcal{E}}_n + 2\left\{\sum_{i=1}^{n}\left(\log\left(\tilde{p}_i\right) - \log\left(\acute{p}_i\right)\right)\right\}, \tag{C.22}$$

where $\tilde{p}_1, \ldots, \tilde{p}_n$ is the solution of the concave SIP optimization problem (3.1). By definition, $2\left\{\sum_{i=1}^{n}\left(\log\left(\tilde{p}_i\right) - \log\left(\acute{p}_i\right)\right)\right\} \leq 0$. Therefore

$$\begin{aligned} 0 \leq \left|2\left\{\sum_{i=1}^{n}\left(\log\left(\tilde{p}_i\right) - \log\left(\acute{p}_i\right)\right)\right\}\right| &= 2\sum_{i=1}^{n}\left(\log\left(\acute{p}_i\right) - \log\left(\tilde{p}_i\right)\right) \\ &= 2\sum_{i=1}^{n}\left(\log\left(1 + \frac{\acute{p}_i - \tilde{p}_i}{\tilde{p}_i}\right)\right). \end{aligned} \tag{C.23}$$

Using the inequality $\log(1+x) \leq x \,\forall x > -1$, and the FONC for $\tilde{p}_1, \ldots, \tilde{p}_n$,

$$2\sum_{i=1}^{n}\left(\log\left(1 + \frac{\acute{p}_i - \tilde{p}_i}{\tilde{p}_i}\right)\right) \leq 2\sum_{i=1}^{n}\frac{\acute{p}_i - \tilde{p}_i}{\tilde{p}_i} = 2\sum_{t\in\Delta(\tilde{\mathbf{p}})}\tilde{\mu}\left(\{t\}\right)\sum_{i=1}^{n}\acute{p}_i g\left(\mathbf{X}_i; t\right). \tag{C.24}$$

Now using Lemma E.2, we have that

$$2\sum_{t\in\Delta(\tilde{\mathbf{p}})}\tilde{\mu}\left(\{t\}\right)\sum_{i=1}^{n}\acute{p}_i g\left(\mathbf{X}_i; t\right) \leq 2\min\left(L\, d_n, \alpha(n)\right)\sum_{t\in\Delta(\tilde{\mathbf{p}})}\tilde{\mu}\left(\{t\}\right), \tag{C.25}$$

and by Lemma E.3, $2\max\left(L\, d_n, \alpha(n)\right)\sum_{t\in\Delta(\tilde{\mathbf{p}})}\tilde{\mu}\left(\{t\}\right) = 2\min\left(L\, d_n, \alpha(n)\right)\|\tilde{\mu}\|_{l^1_{w_n}} = o_P(1)$, where w_n is the cardinality of $\Delta\left(\tilde{\mathbf{p}}\right)$.

Finally, the result for the uniform rate of convergence follows directly since $L\, d_n \leq \frac{L}{N(n)}$. This completes the proof. $\square$

C.5 Further Discussion for Section 3.3

This section presents finite-sample results that are useful for increasing the numerical accuracy of the computation algorithm Section 3.3 proposes. These results are relationships between the constrained and unconstrained estimators of the population moments, which follow from a property of the moment functions.

The form of the moment functions being differences of the same function, as in (2.2) and (2.3), implies that they have the following property.

Property C.1. *[Sign Conditions]*

Let dom(t) *denote the domain of definition of the index parameter.*

1. *Let* $z = \max\left\{z^A, z^B\right\}$ *. For the ranking of distributions over a poverty aversion parameter with given poverty lines:* $\forall \mathbf{x} \in \mathbb{R}_+^2 - [0,z] \times [0,z], g(\mathbf{x};t) = 0 \quad \forall t \in$ *dom*(t) *and for each* $\mathbf{x} \in [0,z] \times [0,z]$*, either* $g(\mathbf{x};t) \leq 0 \, \forall t \in$ *dom*(t)*, or* $g(\mathbf{x};t) \geq 0 \, \forall t \in$ *dom*(t)*, or* $g(\mathbf{x};t) = 0 \, \forall t \in$ *dom*(t).

2. *For the ranking of distributions over poverty lines: for each* $\mathbf{x} \in \mathbb{R}_+^2$ *such that* $x^A \neq x^B$*, either* $g(\mathbf{x};t) \leq 0 \, \forall t \in$ *dom*(t) *or* $g(\mathbf{x};t) \geq 0 \, \forall t \in$ *dom*(t)*, and* $g(\mathbf{x};t) = 0 \, \forall t \in$ *dom*(t) *whenever* $x^A = x^B$.

Property C.1 states that the sign of the functions g is determined by the configuration in its data dimension independently of t.

A consequence of Property C.1 on the Slater event $\mathcal{S}_n$ is the following.

Proposition C.2. 1. *Prob*$_{P_0}\left[\sum_{i=1}^n \tilde{p}_i g(\mathbf{X};t) \leq \frac{1}{n}\sum_{i=1}^n g(\mathbf{X};t) \; \forall t \in [\underline{t}, \overline{t}] \, | \mathcal{S}_n\right] = 1;$

2. *Prob*$_{P_0}\left[\Delta(\tilde{\mathbf{p}}) \subset \left\{t \in [\underline{t}, \overline{t}] : \frac{1}{n}\sum_{i=1}^n g(\mathbf{X};t) \geq 0\right\} | \mathcal{S}_n\right] = 1.$

Proof. **Part 1**:the proof proceeds by the direct method.

Given $t \in [\underline{t}, \overline{t}]$, $\frac{1}{n}\sum_{i=1}^n g(\mathbf{X}_i;t) - \sum_{i=1}^n \tilde{p}_i g(\mathbf{X}_i;t) = \sum_{i=1}^n \tilde{p}_i \sum_{j=1}^\infty \tilde{\mu}_j g(\mathbf{X}_i, t_j) \, g(\mathbf{X}_i, t)$. By Property C.1, for each i $g(\mathbf{X}_i, t_j) \, g(\mathbf{X}_i, t) \geq 0 \quad \forall t \in [\underline{t}, \overline{t}]$, which implies the result.

Part 2:the proof proceeds by the direct method.

Let $t \in \Delta(\tilde{\mathbf{p}})$, then $0 \leq \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{X}_i;t) - \sum_{i=1}^{n} \tilde{p}_i g(\mathbf{X}_i;t) = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{X}_i;t)$, where the non-negativity follows from part 1 of this proposition. $\square$

Part 1 of Proposition C.2 indicates that conditional on the Slater event, $\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{X};t)$ (weakly) stochastically dominates $\sum_{i=1}^{n} \tilde{p}_i g(\mathbf{X};t)$, at first order, uniformly over $[\underline{t}, \overline{t}]$ on a set of probability measure one. Part 2 of Proposition C.2 indicates that with probability one, the index set of active at $\tilde{\mathbf{p}}$ constraints must be at points in $[\underline{t}, \overline{t}]$ where $\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{X};t)$ is non-negative. Proposition C.2 does not require $P_0 \in \mathcal{M}$, which implies that it is solely a consequence of the estimation procedure.

These results are useful for speeding up and checking numerical computation. Because these results are inequality restrictions that relate the constrained and unconstrained estimators of the population moments, they thus hold at the grid-points in $\mathcal{T}_N$; therefore, they can be imposed as constraints in (3.12) when implementing the algorithm.

D Proofs of Results in Section 4

D.1 Proof of Theorem 4.1

Proof. The proof follows the same steps as those in the proof of part (1) of Theorem 3.2. The bootstrapped ELR statistic (4.5) can be decomposed as follows:

$$\overline{\mathcal{E}}_n^\star = \tilde{\mathcal{E}}_n^\star + 2\left\{\sum_{i=1}^{n} \left(\log\left(\tilde{p}_i^\star\right) - \log\left(\overline{p}_i^\star\right)\right)\right\}, \tag{D.1}$$

where $\tilde{p}_1^\star, \ldots, \tilde{p}_n^\star$ is the solution of the concave SIP optimization problem (F.20) in Proposition F.1, and $\tilde{\mathcal{E}}_n^\star$ is the bootstrapped ELR statistic based on them. Proposition F.1 shows that $\tilde{\mathcal{E}}_n^\star$ converges in distribution to the asymptotic distribution of the ELR statistic (3.2) conditional on $\left\{\overline{P}_n : n \geq 1\right\}$ for almost every sample path. Therefore, to complete the proof, we need to show the second term in (D.1) converges to zero.

To that end,

$$0 \leq \left| 2\left\{ \sum_{i=1}^{n} \left(\log\left(\tilde{p}_i^\star\right) - \log\left(\overline{p}_i^\star\right) \right) \right\} \right| = 2\left\{ \sum_{i=1}^{n} \left(\log\left(\overline{p}_i^\star\right) - \log\left(\tilde{p}_i^\star\right) \right) \right\}$$
$$= 2\sum_{i=1}^{n} \left(\log\left(1 + \frac{\overline{p}_i^\star - \tilde{p}_i^\star}{\tilde{p}_i^\star} \right) \right). \tag{D.2}$$

Using the inequality $\log(1+x) \leq x \, \forall x > -1$, and the FONC for $\tilde{p}_1^\star, \ldots, \tilde{p}_n^\star$, (D.2) is less than or equal to

$$2\sum_{i=1}^{n} \frac{\overline{p}_i^\star - \tilde{p}_i^\star}{\tilde{p}_i^\star} = 2 \sum_{t \in \Delta(\tilde{\mathbf{p}}^\star)} \tilde{\mu}^\star\left(\{t\}\right) \left(\sum_{i=1}^{n} \overline{p}_i^\star g\left(\mathbf{X}_i^\star; t\right) + \eta_n\left(t\right) \right) \tag{D.3}$$
$$\leq 2\min\left\{ \frac{L}{N(n)}, \alpha(n) \right\} \|\tilde{\mu}^\star\|_{l^1_{w_n}} + 2\|\tilde{\mu}^\star\|_{l^1_{w_n}}, \tag{D.4}$$

where w_n is the cardinality of $\Delta\left(\tilde{\mathbf{p}}\right)$, $\|\tilde{\mu}^\star\|_{l^1_{w_n}} = \sum_{t\in\Delta(\tilde{\mathbf{p}}^\star)} \tilde{\mu}^\star\left(\{t\}\right)$. by By Lemma F.2, $\|\tilde{\mu}^\star\|_{l^1_{w_n}} \xrightarrow{P} 0$ conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}$. Therefore, under the assumption $\min\left\{\frac{L}{N(n)}, \alpha(n)\right\} = o(1)$, the right side of (D.3) converges to zero in probability conditional on on $\mathcal{A}_n$ uniformly in $\mathcal{M}$. $\square$

D.2 Proof of Corollary 4.1

Proof. The proof proceeds by the direct method. For a large number of bootstrap replications, we have for each $e_0 \in \mathbb{R}_+$ that

$$\limsup_{n\to+\infty} \sup_{P\in\mathcal{M}(e_0)} \mathrm{Prob}_P\left[\Upsilon_{B_n} \leq \beta\right] = \limsup_{n\to+\infty} \sup_{P\in\mathcal{M}(e_0)} \mathrm{Prob}_P\left[\mathrm{Prob}_P\left[\acute{\mathcal{E}}_n \leq \overline{\mathcal{E}}_n^\star | \mathcal{A}_n\right] \leq \beta\right]. \tag{D.5}$$

Then, a direct application of Theorem 4.1 to the right side of (D.5) yields

$$\limsup_{n\to+\infty} \sup_{P\in\mathcal{M}(e_0)} \mathrm{Prob}_P\left[\mathrm{Prob}_P\left[\acute{\mathcal{E}}_n \leq \overline{\mathcal{E}}_n^\star | \mathcal{A}_n\right] \leq \beta\right] \leq \beta, \tag{D.6}$$

since the conditional distribution of $\overline{\mathcal{E}}_n^\star$ and the unconditional distribution of $\acute{\mathcal{E}}_n$ are equal asymptotically in a uniform sense. An inequality holds in (D.6) because the asymptotic null distribution of the ELR statistic can have a mass point at zero in the case $\Delta\left(P_0\right) \neq \emptyset$. $\square$

D.3 Proof of Theorem 4.2

Proof. Under the assumptions of this theorem, Lemma F.4 says the approximate ELR statistic (3.15) diverges to $+\infty$ as the sample size increases i.e. $\acute{\mathcal{E}}_n \to +\infty$. Therefore, to prove the result, all we need to do is to show $\overline{\mathcal{E}}_n^\star = O_P(1)\overline{P}_n -$ a.e..

The bootstrapped approximate ELR statistic can be expressed as

$$\begin{aligned} \overline{\mathcal{E}}_n^\star &= 2\sum_{i=1}^n \log\left(1 + \sum_{t\in\Delta(\overline{\mathbf{p}}^\star)} \acute{\mu}_t^\star\left(g\left(\mathbf{X}_i^\star;t\right) + \eta_n(t)\right)\right) \\ &\leq 2\log\left(1 + \sum_{t\in\Delta(\overline{\mathbf{p}}^\star)} \acute{\mu}_t^\star \sum_{i=1}^n \left(g\left(\mathbf{X}_i^\star;t\right) + \eta_n(t)\right)\right) \end{aligned} \tag{D.7}$$

by Jensen's inequality for a concave function. Adding and subtracting $\overline{\Psi}_n(t) = \sum_{i=1}^n \overline{p}_i g\left(\mathbf{X}_i;t\right)$ under the logarithm on the right side of (D.7) as follows

$$2\log\left(1 + \sum_{t\in\Delta(\overline{\mathbf{p}}^\star)} \acute{\mu}_t^\star \sum_{i=1}^n \left(g\left(\mathbf{X}_i^\star;t\right) - \overline{\Psi}_n(t) + \overline{\Psi}_n(t) + \eta_n(t)\right)\right), \tag{D.8}$$

and multiplying and dividing by $\sqrt{n}$ yields

$$2\log\left(1 + \sqrt{n}\sum_{t\in\Delta(\overline{\mathbf{p}}^\star)} \acute{\mu}_t^\star \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(g\left(\mathbf{X}_i^\star;t\right) - \overline{\Psi}_n(t) + \overline{\Psi}_n(t) + \eta_n(t)\right)\right), \tag{D.9}$$

implies (D.9) is $2\log\left(1 + O_P(1)\right)\overline{P}_n -$ a.e. by Lemma F.6, parts 2 and 3 of Lemma F.5, and the Central Limit Theorem. Therefore, $\overline{\mathcal{E}}_n^\star \leq O_P(1)\overline{P}_n -$ a.e., which completes the proof. $\square$

E Auxiliary Technical Results For Section 3

Let $I_n^- = \left\{i \in \{1,\ldots,n\} : g\left(\mathbf{X}_i; t\right) < 0 \quad \forall t \in [\underline{t}, \overline{t}]\right\}$. We have the following result concerning its large sample behavior.

Lemma E.1. *Let $P \in \mathcal{M}$. Then $\sup_{P\in\mathcal{M}}$ Prob$_P\left[I_n^- \neq \emptyset\right] \to 1$.*

Proof. The proof proceeds by the direct method. We show that the probability of the complement of $\{I_n^- \neq \emptyset\}$ converges to zero. Note that Property C.1 implies

$$\left\{I_n^- \neq \emptyset\right\} = \left\{\text{for each i } g\left(\mathbf{X}_i; t\right) \geq 0 \quad \forall t \in [\underline{t}, \overline{t}]\right\}.$$

We have by the bivariate random sampling assumption on $\{\mathbf{X}_i\}_{i=1}^n$,

$$\sup_{P\in\mathcal{M}} \text{Prob}_P\left[I_n^- \neq \emptyset\right] = \sup_{P\in\mathcal{M}} \left(\text{Prob}_P\left[g\left(\mathbf{X}_1; t\right) \geq 0 \quad \forall t \in [\underline{t}, \overline{t}]\right]\right)^n$$

which must converge to either zero or 1. It can only converge to zero because only a $P \notin \mathcal{M}$ has $g\left(\mathbf{X}_1; t\right) \geq 0$ for each t almost surely. □

Lemma E.2. *Let $\acute{\mathbf{p}}$ be the solution of the exchange algorithm with tolerance parameter $\alpha(n)$. Then,*

$$\sup_{t\in[\underline{t},\overline{t}]} \sum_{i=1}^n \acute{p}_i g\left(\mathbf{X}_i, t\right) \leq \min\left(L\, d_n, \alpha(n)\right), \tag{E.1}$$

where $d_n =$ dist $\left(\mathcal{T}_{N(n)}, [\underline{t}, \overline{t}] \cap \mathbb{Q}\right)$ is the Hausdorff distance (3.13), and L is the Lipschitz constant arising from the Lipschitz continuity of the moment functions, g.

Proof. The proof proceeds by the direct method and follows similar steps as in Lemma 1 of Still (2001). First, we note that L in (E.1) depends only on the class of moment functions that are being used and on $\mathcal{T}$. Let t_d be a solution of

$$\max_{t\in[\underline{t},\overline{t}]} \sum_{i=1}^n \acute{p}_i g\left(\mathbf{X}_i, t\right)$$

and let $\hat{t}_d \in \mathcal{T}_{N(n)}$ such that $|\hat{t}_d - t_d| \leq d_n$. By Lipschitz continuity of the moment functions g and using $\sum_{i=1}^{n} \acute{p}_i g\left(\mathbf{X}_i, \hat{t}_d\right) \leq 0$ we find

$$\sum_{i=1}^{n} \acute{p}_i g\left(\mathbf{X}_i, t_d\right) \leq \sum_{i=1}^{n} \acute{p}_i g\left(\mathbf{X}_i, t_d\right) - \sum_{i=1}^{n} \acute{p}_i g\left(\mathbf{X}_i, \hat{t}_d\right) \leq L|\hat{t}_d - t_d| \leq L\, d_n \quad \forall t \in [\underline{t}, \overline{t}]. \tag{E.2}$$

The result follows directly since $\acute{\mathbf{p}}$ also satisfies

$$\sum_{i=1}^{n} \acute{p}_i g\left(\mathbf{X}_i, t\right) \leq \alpha(n) \quad \forall t \in [\underline{t}, \overline{t}].$$

□

Lemma E.3. *Suppose $P_0 \in \mathcal{M}$. Then*

1. $\Delta\left(\tilde{\mathbf{p}}\right) \subset \Delta\left(P_0\right)$ *for large* n.

2. *Let $\tilde{\mu}$ be the Lagrange multiplier measure in the FONCs (3.5) and (3.6). If $P_0 \in \mathcal{M}$ and $w_n = \left|\Delta\left(\tilde{\mathbf{p}}\right)\right|$, then $\|\tilde{\mu}\|_{l^1_{w_n}} = \sum_{t \in \Delta(\tilde{\mathbf{p}})} \tilde{\mu}\left(\{t\}\right) = o_P(1)$ uniformly in $\mathcal{M}$ at the $\sqrt{n}$ – rate.*

Proof. **Part 1.** Given $t \in [\underline{t}, \overline{t}] \cap \mathbb{Q}$, we prove this result by showing that

$$t \notin \Delta\left(P_0\right) \implies t \notin \Delta\left(\tilde{\mathbf{p}}\right) \quad \text{for large}\, n.$$

Proposition C.2 implies

$$\sum_{i=1}^{n} \tilde{p}_i g\left(\mathbf{X}_i, t\right) \leq \frac{1}{n} \sum_{i=1}^{n} g\left(\mathbf{X}_i, t\right) = \frac{1}{n} \sum_{i=1}^{n} g\left(\mathbf{X}_i, t\right) - E_{P_0}\left[g\left(\mathbf{X}; t\right)\right] + E_{P_0}\left[g\left(\mathbf{X}; t\right)\right]. \tag{E.3}$$

Then, for such t we have $E_{P_0}\left[g\left(\mathbf{X}; t\right)\right] < 0$, which implies $\sum_{i=1}^{n} \tilde{p}_i g\left(\mathbf{X}_i, t\right) < 0$ for large enough n since

$$\frac{1}{n} \sum_{i=1}^{n} g\left(\mathbf{X}_i, t\right) - E_{P_0}\left[g\left(\mathbf{X}; t\right)\right] = O_P(n^{-1/2}) \quad \text{by LLN.}$$

This implies $t \notin \Delta\left(\tilde{\mathbf{p}}\right)$ for large enough n.

Part 2. Let $w_n = |\Delta(\tilde{\mathbf{p}})|$. For ease of exposition, we will use the following notation: $\mu_j = \mu(\{t_j\})$. The the equality constraints are

$$\frac{1}{n}\sum_{i=1}^{n}\frac{g(\mathbf{X}_i;t_k)}{1+\sum_{t_j\in\Delta(\tilde{\mathbf{p}})}\tilde{\mu}_j g(\mathbf{X}_i;t_j)} = 0 \quad \forall t_k \in \Delta(\tilde{\mathbf{p}}). \tag{E.4}$$

Let $\tilde{\mu} = \|\tilde{\mu}\|\theta$ where $\theta \in \mathbb{R}_+^{w_n}$ such that $\sum_j \theta_j = 1$. Because the Lagrange multipliers are nonnegative, the elements of θ must be non-negative and sum to unity, which means they are weights.

If $w_n = \infty$, which means $\Delta(\tilde{\mathbf{p}})$ is countable, then without loss of generality, suppose $\Delta(\tilde{\mathbf{p}}) = \{t_1, t_2, \ldots\}$, and let $\underline{g_i} = [g(\mathbf{X}_i;t_1), g(\mathbf{X}_i;t_2), \ldots]'$ and $\underline{\tilde{\mu}} = [\tilde{\mu}_1, \tilde{\mu}_2, \ldots]'$. Additionally, if w_n is positive and finite, then without loss of generality, suppose $\Delta(\tilde{\mathbf{p}}) = \{t_1, t_2, \ldots, t_{w_n}\}$, and let $\underline{g_i} = [g(\mathbf{X}_i;t_1), g(\mathbf{X}_i;t_2), \ldots, g(\mathbf{X}_i;t_{w_n})]'$ and $\underline{\tilde{\mu}} = [\tilde{\mu}_1, \tilde{\mu}_2, \ldots, \tilde{\mu}_{w_n}]'$.

The system (E.4) implies

$$\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\underline{g_i}}{1+\underline{\tilde{\mu}}'\underline{g_i}}\right) = 0. \tag{E.5}$$

Let $Y_i = \underline{\tilde{\mu}}'\underline{g_i}$ and use $\frac{1}{1+Y_i} = 1 - \frac{Y_i}{1+Y_i}$ to expand (E.5) as follows:

$$\|\tilde{\mu}\|_{l^1_{w_n}}\left(\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\underline{g_i}\left(\underline{g_i}\right)'}{1+Y_i}\right)\theta\right) = \theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}\right). \tag{E.6}$$

The sample analogue estimator of $\Sigma_{w_n}(P_0)$ is $\hat{\Sigma}_{w_n} = \frac{1}{n}\sum_{i=1}^{n}\underline{g_i}\left(\underline{g_i}\right)'$. Since $1+Y_i > 0 \quad \forall i$,

$$\begin{aligned}\|\tilde{\mu}\|_{l^1_{w_n}}\left(\theta'\hat{\Sigma}_{w_n}\theta\right) &\le \|\tilde{\mu}\|_{l^1_{w_n}}\left(\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\underline{g_i}\left(\underline{g_i}\right)'}{1+Y_i}\right)\theta\right)\left(1+\max_{i=1,\ldots,n}|Y_i|\right)\\ &\le \|\tilde{\mu}\|_{l^1_{w_n}}\left(\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\underline{g_i}\left(\underline{g_i}\right)'}{1+Y_i}\right)\theta\right)\left(1+\|\tilde{\mu}\|_{l^1_{w_n}}\right).\end{aligned} \tag{E.7}$$

Using (E.6), we can substitute out $\|\tilde{\mu}\|_{l^1_{w_n}}\left(\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\frac{g_i\left(g_i\right)'}{1+Y_i}\right)\theta\right)$ from (E.7) yielding

$$\|\tilde{\mu}\|_{l^1_{w_n}}\left(\theta'\hat{\Sigma}_{w_n}\theta\right)\leq\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}\right)\left(1+\|\tilde{\mu}\|_{l^1_{w_n}}\right). \tag{E.8}$$

Rewriting (E.8) as follows

$$\|\tilde{\mu}\|_{l^1_{w_n}}\left(\theta'\hat{\Sigma}_{w_n}\theta-\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}\right)\right)\leq\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}\right), \tag{E.9}$$

yields an upper bound on $\|\tilde{\mu}\|_{l^1_{w_n}}$.

Now we prove that $\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}\right)$ converges to zero in probability and at the desired rate, uniformly in $\mathcal{M}$. Note that

$$\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}\right)=\sum_j\theta_j\left(\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i;t_j\right)\right) \tag{E.10}$$

$$=\sum_j\theta_j\left(\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i;t_j\right)-E_{P_0}\left[g\left(\mathbf{X};t_j\right)\right]\right), \tag{E.11}$$

since $t_j\in\Delta\left(\tilde{\mathbf{p}}\right)\implies t_j\in\Delta\left(\dot{P}_0\right)$ for large n by part 1 of this lemma. This implies

$$\begin{aligned}\left|\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}\right)\right|&\leq\left(\sum_j\theta_j\right)\max_j\left|\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i;t_j\right)-E_{P_0}\left[g\left(\mathbf{X};t_j\right)\right]\right|\\&=\sup_{t\in[\underline{t},\overline{t}]}\left|\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i;t\right)-E_{P_0}\left[g\left(\mathbf{X};t\right)\right]\right|.\end{aligned} \tag{E.12}$$

The VC property of the class of moment functions in Assumption 2.1 implies the right side of (E.9) is $o_P(1)$ uniformly in $\mathcal{M}$.

All that is left to conclude the proof is to show that $\theta'\hat{\Sigma}_{w_n}\theta>c$ uniformly in $\mathcal{M}$, where the constant c is defined in Definition 2.1. Part 1 of this lemma implies $\hat{\Sigma}_{w_n}$ is a sub-covariance matrix of $\hat{\Sigma}_w$, where $w = \left|\Delta\left(\dot{P}_0\right)\right|$. Therefore, $\theta'\hat{\Sigma}_{w_n}\theta$ can be expressed as

$$\theta'\hat{\Sigma}_{w_n}\theta = \upsilon'\hat{\Sigma}_w\upsilon, \tag{E.13}$$

where $\upsilon \in \mathbb{R}^w$ is such that

$$\upsilon_j = \begin{cases} \theta_j, & \text{if } t_j \in \Delta\left(\tilde{\mathbf{p}}\right), \\ 0, & \text{if } t_j \in \Delta\left(\dot{P}_0\right) - \Delta\left(\tilde{\mathbf{p}}\right). \end{cases}$$

Now the injectivity condition of Definition 2.1 implies $\upsilon'\hat{\Sigma}_w\upsilon > c$ holds for large enough n since $P_0 \in \mathcal{M}$. Hence,

$$\|\tilde{\mu}\|_{l^1_{w_n}} \leq \frac{o_P(1)}{c + o_P(1)}, \tag{E.14}$$

which implies that $\|\tilde{\mu}\|_{l^1_{w_n}} = o_P(1)$ uniformly in $\mathcal{M}$. Finally, the $\sqrt{n}$ rate of uniform convergence of $\|\tilde{\mu}\|_{l^1_{w_n}}$ is a consequence of the class of moment functions being uniform Donsker. $\square$

Lemma E.4. *Consider the set of binding moments, $\Delta\left(\dot{P}_0\right)$, and recall that $\tilde{\mu} \in ba\left(2^{[\underline{t},\bar{t}]\cap\mathbb{Q}}\right)$ is the Lagrange multiplier measure on $\mathbb{Q} \cap [\underline{t}, \bar{t}]$ described in Section 3.1.*

1. *For large n, $\tilde{\mu}(\{t\}) = 0 \quad \forall t \in \mathbb{Q} \cap [\underline{t}, \bar{t}] - \Delta\left(\dot{P}_0\right)$, uniformly in $\mathcal{M}$.*

2. *Let $\tilde{\mu}_b\left(\tau_{n,b}\right)$ be an element of the set*

$$\underset{\mu_b \in ba_0\left(2^{[\underline{t},\bar{t}]\cap\mathbb{Q}}\right)}{\arg\max} \left\{2\sum_{i=1}^{n} \log\left(1 + \sum_{t \in \Delta(\dot{P}_0)} \mu_b\left(\{t\}\right) g\left(\mathbf{X}_i; t\right)\right) - n \sum_{t \in \Delta(\dot{P}_0)} \mu_b\left(\{t\}\right) \tau_{n,b}\left(\{t\}\right)\right\},$$

where

$$ba_0\left(2^{[\underline{t},\bar{t}]\cap\mathbb{Q}}\right) = \left\{\nu \in ba\left(2^{[\underline{t},\bar{t}]\cap\mathbb{Q}}\right) : supp\left(\nu\right) = \Delta\left(\dot{P}_0\right)\right\},$$

$\tau_{n,b}\left(\{t\}\right) \leq 0 \, \forall t \in \Delta\left(\dot{P}_0\right)$, and $\sum_{t \in \Delta(\dot{P}_0)} \tau_{n,b}\left(\{t\}\right) = O_P\left(n^{-1/2}\right)$ uniformly in $\mathcal{M}$. Then, for $P_0 \in \mathcal{M}$, $\sum_{t \in \Delta(\dot{P}_0)}^{\infty} \left|\tilde{\mu}_b(\{t\}, \tau_{n,b})\right| = O_p(n^{-1/2})$, uniformly in $\mathcal{M}$.

Proof. The proof proceeds by the direct method.

Part 1. For large n, the Slater condition (3.4) holds. Therefore, by Proposition C.2:

$$\sum_{i=1}^{n} \tilde{p}_i g\left(\mathbf{X}_i; t\right) \leq \frac{1}{n} \sum_{i=1}^{n} g\left(\mathbf{X}_i; t\right) \quad \forall t \in \mathbb{Q} \cap [\underline{t}, \overline{t}].$$

Since $\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i; t\right) < 0$ for each t such that $E_{P_0}\left[g\left(\mathbf{X}; t\right)\right] < 0$, by complementary slackness it follows that $\tilde{\mu}(\{t\}) = 0$ for such t. The uniformity in the convergence of $\tilde{\mu}_s$ to zero holds since the set of moment functions is uniform Glivenko-Cantelli, because it is uniform Donsker.

Part 2. First suppose that $\Delta_c\left(P_0\right) \neq \emptyset$. Then $\Delta\left(\dot{P}_0\right) = \left\{t_1^b, t_2^b, \ldots\right\}$ is countable. For ease of exposition, we will use the following notation: $\mu_{b,j} = \mu\left(\left\{t_j^b\right\}\right)$, and $\tau_{n,b,j} = \tau_n\left(\left\{t_j^b\right\}\right)$. The first order condition for $\tilde{\mu}_b\left(\tau_{n,b}\right)$ is

$$\frac{1}{n}\sum_{i=1}^{n} \frac{g\left(\mathbf{X}_i; t_k^b\right)}{1 + \sum_{j=1}^{+\infty} \mu_{b,j} g\left(\mathbf{X}_i; t_j^b\right)} - \tau_{n,b,k} = 0 \quad \forall t_k^b \in \Delta\left(\dot{P}_0\right). \tag{E.15}$$

Let $\tilde{\mu}_b\left(\tau_{n,b}\right) = \|\tilde{\mu}_b\left(\tau_{n,b}\right)\|_{l^1_\infty} \theta$ where $\theta \in l^1_\infty$ such that $\|\theta\|_{l^1_\infty} = 1$. Furthermore, for ease of exposition, we will suppress $\tau_{n,b}$ in the notation the Lagrange multiplier.

Let $\underline{g_i}^b = \left[g\left(\mathbf{X}_i; t_1^b\right), g\left(\mathbf{X}_i; t_2^b\right), \ldots\right]'$, $\underline{\tilde{\mu}_b} = \left[\tilde{\mu}_{b,1}, \tilde{\mu}_{b,2}, \ldots\right]'$ and $\underline{\tau_{n,b}} = \left[\tau_{n,b,1}, \tau_{n,b,2}, \ldots\right]'$. The system (E.15) implies

$$\theta' \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\underline{g_i}^b}{1 + \underline{\tilde{\mu}_b}' \underline{g_i}^b} - \underline{\tau_{n,b}}\right) = 0. \tag{E.16}$$

Let $Y_i = \underline{\tilde{\mu}_b}' \underline{g_i}^b$ and use $\frac{1}{1+Y_i} = 1 - \frac{Y_i}{1+Y_i}$ to expand (E.16) as follows:

$$\|\tilde{\mu}_b\|_{l^1_\infty} \left(\theta' \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\underline{g_i}^b \left(\underline{g_i}^b\right)'}{1 + Y_i}\right) \theta\right) = \theta' \left(\frac{1}{n}\sum_{i=1}^{n} \underline{g_i}^b - \underline{\tau_{n,b}}\right). \tag{E.17}$$

The sample analogue estimator of Σ_∞ is $\hat{\Sigma}_\infty = \frac{1}{n}\sum_{i=1}^{n} \underline{g_i}^b \left(\underline{g_i}^b\right)'$. Since $1 + Y_i > 0 \quad \forall i$,

$$\begin{aligned}
\|\tilde{\mu}_b\|_{l^1_\infty}\left(\theta'\hat{\Sigma}_\infty\theta\right) &\leq \|\tilde{\mu}_b\|_{l^1_\infty}\left(\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\underline{g_i}^b\left(\underline{g_i}^b\right)'}{1+Y_i}\right)\theta\right)\left(1+\max_{i=1,\ldots,n}|Y_i|\right) \\
&\leq \|\tilde{\mu}_b\|_{l^1_\infty}\left(\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\underline{g_i}^b\left(\underline{g_i}^b\right)'}{1+Y_i}\right)\theta\right)\left(1+\|\tilde{\mu}_b\|_{l^1_\infty}\max_{i=1,\ldots,n}\left\|\underline{g_i}^b\right\|_{l^\infty_\infty}\right). \qquad \text{(E.18)}
\end{aligned}$$

Using (E.17), we can substitute out $\|\tilde{\mu}_b\|_{l^1_\infty}\left(\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\underline{g_i}^b\left(\underline{g_i}^b\right)'}{1+Y_i}\right)\theta\right)$ from (E.18) yielding

$$\|\tilde{\mu}_b\|_{l^1_\infty}\left(\theta'\hat{\Sigma}_\infty\theta\right) \leq \theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}^b - \underline{\tau_{n,b}}\right)\left(1+\|\tilde{\mu}_b\|_{l^1_\infty}\max_{i=1,\ldots,n}\left\|\underline{g_i}^b\right\|_{l^\infty_\infty}\right). \qquad \text{(E.19)}$$

Rewriting (E.19) as follows

$$\|\tilde{\mu}_b\|_{l^1_\infty}\left(\theta'\hat{\Sigma}_\infty\theta - \max_{i=1,\ldots,n}\left\|\underline{g_i}^b\right\|_{l^\infty_\infty}\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}^b - \underline{\tau_{n,b}}\right)\right) \leq \theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}^b - \underline{\tau_{n,b}}\right), \qquad \text{(E.20)}$$

yields an upper bound on $\|\tilde{\mu}_b\|_{l^1_\infty}$.

Now we prove that $\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}^b\right)$ converges to zero in probability and at the desired rate, uniformly in $\mathcal{M}$. Note that

$$\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}^b\right) = \sum_j \theta_j\left(\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i; t_j^b\right)\right) \quad = \sum_j \theta_j\left(\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i; t_j^b\right) - E_{P_0}\left[g\left(\mathbf{X}; t_j^b\right)\right]\right). \qquad \text{(E.21)}$$

Therefore,

$$\begin{aligned}
\left|\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}^b\right)\right| &\leq \left(\sum_j |\theta_j|\right)\max_j\left|\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i; t_j^b\right) - E_{P_0}\left[g\left(\mathbf{X}; t_j^b\right)\right]\right| \\
&\leq \sup_{t\in[\underline{t},\bar{t}]}\left|\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i; t\right) - E_{P_0}\left[g\left(\mathbf{X}; t\right)\right]\right|. \qquad \text{(E.22)}
\end{aligned}$$

The VC property of the class of moment functions in Assumption 2.1 implies the right side of (E.19) is $o_P(1)$ uniformly in $\mathcal{M}$. The $\sqrt{n}$ rate of uniform convergence of $\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}^b\right)$ results from the fact that the class of moment function is uniform Donsker.

By similar arguments to those in Lemma 11.2 of Owen (2001), we have $\max_{i=1,\dots,n}\left\|\underline{g_i}^b\right\|_{l_\infty^\infty} = o\left(\sqrt{n}\right)$. This implies

$$\max_{i=1,\dots,n}\left\|\underline{g_i}^b\right\|_{l_\infty^\infty}\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\underline{g_i}^b-\underline{\tau_{n,b}}\right)=o\left(\sqrt{n}\right)O_P\left(n^{-1/2}\right)=o_P(1), \tag{E.23}$$

uniformly in $\mathcal{M}$. Hence, it follows that

$$\|\tilde{\mu}_b\|_{l_\infty^1}\left(\theta'\hat{\Sigma}_\infty\theta-o_P(1)\right)\leq O_P\left(n^{-1/2}\right). \tag{E.24}$$

Finally, since the injectivity condition of Definition 2.1 implies $\theta'\hat{\Sigma}_\infty\theta>0$ for large n, the inequality (E.24) implies $\|\tilde{\mu}_b\|_{l_\infty^1}=O_P\left(n^{-1/2}\right)$.

The proof when $\Delta_c(P_0)=\emptyset$ follows similar steps as in the previous case. The contact set is finite in this case; therefore, the difference is that the norm of l_w^1 with $w<+\infty$ is used instead of the l_∞^1 norm. □

F Auxiliary Technical Results For Section 4

F.1 Asymptotic Validity Under H_0

Lemma F.1. *Let $\overline{\Psi}(t)=\sum_{i=1}^{n}\overline{p}_i g\left(\mathbf{X}_i;t\right)$, where $\left(\overline{p}_1,\dots,\overline{p}_n\right)$ is the bootstrap DGP described in Section 4. If $P_0\in\mathcal{M}$, then the following two statements hold.*

1. *If $t\in[\underline{t},\overline{t}]\cap\mathbb{Q}$ and $E_{P_0}\left[g\left(\mathbf{X};t\right)\right]<0$, then $\overline{\Psi}(t)<-\eta_n(t)$ for large n with probability one.*
2. *If $t\in\Delta\left(P_0\right)$ and $t\in\mathcal{T}_{N(n)}$ for large enough n, then $\overline{\Psi}(t)=-\eta_n(t)$ for large n with probability one.*
3. *If $t\in\Delta\left(P_0\right)$ and $t\notin\mathcal{T}_{N(n)}$ $\forall n$, then $\overline{\Psi}(t)=-\eta_n(t)$ for large n with probability one.*

Proof. The proof follows similar steps in the proof of Lemma B.4 in Canay (2010).

Part 1. The proof proceeds using the direct method. Suppose $t \in [\underline{t}, \overline{t}] \cap \mathbb{Q}$ and $E_{P_0}[g(\mathbf{X}; t)] < 0$. From

$$0 \leq \max_{\mu_j \geq 0, j=1,\ldots,N(n)} 2 \sum_{i=1}^{n} \log \left[1 + \sum_{j=1}^{N(n)} \mu_j \left(g(\mathbf{X}_i; t_j) + \eta_n(t_j) \right) \right]$$
$$2n \log \left[1 + \sum_{j=1}^{N(n)} \overline{\mu}_j \left(\frac{1}{n} \sum_{i=1}^{n} g(\mathbf{X}_i; t) + \eta_n(t_j) \right) \right], \tag{F.1}$$

it follows that $\sum_{j=1}^{N(n)} \overline{\mu}_j \left(\frac{1}{n} \sum_{i=1}^{n} g(\mathbf{X}_i; t_j) + \eta_n(t_j) \right) \geq 0$. Now,

$$\begin{aligned} -\left(\overline{\Psi}(t) + \eta_n(t) \right) &= -\frac{1}{n} \sum_{i=1}^{n} \left[\frac{g(\mathbf{X}_i; t) + \eta_n(t)}{1 + \sum_{j=1}^{N(n)} \overline{\mu}_j \left(g(\mathbf{X}_i; t_j) + \eta_n(t_j) \right)} \right] \\ &\geq \frac{-\left(\frac{1}{n} \sum_{i=1}^{n} g(\mathbf{X}_i; t) + \eta_n(t) \right)}{1 + \sum_{j=1}^{N(n)} \overline{\mu}_j \frac{1}{n} \sum_{i=1}^{n} \left(g(\mathbf{X}_i; t_j) + \eta_n(t_j) \right)} \end{aligned} \tag{F.2}$$

by Jensen's inequality. Since $\sum_{j=1}^{N(n)} \overline{\mu}_j \frac{1}{n} \sum_{i=1}^{n} \left(g(\mathbf{X}_i; t_j) + \eta_n(t_j) \right) \geq 0$ for all n and

$$-\left(\frac{1}{n} \sum_{i=1}^{n} g(\mathbf{X}_i; t) + \eta_n(t) \right) > 0$$

for large n with probability 1, we have

$$-\overline{\Psi}(t) - \eta_n(t) = -\frac{1}{n} \sum_{i=1}^{n} \left[\frac{g(\mathbf{X}_i; t) + \eta_n(t)}{1 + \sum_{j=1}^{N(n)} \overline{\mu}_j \left(g(\mathbf{X}_i; t_j) + \eta_n(t_j) \right)} \right] > 0 \tag{F.3}$$

with probability one, which is equivalent to $\overline{\Psi}(t) < -\eta_n(t)$ with probability one.

Part 2. The proof proceeds by the direct method. First, for large enough n that $t \in \mathcal{T}_{N(n)}$. Then it must be equal to one of t_j $j = 1, \ldots, N(n)$, and suppose $t = t_k \in \mathcal{T}_{N(n)}$ without loss of generality. Set $\mu_j = 0 \, \forall j = 1, \ldots, N(n)$ and consider the FONC for μ_k,

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{g\left(\mathbf{X}_i;t_k\right)+\eta_n(t_k)}{1+\sum_{j=1}^{N(n)}\mu_j\left(g\left(\mathbf{X}_i;t_j\right)+\eta_n(t_j)\right)}\right]\Bigg|_{\mu_j=0\,\forall j=1,\ldots,N(n)} = \frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i;t_k\right)+\eta_n(t_k) \quad \text{(F.4)}$$

We know $\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i;t_k\right)+\eta_n(t_k) \leq 0$ for large n with probability 1. If

$$\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i;t_k\right)+\eta_n(t_k) = 0,$$

we are done since

$$\overline{\Psi}(t_k) = \frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i;t_k\right) = \eta_n(t_k)$$

and $\mu_k = 0$ is optimal. If $\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i;t_k\right)+\eta_n(t_k) < 0$, then the optimal value of μ_k has to increase (so it will be positive) by continuity of the objective function in μ_k. Since the optimal solution has to satisfy $\mu_k\left(\overline{\Psi}(t_k)+\eta_n(t_k)\right) = 0$, it follows that $\overline{\Psi}(t_k) = -\eta_n(t_k)$.

Part 3. The proof proceeds by the direct method. Let $t_d \in \mathcal{T}_{N(n)}$ for large enough n. Then

$$\overline{\Psi}(t)+\eta_n(t) = \left(\overline{\Psi}(t)-\overline{\Psi}(t_d)\right)+\left(\eta_n(t)-\eta_n(t_d)\right)+\overline{\Psi}(t_d)+\eta_n(t_d), \quad \text{(F.5)}$$

where

$$\left(\overline{\Psi}(t)-\overline{\Psi}(t_d)\right) = o(1)\; n \to +\infty \text{ because } \left|\overline{\Psi}(t)-\overline{\Psi}(t_d)\right| \leq L/N(n) \quad \text{(F.6)}$$

by Lipschitz continuity of the moment functions where L is the Lipschitz constant (see Assumption 2.1), $\left(\eta_n(t)-\eta_n(t_d)\right) = o_p(1)$ by property (4.2), and $\overline{\Psi}(t_d)+\eta_n(t_d) = 0$ for large n with probability tending to one by part 2 of this lemma. $\square$

Lemma F.2. *Suppose $P_0 \in \mathcal{M}$. Let $\overline{P}_n$ be the bootstrap DGP described in Section 4, and let*

$\{\mathbf{X}_i^\star\}_{i=1}^n$ be IID $\overline{P}_n$. Furthermore, let $\tilde{\mathbf{p}}^\star$ denote the solution of the following SIP problem

$$\max_{p_1,\ldots,p_n}\left\{\sum_{i=1}^n \log\left(p_i\right);\, p_i \geq 0,\, \sum_{i=1}^n p_i = 1,\, \sum_{i=1}^n p_i g\left(\mathbf{X}_i^\star;t\right) \leq -\eta_n(t) \quad \forall t \in [\underline{t},\overline{t}] \cap \mathbb{Q},\right\}. \tag{F.7}$$

Then,

1. *$\Delta\left(\tilde{\mathbf{p}}^\star\right) \subset \Delta\left(P_0\right)$ for large n, where*

$$\Delta\left(\tilde{\mathbf{p}}^\star\right) = \left\{t \in [\underline{t},\overline{t}] \cap \mathbb{Q} : \sum_{i=1}^n \tilde{p}_i^\star g\left(\mathbf{X}_i^\star;t\right) = -\eta_n(t)\right\}. \tag{F.8}$$

2. *Let $w_n^\star$ denote the cardinality of the set (F.8), and let $\tilde{\mu}^\star$ denote the Lagrange multiplier measure on the inequality constraints in the SIP problem (F.7). Then,*

$$\|\tilde{\mu}^\star\|_{l^1_{w_n^\star}} = \sum_{t\in\Delta(\tilde{\mathbf{p}}^\star)} \tilde{\mu}^\star\left(\{t\}\right) \xrightarrow{P} 0$$

conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}$.

Proof. **Part 1.** The proof proceeds by the direct method. Given $t \in [\underline{t},\overline{t}] \cap \mathbb{Q}$, we prove this result by showing that

$$t \notin \Delta\left(P_0\right) \implies t \notin \Delta\left(\tilde{\mathbf{p}}^\star\right) \quad \text{for large } n.$$

Note that $t \notin \Delta\left(P_0\right) \iff E_{P_0}\left[g\left(\mathbf{X};t\right)\right] < 0$ for $P_0 \in \mathcal{M}$. Then,

$$-\frac{1}{n}\sum_{i=1}^n\left[\frac{g\left(\mathbf{X}_i^\star;t\right)+\eta_n(t)}{1+\sum_{j=1}^{+\infty}\tilde{\mu}_j^\star\left(g\left(\mathbf{X}_i^\star;t_j\right)+\eta_n(t_j)\right)}\right] \geq \frac{-\left(\frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}_i^\star;t\right)+\eta_n(t)\right)}{1+\sum_{j=1}^{+\infty}\tilde{\mu}_j^\star\left(\frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}_i^\star;t_j\right)+\eta_n(t_j)\right)} \tag{F.9}$$

by Jensen's inequality. The right side of (F.9) is equal to

$$-\frac{\left(\frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}_i^\star;t\right)-\sum_{i=1}^n \overline{p}_i g\left(\mathbf{X}_i;t\right)\right)}{1+\sum_{j=1}^{+\infty}\tilde{\mu}_j^\star\left(\frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}_i^\star;t_j\right)+\eta_n(t_j)\right)} - \frac{\left(\sum_{i=1}^n \overline{p}_i g\left(\mathbf{X}_i;t\right)+\eta_n(t)\right)}{1+\sum_{j=1}^{+\infty}\tilde{\mu}_j^\star\left(\frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}_i^\star;t_j\right)+\eta_n(t_j)\right)}, \tag{F.10}$$

where the first term is $O_P\left(n^{-1/2}\right)$ $\overline{P}_n$ a.e., and part 1 of Lemma F.1 implies the second term is positive with probability tending to one. Thus, the left side of (F.9) is positive for n sufficiently large $\overline{P}_n$ a.e., which is equivalent to

$$\sum_{i=1}^{n} \tilde{p}_i^\star g\left(\mathbf{X}_i^\star; t\right) < -\eta_n(t) \tag{F.11}$$

for n sufficiently large $\overline{P}_n$ a.e..

Part 2. The proof follows the same steps as those in the part 2 of Lemma E.3. Let $w_n^\star = \left|\Delta\left(\tilde{\mathbf{p}}^\star\right)\right|$. For ease of exposition, we will use the following notation: $\mu_j^\star = \mu^\star\left(\{t_j\}\right)$. The the equality constraints are

$$\frac{1}{n}\sum_{i=1}^{n} \frac{g\left(\mathbf{X}_i^\star; t_k\right) + \eta_n(t_k)}{1 + \sum_{t_j \in \Delta(\tilde{\mathbf{p}}^\star)} \tilde{\mu}_j^\star \left(g\left(\mathbf{X}_i^\star; t_j\right) + \eta_n(t_j)\right)} = 0 \quad \forall t_k \in \Delta\left(\tilde{\mathbf{p}}^\star\right). \tag{F.12}$$

Let $\tilde{\mu}^\star = \|\tilde{\mu}^\star\|_{l^1_{w_n^\star}} \theta$ where $\theta \in \mathbb{R}_+^{w_n^\star}$ such that $\sum_j \theta_j = 1$. Because the Lagrange multipliers are nonnegative, the elements of θ must be non-negative and sum to unity, which means they are weights.

If $w_n^\star = \infty$, which means $\Delta\left(\tilde{\mathbf{p}}^\star\right)$ is countable, then without loss of generality, suppose $\Delta\left(\tilde{\mathbf{p}}^\star\right) = \{t_1, t_2, \ldots\}$, and let $\underline{g_i^\star} + \underline{\eta_n} = [g\left(\mathbf{X}_i^\star; t_1\right) + \eta_n(t_1), g\left(\mathbf{X}_i^\star; t_2\right) + \eta_n(t_2), \ldots]'$ and $\underline{\tilde{\mu}^\star} = [\tilde{\mu}_1^\star, \tilde{\mu}_2^\star, \ldots]'$. Additionally, if $w_n^\star$ is positive and finite, then without loss of generality, suppose $\Delta\left(\tilde{\mathbf{p}}^\star\right) = \left\{t_1, t_2, \ldots, t_{w_n^\star}\right\}$, and let

$$\underline{g_i^\star} + \underline{\eta_n} = [g\left(\mathbf{X}_i^\star; t_1\right) + \eta_n(t_1), g\left(\mathbf{X}_i^\star; t_2\right) + \eta_n(t_2), \ldots, g\left(\mathbf{X}_i^\star; t_{w_n^\star}\right) + \eta_n(t_{w_n^\star})]'$$

and $\underline{\tilde{\mu}^\star} = [\tilde{\mu}_1^\star, \tilde{\mu}_2^\star, \ldots, \tilde{\mu}_{w_n^\star}^\star]'$.

The system (F.12) implies

$$\theta' \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\underline{g_i^\star} + \underline{\eta_n}}{1 + \underline{\tilde{\mu}^\star}'\left(\underline{g_i^\star} + \underline{\eta_n}\right)}\right) = 0. \tag{F.13}$$

Let $Y_i^\star = \underline{\tilde{\mu}}^{\star\prime}\left(\underline{g_i^\star} + \underline{\eta_n}\right)$ and use $\frac{1}{1+Y_i^\star} = 1 - \frac{Y_i^\star}{1+Y_i^\star}$ to expand (F.13) as follows:

$$\|\tilde{\mu}^\star\|_{l^1_{w_n^\star}} \left(\theta' \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\left(\underline{g_i^\star} + \underline{\eta_n}\right)\left(\underline{g_i^\star} + \underline{\eta_n}\right)'}{1+Y_i^\star}\right)\theta\right) = \theta' \left(\frac{1}{n}\sum_{i=1}^{n}\left(\underline{g_i^\star} + \underline{\eta_n}\right)\right). \tag{F.14}$$

Let $\hat{\Sigma}^\star_{w_n^\star}(\eta) = \frac{1}{n}\sum_{i=1}^{n}\left(\underline{g_i^\star} + \underline{\eta_n}\right)\left(\underline{g_i^\star} + \underline{\eta_n}\right)'$. Since $1 + Y_i^\star > 0 \quad \forall i$,

$$\begin{aligned}\|\tilde{\mu}^\star\|_{l^1_{w_n^\star}} \left(\theta'\hat{\Sigma}^\star_{w_n^\star}(\eta)\theta\right) &\le \|\tilde{\mu}^\star\|_{l^1_{w_n^\star}} \left(\theta' \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\left(\underline{g_i^\star} + \underline{\eta_n}\right)\left(\underline{g_i^\star} + \underline{\eta_n}\right)'}{1+Y_i^\star}\right)\theta\right)\left(1 + \max_{i=1,\ldots,n}|Y_i^\star|\right) \\ &\le \|\tilde{\mu}^\star\|_{l^1_{w_n^\star}} \left(\theta' \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\left(\underline{g_i^\star} + \underline{\eta_n}\right)\left(\underline{g_i^\star} + \underline{\eta_n}\right)'}{1+Y_i^\star}\right)\theta\right)\left(1 + \|\tilde{\mu}^\star\|\right).\end{aligned} \tag{F.15}$$

Using (F.14), we can substitute out $\|\tilde{\mu}^\star\|_{l^1_{w_n^\star}} \left(\theta' \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\left(\underline{g_i^\star} + \underline{\eta_n}\right)\left(\underline{g_i^\star} + \underline{\eta_n}\right)'}{1+Y_i^\star}\right)\theta\right)$ from (F.15) yielding

$$\|\tilde{\mu}^\star\|_{l^1_{w_n^\star}} \left(\theta'\hat{\Sigma}^\star_{w_n^\star}(\eta)\theta\right) \le \theta' \left(\frac{1}{n}\sum_{i=1}^{n}\left(\underline{g_i^\star} + \underline{\eta_n}\right)\right)\left(1 + \|\tilde{\mu}^\star\|_{l^1_{w_n^\star}}\right). \tag{F.16}$$

Rewriting (F.16) as follows

$$\|\tilde{\mu}^\star\|_{l^1_{w_n^\star}} \left(\theta'\hat{\Sigma}^\star_{w_n^\star}(\eta)\theta - \theta' \left(\frac{1}{n}\sum_{i=1}^{n}\left(\underline{g_i^\star} + \underline{\eta_n}\right)\right)\right) \le \theta' \left(\frac{1}{n}\sum_{i=1}^{n}\left(\underline{g_i^\star} + \underline{\eta_n}\right)\right), \tag{F.17}$$

yields an upper bound on $\|\tilde{\mu}^\star\|_{l^1_{w_n^\star}}$.

Now we prove that $\theta'\left(\frac{1}{n}\sum_{i=1}^{n}\left(\underline{g_i^\star} + \underline{\eta_n}\right)\right) \xrightarrow{P} 0$ conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}$. Since the elements of θ are non-negative and sum to unity, it suffices to show that

$$\text{Prob}_P\left[\left|\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i^\star; t\right) + \eta_n(t)\right| > \epsilon|\mathcal{A}_n\right] \xrightarrow{P} 0 \quad \text{uniformly in} \quad \mathcal{M}, \tag{F.18}$$

for each $t \in \Delta\left(P_0\right)$, which is the content of Lemma G.3.

All that is left to complete the proof is to show that $\theta'\hat{\Sigma}^\star_{w_n^\star}(\eta)\theta > 0$ for large n conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}$. From part 1 of this lemma, we can conclude that $\hat{\Sigma}^\star_{w_n^\star}(\eta)$ is a sub-covariance matrix of $\hat{\Sigma}^\star_w(\eta)$, where $w = \left|\Delta\left(\dot{P}_0\right)\right|$. for large enough n. Therefore, $\theta'\hat{\Sigma}^\star_{w_n^\star}(\eta)\theta$ can be expressed as

$$\theta'\hat{\Sigma}^\star_{w_n^\star}(\eta)\theta = \upsilon'\hat{\Sigma}^\star_w(\eta)\upsilon, \tag{F.19}$$

where $\upsilon \in \mathbb{R}^w$ is such that

$$\upsilon_j = \begin{cases} \theta_j, & \text{if } t_j \in \Delta\left(\tilde{\mathbf{p}}^\star\right), \\ 0, & \text{if } t_j \in \Delta\left(\dot{P}_0\right) - \Delta\left(\tilde{\mathbf{p}}^\star\right). \end{cases}$$

By Part 1 of LemmaG.4, for large n, $\hat{\Sigma}^\star_w(\eta)$ gets close to $\Sigma_w\left(P_0\right)$ in operator norm conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}$. Then, the injectivity condition from Definition 2.1 implies $\theta'\hat{\Sigma}^\star_{w_n^\star}\theta = \upsilon'\hat{\Sigma}^\star_w\upsilon > c > 0$ conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}$. □

Proposition F.1. *Let $\overline{P}_n$ be the bootstrap DGP described in Section 4, and let $\{\mathbf{X}_i^\star\}_{i=1}^n$ be IID $\overline{P}_n$. Furthermore, let $\tilde{\mathcal{E}}_n^\star = 2\left\{-n\log(n) - \tilde{l}^{r,\star}\right\}$, where*

$$\tilde{l}^{r,\star} = \max_{p_1,\ldots,p_n}\left\{\sum_{i=1}^n \log\left(p_i\right);\, p_i \geq 0,\, \sum_{i=1}^n p_i = 1,\, \sum_{i=1}^n p_i g\left(\mathbf{X}_i^\star; t\right) \leq -\eta_n(t) \quad \forall t \in [\underline{t}, \overline{t}] \cap \mathbb{Q}, \right\}. \tag{F.20}$$

For every $e_0 \in \mathbb{R}_+$,

$$\tilde{\mathcal{E}}_n^\star \rightsquigarrow \begin{cases} 0, & \textit{if } w = 0, \\ \min_{\mathbf{U}\in l^\infty_{w,-}}\left(\mathbf{G} - \mathbf{U}\right)'\Sigma_w^{-1}\left(\mathbf{G} - \mathbf{U}\right), & \textit{if } w \neq 0, \end{cases}$$

conditional on $\mathcal{A}_n$ in P_0 uniformly in $\mathcal{M}(e_0)$.

Proof. The proof proceeds by the direct method. The ELR statistic $\tilde{\mathcal{E}}_n^\star$ can be expressed as

$$\min_{\tau\in ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)_-}\ \max_{\mu(\tau)\in ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)} 2\left\{\sum_{i=1}^n \log\left(1 + \sum_{t\in[\underline{t},\overline{t}]\cap\mathbb{Q}} g\left(\mathbf{X}_i^\star; t\right)\mu\left(\{t\}\right)\right) - n\sum_{t\in[\underline{t},\overline{t}]\cap\mathbb{Q}} \tau\left(\{t\}\right)\mu\left(\{t\}\right)\right\} \tag{F.21}$$

From part 1 of Lemma F.2, for $t \in [\underline{t}, \overline{t}] \cap \mathbb{Q}$ such that $E_{P_0}[g(\mathbf{X}; t)] < 0$, we have

$$\sum_{i=1}^{n} \tilde{p}_i^{\star} g(\mathbf{X}_i^{\star}; t) < -\eta_n(t) \tag{F.22}$$

for n sufficiently large $\overline{P}_n$ a.e.. Then equation (F.22) implies $\tilde{\mu}^{\star}(\{t\}) = 0$ for n sufficiently large $\overline{P}_n$ a.e. by complementary slackness. So the ELR statistic is equal to

$$\min_{\tau \in ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)_-} \max_{\mu(\tau) \in ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)} 2\left\{\sum_{i=1}^{n} \log\left(1 + \sum_{t \in \Delta(\dot{P}_0)} g(\mathbf{X}_i^{\star}; t)\, \mu(\{t\})\right) - n \sum_{t \in \Delta(\dot{P}_0)} \tau(\{t\})\, \mu(\{t\})\right\} \tag{F.23}$$

for large n $\overline{P}_n$ a.e..

By similar arguments as those in the proof of part 2 of Lemma E.4 it follows $\|\tilde{\mu}_b^{\star}\|_{l_w^1} = O_P\left(n^{-1/2}\right)$ conditional on $\mathcal{A}_n$ in P uniformly in $\mathcal{M}$, where $w = \left|\Delta(\dot{P}_0)\right|$. Suppose that $\Delta_c(P_0) \neq \emptyset$. Then by following the steps in Theorem 3.1, the corresponding first order condition for $\tilde{\mu}_b^{\star}(\tau_b)$ is

$$\sum_{i=1}^{n} \left[\frac{g\left(\mathbf{X}_i^{\star}; t_k^b\right) + \eta_n(t_k^b)}{1 + \gamma_i^{\star}}\right] = n\tau_{b,k} \quad \forall t_k^b \in \Delta(\dot{P}_0), \tag{F.24}$$

where $\|\tilde{\mu}_b^{\star}\|_{l_\infty^1} = O_P\left(n^{-1/2}\right)$, $\gamma_i^{\star} = \sum_{j=1}^{+\infty} \tilde{\mu}_{b,j}^{\star}(\tau_b)\left(g\left(\mathbf{X}_i^{\star}; t_j^b\right) + \eta_n(t_j^b)\right)$, and $\max_{i=1,\ldots,n} |\gamma_i^{\star}| = o_P(1)$. Therefore, using the same expansion in (C.12), the statistic $\tilde{\mathcal{E}}_n^{\star}$ is given by

$$\begin{aligned}
\tilde{\mathcal{E}}_n^{\star} &= \min_{\tau \in ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)_-} \left\{n\left[\underline{\hat{\Psi}_b}^{\star} + \underline{\eta_b} - \underline{\tau_b}\right]'\left(\hat{\Sigma}_\infty^{\star}\right)^{-1}\left[\underline{\hat{\Psi}_b}^{\star} + \underline{\eta_b} - \underline{\tau_b}\right]\right\} + o_P(1) \\
&= \min_{\tau \in ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)_-} \left\{n\left[\left(\underline{\hat{\Psi}_b}^{\star} - \underline{\overline{\Psi}_b}\right) + \left(\underline{\eta_b} + \underline{\overline{\Psi}_b}\right) - \underline{\tau_b}\right]'\left(\hat{\Sigma}_\infty^{\star}\right)^{-1}\left[\left(\underline{\hat{\Psi}_b}^{\star} - \underline{\overline{\Psi}_b}\right) + \left(\underline{\eta_b} + \underline{\overline{\Psi}_b}\right) - \underline{\tau_b}\right]\right\}, \\
&\qquad + o_P(1) \\
&= T_n^{\star}\left(\underline{\eta_b}\right) + o_P(1),
\end{aligned} \tag{F.25}$$

where

$$\hat{\Sigma}_\infty^{\star}(\eta) = \frac{1}{n}\sum_{i=1}^{n}\left(\underline{g_{i_b}}^{\star} + \underline{\eta}\right)\left(\underline{g_{i_b}}^{\star} + \underline{\eta}\right)', \quad \underline{g_{i_b}}^{\star} = \left[g\left(\mathbf{X}_i^{\star}; t_1^b\right), g\left(\mathbf{X}_i^{\star}; t_2^b\right), \ldots\right]',$$

$\underline{\overline{\Psi}_b} = \left[\overline{\Psi}(t_1^b), \overline{\Psi}(t_2^b), \ldots\right]'$ with $\overline{\Psi}(t) = \sum_{i=1}^n \overline{p}_i g\left(\mathbf{X}_i; t\right)$. Now define the stochastic process $\zeta_{n,b}^\star$ and note that,

$$\zeta_{n,b}^\star = \sqrt{n}\left(\underline{\hat{\Psi}_b}^\star - \underline{\overline{\Psi}_b}\right) \xrightarrow{p} \text{MVN}\left(\mathbf{0}_\infty, \Sigma_\infty\right) \tag{F.26}$$

conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}$. Then, letting $\underline{u}_b = \sqrt{n}\left(\underline{\tau_b} - \left(\underline{\eta_b} + \underline{\overline{\Psi}_b}\right)\right)$

$$T_n^\star\left(\underline{\eta_b}\right) = \min_{\tau \in ba_0\left(2^{[\underline{t},\overline{t}]\cap\mathbb{Q}}\right)_-} \left\{\left(\zeta_{n,b}^\star - \sqrt{n}\left(\underline{\tau_b} - \underline{\overline{\Psi}_b} + \underline{\eta_b}\right)\right)'\left(\hat{\Sigma}_\infty^\star\right)^{-1}\left(\zeta_{n,b}^\star - \sqrt{n}\left(\underline{\tau_b} - \underline{\overline{\Psi}_b} + \underline{\eta_b}\right)\right)\right\} \tag{F.27}$$

$$= \min_{\left\{\underline{u}_b \in l_\infty^\infty : \underline{u}_b \leq \sqrt{n}\left(\underline{\eta_b} + \underline{\overline{\Psi}_b}\right)\right\}} \left\{\left(\zeta_{n,b}^\star - \underline{u}_b\right)'\left(\hat{\Sigma}_\infty^\star\right)^{-1}\left(\zeta_{n,b}^\star - \underline{u}_b\right)\right\}. \tag{F.28}$$

By Lemma F.1, $\sqrt{n}\left(\underline{\eta_b} + \underline{\overline{\Psi}_b}\right) \xrightarrow{P} \mathbf{0}_\infty$ with probability 1 for n sufficiently large, which implies the following equality holds with probability tending to one:

$$T_n^\star\left(\eta_n\right) = \min_{\underline{u}_b \in l_{\infty,-}^\infty} \left\{\left(\zeta_{n,b}^\star - \underline{u}_b\right)'\left(\hat{\Sigma}_\infty^\star\right)^{-1}\left(\zeta_{n,b}^\star - \underline{u}_b\right)\right\}. \tag{F.29}$$

Since $\left(\hat{\Sigma}_\infty^\star(\eta)\right)^{-1} \xrightarrow{p} \Sigma_\infty^{-1}$ in the operator norm conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}(e_0)$, it follows that $T_n^\star\left(\eta_n\right) = T + o_P(1)$ conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}(e_0)$, where

$$T = \min_{\underline{u}_b \in l_{\infty,-}^\infty} \left\{\left(\zeta_b - \underline{u}_b\right)'\left(\Sigma_\infty\right)^{-1}\left(\zeta_b - \underline{u}_b\right)\right\}. \tag{F.30}$$

This is exactly the same asymptotic distribution of $\tilde{\mathcal{E}}_n$ and the result follows.

The case $\Delta\left(P_0\right) = \Delta_d\left(P_0\right) = \left\{t_1^b, \ldots, t_m^b\right\}, m \in \mathbb{Z}_+$ follows similar steps as in the previous case while keeping track of the fact that we now have a finite dimensional problem. □

F.2 Test Consistency

Lemma F.3. *Suppose Assumptions 2.1 and 4.1 hold. Then,*

1. $\left\{t \in [\underline{t}, \overline{t}] \cap \mathbb{Q} : E_{P_0}\left[g\left(\mathbf{X}; t\right)\right] > 0\right\} \subset \Delta\left(\tilde{\mathbf{p}}\right)$ *for large enough* n *with probability tending to one.*

2. $\|\tilde{\mu}\|_{l^1_{w_n}} = O_p(1)$.

3. $\tilde{\mathcal{E}}_n \to +\infty$ *as* $n \to +\infty$.

Proof. **Part 1**: The Lipschitz continuity of the moment functions (i.e. Assumption 2.1(i)) implies $\exists t \in [\underline{t}, \overline{t}] \cap \mathbb{Q}$ such that $E_{P_0}[g(\mathbf{X}; t)] > 0$. Without loss of generality, suppose that $t_a \in \left\{t \in [\underline{t}, \overline{t}] \cap \mathbb{Q} : E_{P_0}[g(\mathbf{X}; t)] > 0\right\}$. Recall that $w_n = |\Delta(\tilde{\mathbf{p}})|$. For ease of exposition, we will use the following notation: $\mu_j = \mu(\{t_j\})$. The the equality constraints are

$$\frac{1}{n}\sum_{i=1}^{n} \frac{g(\mathbf{X}_i; t_k)}{1 + \sum_{t_j \in \Delta(\tilde{\mathbf{p}})} \tilde{\mu}_j g(\mathbf{X}_i; t_j)} = 0 \quad \forall t_k \in \Delta(\tilde{\mathbf{p}}). \tag{F.31}$$

Evaluating at $k = a$ and at $\mu_j = 0\, \forall t_j \in \Delta(\tilde{\mathbf{p}})$ in (F.31) yields

$$\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{X}_i; t_a) = \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{X}_i; t_a) - \eta_n(t_a) + \eta_n(t_a). \tag{F.32}$$

The Law of the Iterated Logarithm implies $\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{X}_i; t_a) - \eta_n(t_a) > 0$ for large enough n with probability tending to one, and we also have $\eta_n(t_a) \geq 0$. Therefore, the optimal value of $\mu(\{t_a\})$ must increase for large enough n with probability tending to one. Hence, implies $t_a \in \text{supp}(\tilde{\mu})$ for large enough n with probability tending to one. Finally, by the FONCs (3.6), we have $\text{supp}(\tilde{\mu}) \subset \Delta(\tilde{\mathbf{p}})$ so that $t_a \in \Delta(\tilde{\mathbf{p}})$ for large enough n with probability tending to one.

Part 2: Using the same notation and steps from part 2 of Lemma E.3, we can arrive at equation (E.6), which is repeated here for convenience:

$$\|\tilde{\mu}\|_{l^1_{w_n}} \left(\theta' \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\underline{g_i}\left(\underline{g_i}\right)'}{1 + Y_i}\right)\theta\right) = \theta' \left(\frac{1}{n}\sum_{i=1}^{n} \underline{g_i}\right), \tag{F.33}$$

where $\tilde{\mu} = \|\tilde{\mu}\|_{l^1_{w_n}} \theta$ where $\theta \in \mathbb{R}^{w_n}_+$ such that $\sum_j \theta_j = 1$, and $Y_i = \underline{\tilde{\mu}}' \underline{g_i}$. Letting $\underline{E}$ denote the w_n vector of the population moments whose index set is supp $(\tilde{\mu})$, we have

$$\theta' \left(\frac{1}{n} \sum_{i=1}^{n} \underline{g_i} - \underline{E} \right) + \theta' \underline{E} = \|\tilde{\mu}\|_{l^1_{w_n}} \left(\theta' \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\underline{g_i} \left(\underline{g_i} \right)'}{1 + Y_i} - \frac{1}{n} \sum_{i=1}^{n} \underline{g_i} \left(\underline{g_i} \right)' \right) \theta \right) \tag{F.34}$$

$$+ \|\tilde{\mu}\|_{l^1_{w_n}} \theta' \left(\frac{1}{n} \sum_{i=1}^{n} \underline{g_i} \left(\underline{g_i} \right)' \right) \theta. \tag{F.35}$$

By the Law of Large Numbers, for large enough n (F.34) becomes

$$o_P(1) + \theta' \underline{E} = -\|\tilde{\mu}\|_{l^1_{w_n}} \left(\theta' \left(\sum_{i=1}^{n} \tilde{p}_i Y_i \underline{g_i} \left(\underline{g_i} \right)' \right) \theta \right) + \|\tilde{\mu}\|_{l^1_{w_n}} \left(\upsilon' \hat{\Omega}_\infty \left(P_0 \right) \upsilon \right), \tag{F.36}$$

(F.37)

where $\upsilon \in l^1_\infty$ is such that

$$\upsilon_j = \begin{cases} \theta_j, & \text{if } t_j \in \Delta \left(\tilde{\mathbf{p}} \right), \\ 0, & \text{if } t_j \in \Delta^+ \dot{\left(P_0 \right)} - \Delta \left(\tilde{\mathbf{p}} \right). \end{cases}$$

Since $0 <$ Plim $(\theta' \underline{E})$ by part 1 of this lemma, we must have

$$\text{Plim} \left(\upsilon' \Omega_\infty \left(P_0 \right) \upsilon \right) - \text{Plim} \left(\theta' \left(\sum_{i=1}^{n} \tilde{p}_i Y_i \underline{g_i} \left(\underline{g_i} \right)' \right) \theta \right) > 0, \tag{F.38}$$

Furthermore, $1 \geq$ Plim $(\theta' \underline{E})$ by condition (ii) of Assumption 2.1 implies

$$\|\tilde{\mu}\|_{l^1_{w_n}} \leq \frac{1}{\upsilon' \hat{\Omega}_\infty \left(P_0 \right) \upsilon - \theta' \left(\sum_{i=1}^{n} \tilde{p}_i Y_i \underline{g_i} \left(\underline{g_i} \right)' \right) \theta}. \tag{F.39}$$

for large enough n. Hence, $\|\tilde{\mu}\|_{l^1_{w_n}} = O_P(1)$.

Part 3: The ELR statistic can be expressed as the following maximization over the Lagrange multiplier measure:

$$\tilde{\mathcal{E}}_n = 2 \max_{\mu \in l^1_{w_n,+}} \sum_{i=1}^{n} \log\left(1 + \underline{\tilde{\mu}}' \underline{g_i}\right) \geq 2 \sum_{i=1}^{n} \log\left(1 + g\left(\mathbf{X}_i; t'\right)\right), \tag{F.40}$$

where the inequality follows from part 1 of this lemma with $t' \in \Delta\left(\tilde{\mathbf{p}}\right)$ such that $E_{P_0}\left[g\left(\mathbf{X}; t'\right)\right] > 0$. By condition (ii) of Assumption 2.1, the range of the moment functions is a subset of the interval $(-1, 1)$; therefore, we can apply the Taylor series expansion of $\log(1 + x)$ around $x = 0$ to the function $\log\left(1 + g\left(\mathbf{X}_i; t'\right)\right)$:

$$2 \sum_{i=1}^{n} \log\left(1 + g\left(\mathbf{X}_i; t'\right)\right) = 2 \sum_{i=1}^{n} g\left(\mathbf{X}_i; t'\right) - \sum_{i=1}^{n} g^2\left(\mathbf{X}_i; t'\right) + O_p\left(\sum_{i=1}^{n} g^3\left(\mathbf{X}_i; t'\right)\right). \tag{F.41}$$

Appropriately adding and subtracting $E_{P_0}\left[g\left(\mathbf{X}; t'\right)\right]$ into (F.41) yields

$$\begin{aligned} &2n\left(\frac{1}{n}\sum_{i=1}^{n}\left(g\left(\mathbf{X}_i; t'\right) - E_{P_0}\left[g\left(\mathbf{X}; t'\right)\right]\right)\right) + nE_{P_0}\left[g\left(\mathbf{X}; t'\right)\right] \\ &\quad + n\left(E_{P_0}\left[g\left(\mathbf{X}; t'\right)\right] - E_{P_0}\left[g^2\left(\mathbf{X}; t'\right)\right]\right) + \frac{n}{3}E_{P_0}\left[g^3\left(\mathbf{X}; t'\right)\right] + O_p\left(n^{1/2}\right) \\ &\quad - n\left(\frac{1}{n}\sum_{i=1}^{n}\left(g^2\left(\mathbf{X}_i; t'\right) - E_{P_0}\left[g^2\left(\mathbf{X}; t'\right)\right]\right)\right) \end{aligned} \tag{F.42}$$

, which equals

$$\begin{aligned} &O_p\left(n^{1/2}\right) + \frac{2n}{3}E_{P_0}\left[g\left(\mathbf{X}; t'\right)\right] + \frac{n}{3}\left(E_{P_0}\left[g\left(\mathbf{X}; t'\right)\right] + E_{P_0}\left[g^3\left(\mathbf{X}; t'\right)\right]\right) \\ &\quad + n\left(E_{P_0}\left[g\left(\mathbf{X}; t'\right)\right] - E_{P_0}\left[g^2\left(\mathbf{X}; t'\right)\right]\right). \end{aligned} \tag{F.43}$$

Since $-1 < g \leq -1$, we have

$$\left(E_{P_0}\left[g\left(\mathbf{X}; t'\right)\right] + E_{P_0}\left[g^3\left(\mathbf{X}; t'\right)\right]\right), \left(E_{P_0}\left[g\left(\mathbf{X}; t'\right)\right] - E_{P_0}\left[g^2\left(\mathbf{X}; t'\right)\right]\right) \geq 0,$$

and therefore

$$\frac{2n}{3} E_{P_0}\left[g\left(\mathbf{X}; t'\right)\right] + \frac{n}{3}\left(E_{P_0}\left[g\left(\mathbf{X}; t'\right)\right] + E_{P_0}\left[g^3\left(\mathbf{X}; t'\right)\right]\right) \to +\infty \quad \text{as} \quad n \to +\infty, \tag{F.44}$$

at the n-th rate, which dominates the $O_p\left(n^{1/2}\right)$ terms in (F.43). Using this result in the inequality (F.40), it follows that $\tilde{\mathcal{E}}_n \to +\infty$ as $n \to +\infty$. □

From Lemma F.3, we have the following result regarding the large sample behavior of $\check{\mathcal{E}}_n$.

Lemma F.4. *Suppose Assumptions 2.1 and 4.1 hold. If* $\min\left\{\frac{L}{N(n)}, \alpha(n)\right\} \to 0$ *as* $n \to +\infty$, *then* $\check{\mathcal{E}}_n \to +\infty$.

Proof. The approximate ELR statistic (3.15) can be decomposed as in (C.22), which is repeated now for convenience:

$$\acute{\mathcal{E}}_n = \tilde{\mathcal{E}}_n + 2\left\{\sum_{i=1}^{n}\left(\log\left(\tilde{p}_i\right) - \log\left(\acute{p}_i\right)\right)\right\}. \tag{F.45}$$

Furthermore, in the proof of Theorem 3.1

$$\left|2\left\{\sum_{i=1}^{n}\left(\log\left(\tilde{p}_i\right) - \log\left(\acute{p}_i\right)\right)\right\}\right| \leq 2\|\tilde{\mu}\|_{l^1_{w_n}} \min\left\{\frac{L}{N(n)}, \alpha(n)\right\}. \tag{F.46}$$

As part 2 of Lemma F.3 indicates $\|\tilde{\mu}\|_{l^1_{w_n}} = O_P(1)$, the desired result follows from letting $\min\left\{\frac{L}{N(n)}, \alpha(n)\right\} \to 0$ as $n \to +\infty$. □

The next result describes the behavior of $\overline{\Psi}(t) = \sum_{i=1}^{n} \overline{p}_i g\left(\mathbf{X}_i; t\right)$, where $\left(\overline{p}_1, \ldots, \overline{p}_n\right)$ is the bootstrap DGP described in Section 4, under the alternative.

Lemma F.5. *Let* $\overline{\Psi}(t) = \sum_{i=1}^{n} \overline{p}_i g\left(\mathbf{X}_i; t\right)$, *where* $\left(\overline{p}_1, \ldots, \overline{p}_n\right)$ *is the bootstrap DGP described in Section 4. Suppose* P_0 *satisfies the alternative hypothesis, then the following two statements hold.*

1. *If* $t \in [\underline{t}, \overline{t}] \cap \mathbb{Q}$ *and* $E_{P_0}\left[g\left(\mathbf{X}; t\right)\right] < 0$, *then* $\overline{\Psi}(t) < -\eta_n(t)$ *for large* n *with probability one.*

2. *If* $t \in \Delta_+(P_0)$ *and* $t \in \mathcal{T}_{N(n)}$ *for large enough* n, *then* $\overline{\Psi}(t) = -\eta_n(t)$ *for large* n *with probability one.*

3. *If* $t \in \Delta_+(P_0)$ *and* $t \notin \mathcal{T}_{N(n)}$ $\forall n$, *then* $\overline{\Psi}(t) = -\eta_n(t)$ *for large* n *with probability one.*

Proof. The proof of this lemma follows identical steps as those in Lemma F.1; therefore, we omit it for brevity. □

The next result describes the behavior of the bootstrapped Lagrange multiplier vector that arises in the modified Exchange Algorithm.

Lemma F.6. *Suppose Assumptions 2.1 and 4.1 hold. Let* $\overline{P}_n$ *be the bootstrap DGP described in Section 4, and let* $\{\mathbf{X}_i^\star\}_{i=1}^n$ *be IID* $\overline{P}_n$. *Also let* $\overline{\mathbf{p}}^\star$ *denote the solution of the modified Exchange algorithm, but with the bootstrap sample instead of the data. The following statements hold.*

1. $\Delta(\overline{\mathbf{p}}^\star) \subset \Delta_+(P_0)$ *for large* n, *where* $\Delta(\overline{\mathbf{p}}^\star)$ *denotes the index set of binding moments from the modified Exchange Algorithm that uses the bootstrap sample.*

2. *Let* $\acute{\mu}^\star$ *be the optimal value of the Lagrange multiplier vector on the inequality constraints in the modified Exchange Algorithm that uses the bootstrap sample. Also let* $w_n^\star = |\Delta(\overline{\mathbf{p}}^\star)|$. *Then* $\|\acute{\mu}^\star\|_{l^1_{w_n^\star}} = \sum_{t\in\Delta(\overline{\mathbf{p}}^\star)} \acute{\mu}_t^\star = o_P(1)\overline{P}_n -$ *a.e. at the* $\sqrt{n}-$ *rate.*

Proof. **Part 1**. The proof of this part follows identical steps as those in part 1 of Lemma F.2, but with $\Delta(P_0)$ and $\Delta(\tilde{\mathbf{p}}^\star)$, replaced by $\Delta_+(P_0)$ and $\Delta(\overline{\mathbf{p}}^\star)$ respectively, and by using part 1 of Lemma F.5.

Part 2. The proof of this part follows identical steps as those in part 2 of Lemma F.2, but with $\Delta(\tilde{\mathbf{p}}^\star)$ replaced by $\Delta(\overline{\mathbf{p}}^\star)$. Now we focus on the convergence of
$\hat{\Sigma}^\star_{w_n^\star} = \frac{1}{n}\sum_{i=1}^n \left(\underline{g_i^\star} + \underline{\eta_n}\right)\left(\underline{g_i^\star} + \underline{\eta_n}\right)'$ to complete the proof. Let θ be a non-negative unit vector in $l^1_{w_n^\star}$, and let $\upsilon \in l^1_\infty$ be such that

$$\upsilon_j = \begin{cases} \theta_j, & \text{if } t_j \in \Delta(\overline{\mathbf{p}}^\star), \\ 0, & \text{if } t_j \in \Delta_+(P_0) - \Delta(\overline{\mathbf{p}}^\star). \end{cases}$$

Then, $\theta'\hat{\Sigma}^{\star}_{w_n^{\star}}\theta = \upsilon'\hat{\Sigma}^{\star}_{\infty}\upsilon$. Now we will show that $\upsilon'\hat{\Sigma}^{\star}_{\infty}\upsilon > 0\,\overline{P}_n -$ a.e.. By the LLN, it follows that $\hat{\Sigma}^{\star}_{\infty} \xrightarrow{P} \lim_{n\to+\infty} E_{\overline{P}_n}\left[\hat{\Sigma}^{\star}_{\infty}\right]\,\overline{P}_n -$ a.e.. Therefore, we can focus on the large sample behavior of $\upsilon' E_{\overline{P}_n}\left[\hat{\Sigma}^{\star}_{\infty}\right]\upsilon$. conditional on the data.

Conditionally on the data,

$$E_{\overline{P}_n}\left[\hat{\Sigma}^{\star}_{\infty}\right] = \sum_{i=1}^{n}\overline{p}_i\left(\underline{g_i}+\underline{\eta_n}\right)\left(\underline{g_i}+\underline{\eta_n}\right)',$$

where $\left(\underline{g_i}+\underline{\eta_n}\right)$ is now an infinite vector of the moment functions are indexed by $\Delta_+\dot{}\,(P_0)$. Then, adding and subtracting as follows

$$\sum_{i=1}^{n}\overline{p}_i\left(\underline{g_i}+\underline{\eta_n}\right)\left(\underline{g_i}+\underline{\eta_n}\right)' = \sum_{i=1}^{n}\left(\overline{p}_i-\frac{1}{n}\right)\left(\underline{g_i}+\underline{\eta_n}\right)\left(\underline{g_i}+\underline{\eta_n}\right)' + \frac{1}{n}\sum_{i=1}^{n}\left(\underline{g_i}+\underline{\eta_n}\right)\left(\underline{g_i}+\underline{\eta_n}\right)',$$

makes the right side asymptotically equivalent to

$$\begin{aligned}\sum_{i=1}^{n}\left(\overline{p}_i-\frac{1}{n}\right)\left(\underline{g_i}+\underline{\eta_n}\right)\left(\underline{g_i}+\underline{\eta_n}\right)' + \frac{1}{n}\sum_{i=1}^{n}\underline{g_i g_i}' &= \sum_{i=1}^{n}\left(\overline{p}_i-\frac{1}{n}\right)\left(\underline{g_i}+\underline{\eta_n}\right)\left(\underline{g_i}+\underline{\eta_n}\right)' \\ &\quad + o_P(1) + \Omega\left(P_0\right).\end{aligned} \tag{F.47}$$

Now we show that $A_n = \sum_{i=1}^{n}\left(\overline{p}_i-\frac{1}{n}\right)\left(\underline{g_i}+\underline{\eta_n}\right)\left(\underline{g_i}+\underline{\eta_n}\right)'$ is $o_P(1)$. It is equal to

$$-\sum_{i=1}^{n}\overline{p}_i\left(\sum_{t\in\Delta(\overline{\mathbf{p}})}\overline{\mu}_t\left(g\left(\mathbf{X}_i;t\right)+\eta_n(t)\right)\right)\left(\underline{g_i}+\underline{\eta_n}\right)\left(\underline{g_i}+\underline{\eta_n}\right)', \tag{F.48}$$

which can be re-arranged into

$$\sum_{t\in\Delta(\overline{\mathbf{p}})}\overline{\mu}_t\left((-)\sum_{i=1}^{n}\overline{p}_i\left(\left(g\left(\mathbf{X}_i;t\right)+\eta_n(t)\right)\left(\underline{g_i}+\underline{\eta_n}\right)\left(\underline{g_i}+\underline{\eta_n}\right)'\right)\right). \tag{F.49}$$

Therefore, $|A_n\|$ is bounded from above by

$$\|\acute{\mu}\|_{l^1_{w_n}}(-)\sum_{i=1}^{n}\overline{p}_i\left(g\left(\mathbf{X}_i;t\right)+\eta_n(t)\right), \tag{F.50}$$

where $w_n = |\Delta(\overline{\mathbf{p}})|$, and $\left(\underline{g_i} + \underline{\eta_n}\right)\left(\underline{g_i} + \underline{\eta_n}\right)'$ is non-negative for each i (by Property C.1) and bounded from above by 1 (part 2 of Assumption 2.1). Now parts 2 and 3 of Lemma F.5 show $(-)\sum_{i=1}^{n} \overline{p}_i \left(g\left(\mathbf{X}_i; t\right) + \eta_n(t)\right)$ tends to zero for large enough n with probability tending to one. Finally, by following similar arguments as in the proof of part 2 of LemmaF.3, it follows that $\|\acute{\mu}\|_{l^1_{w_n}} = O_P(1)$. Therefore, $A_n = o_P(1)$.

Putting these two parts together, implies $\upsilon' E_{\overline{P}_n}\left[\hat{\Sigma}^{\star}_{\infty}\right]\upsilon = o_P(1) + \upsilon'\Omega\left(P_0\right)\upsilon + \upsilon' A_n \upsilon$, which is equal to $\upsilon'\Omega\left(P_0\right)\upsilon + o_P(1)$. Finally, $\upsilon'\Omega\left(P_0\right)\upsilon > 0$ follows from part 1 of Assumption 4.1. This completes the proof. $\square$

G Technical Uniform Weak Convergence Results

G.1 Asymptotic Null Distribution

In this section, we present large sample results that are used to prove that the statistic T_n, defined in (C.20), is uniformly weakly convergent to the QLR statistic in Theorem 3.1. The first result is an immediate consequence of Lemma B.1.

Corollary G.1. *Let $\Sigma_w(P)$ be the covariance matrix defined in Section 2.4, and let $\hat{\Sigma}_w(P)$ be its sample analogue estimator. Then,*

1. $\|\hat{\Sigma}_w(P) - \Sigma_w(P)\| = o_P(1)$ *uniformly in $\mathcal{M}$, where $\|\cdot\|$ is the operator norm (2.9).*

2. *For every $e_0 \in \mathbb{R}_+$, $\|\hat{\Sigma}_w^{-1}(P) - \Sigma_w^{-1}(P)\| = o_P(1)$ uniformly in $\mathcal{M}(e_0)$, where $\|\cdot\|$ is the operator norm (2.10).*

Proof. **Part 1:** As $\mathcal{G}$ being uniform Donsker implies that it is also uniform Glivenko-Cantelli, we have $\|\hat{\Sigma}_w(P) - \Sigma_w(P)\| =\leq \sup_{i,j}\left|\hat{\Sigma}_{w,i,j}(P) - \Sigma_{w,i,j}(P)\right| = o_P(1)$ uniformly in $\mathcal{M}$, since the uniform Donsker property is preserved under the multiplication transform and that the moment functions are uniformly bounded (Assumption 2.1).

Part 2: The result follows directly from Part 1 of this corollary upon realizing that

$$\|\hat{\Sigma}_w^{-1}(P) - \Sigma_w^{-1}(P)\| \leq \|\Sigma_w^{-1}\|\|\hat{\Sigma}_w^{-1}\| \left\|\hat{\Sigma}_w - \Sigma_w\right\|, \tag{G.1}$$

which is less than or equal to $e_0^2 o_P(1)$ uniformly in $\mathcal{M}(e_0)$ for large enough n. $\square$

Let T_P be the Gaussian QLR statistic in Theorem 3.1:

$$T_P = \min_{\underline{u_b} \in l^\infty_{w,-}} \left\{ \left[\mathbf{G}_w - \underline{u_b}\right]' \Sigma_w^{-1}(P) \left[\mathbf{G}_w - \underline{u_b}\right] \right\}, \tag{G.2}$$

where $\mathbf{G}_w \sim$ MVN $(\mathbf{0}_w, \Sigma_w(P))$, and for convenience, we repeat the definition of T_n in (C.20)

$$T_n = \min_{\underline{u_b} \in l^\infty_{w,-}} \left\{ \left[\sqrt{n}\underline{\hat{\Psi}_b} - \underline{u_b}\right]' \hat{\Sigma}_w^{-1} \left[\sqrt{n}\underline{\hat{\Psi}_b} - \underline{u_b}\right] \right\}.$$

Our objective is to show that for each $e_0 \in \mathbb{R}_+$, $T_n \rightsquigarrow T_P$ uniformly in $\mathcal{M}(e_0)$.

T_P and T_n are optimal values of quadratic optimization problems whose Lagrangians are respectively

$$L\left(\underline{u_b}, \lambda\right) = \left[\mathbf{G}_w - \underline{u_b}\right]' \Sigma_w^{-1}(P) \left[\mathbf{G}_w - \underline{u_b}\right] + \lambda' \underline{u_b}, \tag{G.3}$$

$$L\left(\underline{u_b}, \gamma\right) = \left[\sqrt{n}\underline{\hat{\Psi}_b} - \underline{u_b}\right]' \hat{\Sigma}_w^{-1}(P) \left[\sqrt{n}\underline{\hat{\Psi}_b} - \underline{u_b}\right] + \gamma' \underline{u_b},, \tag{G.4}$$

where λ and γ are the non-negative Lagrange multiplier vector with dimension w. The FONCs for these optimization problems are

$$\underline{u_b} = -\frac{1}{2}\Sigma_w(P)\lambda + \mathbf{G}_w, \quad \underline{u_b} \in l^\infty_{w,-}, \quad \text{and} \quad \underline{u_b}\left(t_i^b\right)\lambda\left(t_i^b\right) = 0 \; \forall t_i^b \in \Delta\left(\dot{P}\right) \tag{G.5}$$

$$\underline{u_b} = -\frac{1}{2}\hat{\Sigma}_w(P)\gamma + \sqrt{n}\underline{\hat{\Psi}_b}, \quad \underline{u_b} \in l^\infty_{w,-}, \quad \text{and} \quad \underline{u_b}\left(t_i^b\right)\gamma\left(t_i^b\right) = 0 \; \forall t_i^b \in \Delta\left(\dot{P}\right). \tag{G.6}$$

Let $\lambda^\star$ and $\gamma^\star$ denote the optimal values. Then, using the complementary-slackness conditions we have $\lambda'\mathbf{G}_w = \frac{1}{2}\lambda'\Sigma_w\lambda$ and $\gamma'\underline{\hat{\Psi}_b} = \frac{1}{2}\gamma'\hat{\Sigma}_w\gamma$, which implies

$$\|\mathbf{G}_w\|_{l_w^\infty} \|\lambda^\star\|_{l_w^1} \geq (\lambda^\star)'\,\mathbf{G}_w = \frac{1}{2}(\lambda^\star)'\,\Sigma_w\lambda^\star \geq \frac{1}{2}\inf_{i,j}\Sigma_{w,i,j}(P)\|\lambda^\star\|_{l_w^1}^2, \quad \text{and} \tag{G.7}$$

$$\left\|\sqrt{n}\underline{\hat{\Psi}_b}\right\|_{l_w^\infty}\|\gamma^\star\|_{l_w^1} \geq (\gamma^\star)'\,\hat{\Psi}_b = \frac{1}{2}(\gamma^\star)'\,\hat{\Sigma}_w\gamma^\star \geq \frac{1}{2}\inf_{i,j}\hat{\Sigma}_{w,i,j}(P)\|\gamma^\star\|_{l_w^1}^2. \tag{G.8}$$

Therefore,

$$\|\lambda^\star\|_{l_w^1} \leq \frac{2\|\mathbf{G}_w\|_{l_w^\infty}}{\inf_{i,j}\Sigma_{w,i,j}(P)} \leq \frac{2\|\mathbf{G}_w\|_{l_w^\infty}}{c}, \quad \text{and} \quad \|\gamma^\star\|_{l_w^1} \leq \frac{2\left\|\sqrt{n}\underline{\hat{\Psi}_b}\right\|_{l_w^\infty}}{\inf_{i,j}\hat{\Sigma}_{w,i,j}(P)}, \tag{G.9}$$

where c is the constant in Definition 2.1. The first result concerning the Lagrange multipliers is the following.

Lemma G.1. *1. $\|\lambda^\star\|_{l_w^1} = O_P(1)$ and $\|\gamma^\star\|_{l_w^1} = O_P(1)$, uniformly in $\mathcal{M}$.*

2. $\gamma^\star \rightsquigarrow \lambda^\star$ in $\|\cdot\|_{l_w^1}$, uniformly in $\mathcal{M}$.

Proof. **Part 1:** To show that $\|\lambda^\star\|_{l_w^1} = O_P(1)$ uniformly in $\mathcal{M}$, we need to prove $\|\mathbf{G}_w\|_{l_w^\infty} = O_P(1)$ uniformly in $\mathcal{M}$. Let $\mathcal{G}(\delta)$ be a δ-net for $(\mathcal{G}, \rho_P)$ with $k = |\mathcal{G}(\delta)|$ independent of $P \in \mathcal{M}$, where

$$\rho_P^2(t, t') = \mathrm{Var}_P\left(g(\mathbf{X}; t) - g(\mathbf{X}; t')\right) \tag{G.10}$$

and X has distribution P. Additionally, let Π_δ denote a map from $\mathcal{G}$ to a nearest point in $\mathcal{G}(\delta)$. Then,

$$\begin{aligned}\|\mathbf{G}_w\|_{l_w^\infty} &\leq \sup_{t\in[\underline{t},\overline{t}]}|G(t)| \\ &\leq \sup_{t\in[\underline{t},\overline{t}]}|G(t) - (G\circ\Pi_\delta)(t)| + \|G(t)\|_{\mathcal{G}(\delta)} \leq \sup_{\rho_P(t,t')<\delta}|G(t) - G(t')| + \|G(t)\|_{\mathcal{G}(\delta)}.\end{aligned} \tag{G.11}$$

We have $\|G(t)\|_{\mathcal{G}(\delta)} = O_P(1)$ uniformly in $\mathcal{M}$ since it is the maximum over at most k Gaussian random variables, each having zero mean and variance not exceeding unity. The uniform pre-

Gaussianity of $\mathcal{G}$ implies for every $\epsilon > 0$,

$$\lim_{\delta \downarrow 0} \sup_{P \in \mathcal{M}} \text{Prob}_P \left[\sup_{\rho_P(t,t') < \delta} |G(t) - G(t')| > \epsilon \right] = 0. \tag{G.12}$$

Therefore, letting $\delta \downarrow 0$, the inequality (G.11) implies $\|\mathbf{G}_w\|_{l_w^\infty} = O_P(1)$ uniformly in $\mathcal{M}$.

Now we will show that $\|\gamma^\star\|_{l_w^1} = O_P(1)$ uniformly in $\mathcal{M}$. Lemma B.1 says that $\mathcal{G}$ is Donsker and pre-Gaussian uniformly in $\mathcal{M}$. Since the envelope function of $\mathcal{G}$ is the constant function equal to one, it satisfies $\lim_{\nu \to \infty} \sup_{P \in \mathcal{M}} \text{Prob}_P [1 \geq \nu] = 0$. Therefore, Theorem 2.1 of Sheehy and Wellner (1992) implies the empirical process $\frac{1}{\sqrt{n}} \sum_{i=1}^n (g(\mathbf{X}_i; t) - E_P [g(\mathbf{X}; t)])$ admits a weak Gaussian approximation uniformly in $\mathcal{M}$. This means that there exists a sequence of G coherent stochastic processes, $\{G^{(1)}, G^{(2)}, \ldots\}$ such that for every $\epsilon > 0$

$$\lim_{n \to \infty} \sup_{P \in \mathcal{M}} \text{Prob}_P \left[\sup_{t \in [\underline{t}, \bar{t}]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (g(\mathbf{X}_i; t) - E_P [g(\mathbf{X}; t)]) - G^{(n)}(t) \right| > \epsilon \right] = 0. \tag{G.13}$$

Using

$$\begin{aligned}
\left\| \sqrt{n} \underline{\hat{\Psi}_b} \right\|_{l_w^\infty} &= \left\| \sqrt{n} \underline{\hat{\Psi}_b} - \mathbf{G}_w + \mathbf{G}_w \right\|_{l_w^\infty} && \text{(G.14)} \\
&\leq \left\| \sqrt{n} \underline{\hat{\Psi}_b} - \mathbf{G}_w \right\|_{l_w^\infty} + \|\mathbf{G}_w\|_{l_w^\infty} && \text{(G.15)} \\
&= \left\| \sqrt{n} \underline{\hat{\Psi}_b} - \mathbf{G}_w \right\|_{l_w^\infty} + O_P(1) \quad \text{uniformly in} \quad \mathcal{M} && \text{(G.16)} \\
&\leq \left\| \sqrt{n} \underline{\hat{\Psi}_b} - \mathbf{G}_w^{(n)} \right\|_{l_w^\infty} + \left\| \mathbf{G}_w^{(n)} - \mathbf{G}_w \right\|_{l_w^\infty} + O_P(1) \quad \text{uniformly in} \quad \mathcal{M} && \text{(G.17)} \\
&= o_P(1) + O_P(1) \quad \text{uniformly in} \quad \mathcal{M}. && \text{(G.18)}
\end{aligned}$$

Furthermore, for large n, $\inf_{i,j} \hat{\Sigma}_{w,i,j}(P) > c$ since $\hat{\Sigma}_w(P)$ converges to $\Sigma_w(P)$ in the operator norm (2.9), uniformly in $\mathcal{M}$.

Part 2: We prove this part by showing that the FONCs and second order conditions of the dual problems that define T_P and T_n are the same in distribution. The dual optimization problems are the following:

$$\max_{\lambda\in l^1_{w,+}}\left\{-\frac{1}{4}\lambda'\Sigma_w(P)\lambda-\lambda'\mathbf{G}_w\right\},\quad\text{and}\quad\max_{\gamma\in l^1_{w,+}}\left\{-\frac{1}{4}\gamma'\hat{\Sigma}_w(P)\lambda-\lambda'\sqrt{n}\underline{\hat{\Psi}_b}\right\},\tag{G.19}$$

and their respective FONCs and second order conditions are

$$-\frac{1}{2}\lambda'\Sigma_w(P)-\mathbf{G}_w\geq\mathbf{0}_w\quad-\lambda'\Sigma_w(P)\lambda\leq 0,\quad\text{and}\tag{G.20}$$

$$-\frac{1}{2}\gamma'\hat{\Sigma}_w(P)-\sqrt{n}\underline{\hat{\Psi}_b}\geq\mathbf{0}_w\quad-\gamma'\hat{\Sigma}_w(P)\gamma\leq 0.\tag{G.21}$$

Lemma B.1 and Part 1 of Corollary G.1 imply that these conditions are asymptotically the same in distribution, uniformly in $\mathcal{M}$. Since the objective functions in these optimization problems are strictly concave in the Lagrange multipliers, these optimization problems have a unique solution, and therefore, $\|\gamma^\star-\lambda^\star\|_{l^1_w}\rightsquigarrow 0$ uniformly in $\mathcal{M}$. □

We can now prove the main result in this section.

Proposition G.1. *For each $e_0\in\mathbb{R}_+$, we have $T_n\rightsquigarrow T_P$ uniformly in $\mathcal{M}(e_0)$.*

Proof. The proof proceeds by the direct method. Let $\underline{\hat{\Psi}_b}$ be as in equation (C.7). Let $\underline{u_b}$ and $\underline{\hat{u}_b}$ are respectively the solutions of the minimization problems (G.2) and (C.20). Then, difference T_P-T_n is given by:

$$\begin{aligned}
T_P-T_n&=\left[\mathbf{G}_w-\underline{u_b}\right]'\Sigma_w^{-1}(P)\left[\mathbf{G}_w-\underline{u_b}\right]-\left[\sqrt{n}\underline{\hat{\Psi}_b}-\underline{\hat{u}_b}\right]'\hat{\Sigma}_w^{-1}\left[\sqrt{n}\underline{\hat{\Psi}_b}-\underline{\hat{u}_b}\right]\\
&=\left[\mathbf{G}_w-\underline{u_b}\right]'\Sigma_w^{-1}(P)\left[\mathbf{G}_w-\underline{u_b}\right]-\left[\sqrt{n}\underline{\hat{\Psi}_b}-\underline{\hat{u}_b}\right]'\left(\hat{\Sigma}_w^{-1}-\Sigma_w^{-1}+\Sigma_w^{-1}\right)\left[\sqrt{n}\underline{\hat{\Psi}_b}-\underline{\hat{u}_b}\right]\\
&=\left[\mathbf{G}_w-\sqrt{n}\underline{\hat{\Psi}_b}+\underline{\hat{u}_b}-\underline{u_b}\right]'\Sigma_w^{-1}(P)\left[\mathbf{G}_w-\sqrt{n}\underline{\hat{\Psi}_b}+\underline{\hat{u}_b}-\underline{u_b}\right]\qquad\text{(G.22)}\\
&\quad-\left[\sqrt{n}\underline{\hat{\Psi}_b}-\underline{\hat{u}_b}\right]'\left(\hat{\Sigma}_w^{-1}-\Sigma_w^{-1}\right)\left[\sqrt{n}\underline{\hat{\Psi}_b}-\underline{\hat{u}_b}\right].\qquad\text{(G.23)}
\end{aligned}$$

First, we prove that (G.23) is $o_P(1)$ uniformly in $\mathcal{M}(e_0)$. In modulus, (G.23) is

$$\left|\left[\sqrt{n}\underline{\hat{\Psi}_b} - \underline{\hat{u}_b}\right]'\left(\hat{\Sigma}_w^{-1} - \Sigma_w^{-1}\right)\left[\sqrt{n}\underline{\hat{\Psi}_b} - \underline{\hat{u}_b}\right]\right| \leq \left\|\sqrt{n}\underline{\hat{\Psi}_b} - \underline{\hat{u}_b}\right\|_{l_w^\infty}^2 \|\Sigma_w^{-1}\|\|\hat{\Sigma}_w^{-1}\|\left\|\hat{\Sigma}_w - \Sigma_w\right\| \tag{G.24}$$

$$= \frac{1}{4}\left\|\hat{\Sigma}_w(P)\gamma^\star\right\|_{l_w^\infty}^2 \|\Sigma_w^{-1}\|\|\hat{\Sigma}_w^{-1}\|\left\|\hat{\Sigma}_w - \Sigma_w\right\| \tag{G.25}$$

$$\leq \frac{1}{4}\left\|\hat{\Sigma}_w(P)\right\|^2 \|\gamma^\star\|_{l_w^1}^2 \|\|\Sigma_w^{-1}\|\|\hat{\Sigma}_w^{-1}\|\left\|\hat{\Sigma}_w - \Sigma_w\right\|. \tag{G.26}$$

Now, by Lemmas B.1 and G.1, and for large n, (G.26) is bounded above by

$$\frac{e_0^2}{4}O_P(1)o_P(1) \quad \text{uniformly in} \quad \mathcal{M}(e_0), \tag{G.27}$$

which implies the desired result.

Now we will show that (G.22) converges to zero in distribution, uniformly in $\mathcal{M}(e_0)$. First, note that (G.22) is equal to

$$\frac{1}{4}\left[\hat{\Sigma}_w(P)\gamma^\star - \Sigma_w(P)\lambda^\star\right]'\Sigma_w^{-1}\left[\hat{\Sigma}_w(P)\gamma^\star - \Sigma_w(P)\lambda^\star\right]. \tag{G.28}$$

Furthermore, $\hat{\Sigma}_w(P)\gamma^\star - \Sigma_w(P)\lambda^\star = \left(\hat{\Sigma}_w(P) - \Sigma_w(P)\right)\gamma^\star - \Sigma_w(P)\left(\lambda^\star - \gamma^\star\right)$. Corollary G.1 and Part 1 of Lemma G.1 implies $\left(\hat{\Sigma}_w(P) - \Sigma_w(P)\right)\gamma^\star = o_P(1)$ in $\|\cdot\|_{l_w^\infty}$, uniformly in $\mathcal{M}(e_0)$, and Part 2 of Lemma G.1 implies that $-\Sigma_w^{-1}\left[\Sigma_w(P)\left(\lambda^\star - \gamma^\star\right)\right] = -\left(\lambda^\star - \gamma^\star\right) \rightsquigarrow 0$ in $\|\cdot\|_{l_w^1}$, uniformly in $\mathcal{M}(e_0)$.

Hence, putting these two parts together implies $T_P - T_n \rightsquigarrow 0$ uniformly in $\mathcal{M}(e_0)$, which completes the proof. ☐

G.2 Bootstrap Validity

Lemma G.2. *Suppose $P_0 \in \mathcal{M}$, and let $\overline{w}_n = |\Delta(\overline{\mathbf{p}})|$ where $\overline{\mathbf{p}}$ is the bootstrap DGP from Section 4. Additionally, let $\overline{\mu}$ denote the Lagrange multiplier vector from the modified exchange algorithm with optimization problem (4.1). Then, $\|\overline{\mu}\|_{l_{\overline{w}_n}^1} = o_P(1)$ uniformly in $\mathcal{M}$.*

Proof. Firstly, a consequence of Lemma F.1 is that $\Delta(\overline{\mathbf{p}}) \subset \Delta(P_0)$ for large n. Now the proof follows the same steps as in the proof of part 2 of Lemma E.3. Let $\underline{\mu} = \|\underline{\mu}\|_{l^1_{\overline{w}_n}} \theta$ where $\theta \in \mathbb{R}_+^{\overline{w}_n}$ is such that $\sum_j \theta_j = 1$. Using the same vector notation, we have

$$\theta' \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\underline{g_i} + \underline{\eta}}{1 + \underline{\mu}' \left(\underline{g_i} + \underline{\eta} \right)} \right) = 0. \tag{G.29}$$

Let $Y_i = \underline{\mu}' \left(\underline{g_i} + \underline{\eta} \right)$ and use $\frac{1}{1+Y_i} = 1 - \frac{Y_i}{1+Y_i}$ to expand (G.29) as follows:

$$\|\underline{\mu}\|_{l^1_{\overline{w}_n}} \left(\theta' \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\left(\underline{g_i} + \underline{\eta} \right) \left(\underline{g_i} + \underline{\eta} \right)'}{1 + Y_i} \right) \theta \right) = \theta' \left(\frac{1}{n} \sum_{i=1}^{n} \left(\underline{g_i} + \underline{\eta} \right) \right). \tag{G.30}$$

The sample analogue estimator of $\Sigma_{\overline{w}_n}(P_0)$ is $\hat{\Sigma}_{\overline{w}_n} \left(\underline{\eta} \right) = \frac{1}{n} \sum_{i=1}^{n} \left(\underline{g_i} + \underline{\eta} \right) \left(\underline{g_i} + \underline{\eta} \right)'$. Since $1 + Y_i > 0 \quad \forall i$,

$$\begin{aligned} \|\underline{\mu}\|_{l^1_{\overline{w}_n}} \left(\theta' \hat{\Sigma}_{\overline{w}_n} \left(\underline{\eta} \right) \theta \right) &\leq \|\underline{\mu}\|_{l^1_{\overline{w}_n}} \left(\theta' \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\left(\underline{g_i} + \underline{\eta} \right) \left(\underline{g_i} + \underline{\eta} \right)'}{1 + Y_i} \right) \theta \right) \left(1 + \max_{i=1,\ldots,n} |Y_i| \right) \\ &\leq \|\underline{\mu}\|_{l^1_{\overline{w}_n}} \left(\theta' \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\left(\underline{g_i} + \underline{\eta} \right) \left(\underline{g_i} + \underline{\eta} \right)'}{1 + Y_i} \right) \theta \right) \left(1 + \|\underline{\mu}\|_{l^1_{\overline{w}_n}} \right). \end{aligned} \tag{G.31}$$

Therefore, we have the following inequality

$$\|\underline{\mu}\|_{l^1_{\overline{w}_n}} \left(\theta' \hat{\Sigma}_{\overline{w}_n} \left(\underline{\eta} \right) \theta - \theta' \left(\frac{1}{n} \sum_{i=1}^{n} \left(\underline{g_i} + \underline{\eta} \right) \right) \right) \leq \theta' \left(\frac{1}{n} \sum_{i=1}^{n} \left(\underline{g_i} + \underline{\eta} \right) \right). \tag{G.32}$$

Now by arguments similar to those in the proof of part 2 of Lemma E.3, we have

$$\theta' \left(\frac{1}{n} \sum_{i=1}^{n} \left(\underline{g_i} + \underline{\eta} \right) \right) = o_P(1) \quad \text{uniformly in} \quad \mathcal{M}, \tag{G.33}$$

$$\theta' \hat{\Sigma}_{\overline{w}_n} \left(\underline{\eta} \right) \theta > c, \quad \text{with probability approaching 1 uniformly in} \quad \mathcal{M}, \tag{G.34}$$

where we made use of the following expansion

$$\hat{\Sigma}_{\overline{w}_n} \left(\underline{\eta} \right) = \hat{\Sigma}_{\overline{w}_n} + \underline{\eta \eta'} + \frac{1}{n} \sum_{i=1}^{n} \underline{g_i \eta'} + \frac{1}{n} \sum_{i=1}^{n} \underline{\eta g_i}', \tag{G.35}$$

and that $\underline{\eta}$ tends to zero in probability uniformly in $\mathcal{M}$. This implies $\|\overline{\mu}\|_{l^1_{\overline{w}_n}} = o_P(1)$ uniformly in $\mathcal{M}$, which is the desired result. $\square$

Let $\mathcal{A}_n$ denotes the Borel sigma algebra generated by the random sample $\{\mathbf{X}_i\}_{i=1}^n$.

Lemma G.3. *Suppose $P_0 \in \mathcal{M}$ and that $t \in \Delta(P_0)$. Additionally, let $\{\mathbf{X}_i^\star\}_{i=1}^n$ denote the bootstrap sample. Then,* $\forall \epsilon > 0$

$$\textit{Prob}_P \left[\left| \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_i^\star; t) + \eta_n(t) \right| > \epsilon | \mathcal{A}_n \right] \xrightarrow{P} 0 \quad \textit{uniformly in} \quad \mathcal{M}. \tag{G.36}$$

Proof. Let $\overline{\Psi}_n(t) = \sum_{i=1} \overline{p}_i \left(g(\mathbf{X}_i; t) + \eta_n(t) \right)$. Given $\epsilon > 0$, by Markov's inequality and the triangular inequality, we have

$$\begin{aligned} \text{Prob}_P \left[\left| \frac{1}{n} \sum_{i=1}^n g(\mathbf{X}_i^\star; t) + \eta_n(t) \right| > \epsilon | \mathcal{A}_n \right] &\leq \epsilon^{-1} E_P \left[\left| \frac{1}{n} \sum_{i=1}^n \left(g(\mathbf{X}_i^\star; t) + \eta_n(t) \right) - \overline{\Psi}_n(t) \right| | \mathcal{A}_n \right] \\ &\quad + \epsilon^{-1} \left| \overline{\Psi}_n(t) \right|. \end{aligned} \tag{G.37}$$

Concentrating on the second term on the RHS of the inequality (G.37), by the triangular inequality we have

$$\left|\overline{\Psi}_n(t)\right| \leq \left| \overline{\Psi}_n(t) - \frac{1}{n} \sum_{i=1} g(\mathbf{X}_i; t) \right| + \left| \frac{1}{n} \sum_{i=1} g(\mathbf{X}_i; t) \right| \leq \|\overline{\mu}\|_{l^1_{\overline{w}_n}} + \left| \frac{1}{n} \sum_{i=1} g(\mathbf{X}_i; t) \right|. \tag{G.38}$$

By Lemma B.1, $\sup_{t \in [\underline{t}, \overline{t}]} \left| \frac{1}{n} \sum_{i=1} g(\mathbf{X}_i; t) - E_P[g(\mathbf{X}; t)] \right| \xrightarrow{P} 0$ uniformly in $\mathcal{M}$. Therefore, under $t \in \Delta(P_0)$, we must have $\left| \frac{1}{n} \sum_{i=1} g(\mathbf{X}_i; t) \right|$ converging to zero in probability uniformly in $\mathcal{M}$. Furthermore, Lemma G.2 implies $\|\overline{\mu}\|_{l^1_{\overline{w}_n}} \xrightarrow{P} 0$ uniformly in $\mathcal{M}$. Therefore, the second term on the RHS of the inequality (G.37) tends zero in probability uniformly in $\mathcal{M}$.

Now we concentrate on the first term on the RHS of the inequality (G.37). By the triangular inequality, we have

$$E_P\left[\left|\frac{1}{n}\sum_{i=1}^{n}\left(g\left(\mathbf{X}_i^{\star};t\right)+\eta_n(t)\right)-\overline{\Psi}_n(t)\right||\mathcal{A}_n\right]\leq E_P\left[\left|\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i^{\star};t\right)-\overline{\Psi}_n(t)\right||\mathcal{A}_n\right] + \eta_n(t) \tag{G.39}$$

and by the Cauchy-Schwartz inequality, the RHS of (G.39) is less than or equal to

$$\left(E_P\left[\left(\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i^{\star};t\right)-\overline{\Psi}_n(t)\right)^2|\mathcal{A}_n\right]\right)^{1/2}+\eta_n(t). \tag{G.40}$$

By assumption, $\eta_n(t)\xrightarrow{P}0$ uniformly in $\mathcal{M}$. Furthermore,

$$E_P\left[\left(\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i^{\star};t\right)-\overline{\Psi}_n(t)\right)^2|\mathcal{A}_n\right]=E_P\left[\frac{1}{n^2}\sum_{i\neq j}\left(g\left(\mathbf{X}_i^{\star};t\right)-\overline{\Psi}_n(t)\right)\left(g\left(\mathbf{X}_j^{\star};t\right)-\overline{\Psi}_n(t)\right)|\mathcal{A}_n\right] + E_P\left[\frac{1}{n^2}\sum_{i=1}^{n}\left(g\left(\mathbf{X}_i^{\star};t\right)-\overline{\Psi}_n(t)\right)^2|\mathcal{A}_n\right], \tag{G.41}$$

and

$$E_P\left[\frac{1}{n^2}\sum_{i\neq j}\left(g\left(\mathbf{X}_i^{\star};t\right)-\overline{\Psi}_n(t)\right)\left(g\left(\mathbf{X}_j^{\star};t\right)-\overline{\Psi}_n(t)\right)|\mathcal{A}_n\right]=0 \tag{G.42}$$

as $\left\{\mathbf{X}_j^{\star}\right\}_{j=1}^{n}$ is a random sample conditional $\mathcal{A}_n$. Hence,

$$\begin{aligned}E_P\left[\left(\frac{1}{n}\sum_{i=1}^{n}g\left(\mathbf{X}_i^{\star};t\right)-\overline{\Psi}_n(t)\right)^2|\mathcal{A}_n\right]&=E_P\left[\frac{1}{n^2}\sum_{i=1}^{n}\left(g\left(\mathbf{X}_i^{\star};t\right)-\overline{\Psi}_n(t)\right)^2|\mathcal{A}_n\right] && \text{(G.43)}\\ &=\frac{1}{n}E_P\left[\left(g\left(\mathbf{X}_i^{\star};t\right)-\overline{\Psi}_n(t)\right)^2|\mathcal{A}_n\right] && \text{(G.44)}\\ &\leq\frac{4}{n}\quad\text{by Assumption 2.1.} && \text{(G.45)}\end{aligned}$$

Therefore,

$$\left(E_P\left[\left(\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i^\star;t\right)-\overline{\Psi}_n(t)\right)^2|\mathcal{A}_n\right]\right)^{1/2}+\eta_n(t)\leq\frac{2}{\sqrt{n}}+\eta_n(t)\xrightarrow{P}0, \tag{G.46}$$

uniformly in $\mathcal{M}$.

Finally, putting together the above uniform convergence results in two parts from the RHS of the inequality (G.37), we have

$$\mathrm{Prob}_P\left[\left|\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i^\star;t\right)+\eta_n(t)\right|>\epsilon|\mathcal{A}_n\right]\xrightarrow{P}0\quad\text{uniformly in}\quad\mathcal{M}, \tag{G.47}$$

which is the desired result. □

Let Σ_w be the covariance matrix defined in Section 2.4, and let $\hat{\Sigma}_w^\star(\eta)$ be its sample analogue estimator based on the bootstrap sample.

Lemma G.4. *Suppose $P_0\in\mathcal{M}$. Then the following statements hold.*

1. $\forall\epsilon>0,$ *Prob*$_P\left[\|\hat{\Sigma}_w^\star(\eta)-\Sigma_w\|>\epsilon|\mathcal{A}_n\right]\xrightarrow{P}0$ *uniformly in $\mathcal{M}$, where $\|\cdot\|$ is the operator norm (2.9).*
2. *Let $e_0\in\mathbb{R}_+$, $\forall\epsilon>0,$ Prob$_P\left[\left\|\left(\hat{\Sigma}_w^\star(\eta)\right)^{-1}-\Sigma_w^{-1}\right\|>\epsilon|\mathcal{A}_n\right]\xrightarrow{P}0$ uniformly in $\mathcal{M}(e_0)$, where $\|\cdot\|$ is the operator norm (2.10).*

Proof. **Part 1**. Firstly, under Assumption 2.1, the class of functions

$$\mathcal{GG}=\left\{\mathbf{x}\mapsto g\left(\mathbf{x};t\right)g\left(\mathbf{x};t'\right),\quad t,t'\in[\underline{t},\overline{t}]\right\} \tag{G.48}$$

is also a uniformly bounded VC-class. We will use this result in the proof. We have

$$\hat{\Sigma}_w^\star(\eta)-\Sigma_w=\hat{\Sigma}_w^\star(\eta)-\overline{\Sigma}_w+\overline{\Sigma}_w-\hat{\Sigma}_w+\hat{\Sigma}_w-\Sigma_w, \tag{G.49}$$

where $\overline{\Sigma}_w$ is the matrix with $\overline{\Sigma}_{w,k,j} = \sum_{i=1}^n \overline{p}_i g\left(\mathbf{X}_i; t_k\right) g\left(\mathbf{X}_i; t_j\right)$, and $\hat{\Sigma}_w$ is the matrix with $\hat{\Sigma}_{w,k,j} = \frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}_i; t_k\right) g\left(\mathbf{X}_i; t_j\right)$. Then by the triangular inequality

$$\|\hat{\Sigma}^\star_w(\eta) - \Sigma_w\| \leq \|\hat{\Sigma}^\star_w(\eta) - \overline{\Sigma}_w\| + \|\overline{\Sigma}_w - \hat{\Sigma}^\star_w\| + \|\hat{\Sigma}_w - \Sigma_w\|. \tag{G.50}$$

We will prove the result of this part of the lemma by showing that each part on the RHS of (G.50) converges to zero. We have that

$$\|\hat{\Sigma}_w - \Sigma_w\| \leq \sup_{t_k, t_j \in \Delta(P)} \left|\hat{\Sigma}_{w,k,j} - \Sigma_{w,k,j}\right| \leq \sup_{t_k, t_j \in [\underline{t}, \overline{t}] \cap \mathbb{Q}} \left|\hat{\Sigma}_{\infty,k,j} - \Sigma_{\infty,k,j}\right| \xrightarrow{P} 0 \tag{G.51}$$

uniformly in $\mathcal{M}$, since the class of moment functions (G.48) is a uniformly bounded VC class, and hence, Glivenko-Cantelli uniformly in $\mathcal{M}$. We also have that

$$\|\overline{\Sigma}_w - \hat{\Sigma}_w\| \leq \sup_{t_k, t_j \in \Delta(P)} \left|\overline{\Sigma}_{w,k,j} - \hat{\Sigma}_{w,k,j}\right| \leq \|\overline{\mu}\|_{l_{\frac{1}{w}n}} \xrightarrow{P} 0 \quad \text{uniformly in} \quad \mathcal{M}, \tag{G.52}$$

which follows from Lemma G.2.

This leaves us with the first term on the RHS of (G.50). We have that

$$\begin{aligned}
\|\hat{\Sigma}^\star_w(\eta) - \overline{\Sigma}_w\| &\leq \sup_{t_k, t_j \in \Delta(P)} \left|\Sigma^\star_w(\eta) - \overline{\Sigma}_w\right| \\
&\leq 2 \sup_{t \in [\underline{t}, \overline{t}]} \eta_n(t) + \left(\sup_{t \in [\underline{t}, \overline{t}]} \eta_n(t)\right)^2 \\
&\quad + \sup_{t_k, t_j \in \Delta(P)} \left|\frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}^\star_i; t_k\right) g\left(\mathbf{X}^\star_i; t_j\right) - \sum_{i=1}^n \overline{p}_i g\left(\mathbf{X}_i; t_k\right) g\left(\mathbf{X}_i; t_j\right)\right|,
\end{aligned} \tag{G.53}$$

and that $\sup_{t \in [\underline{t}, \overline{t}]} \eta_n(t) \xrightarrow{P} 0$ uniformly in $\mathcal{M}$ by assumption. Therefore, to conclude the proof of this part of the lemma, all that remains is to show that

$$\sup_{t_k, t_j \in \Delta(P)} \left|\frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}^\star_i; t_k\right) g\left(\mathbf{X}^\star_i; t_j\right) - \sum_{i=1}^n \overline{p}_i g\left(\mathbf{X}_i; t_k\right) g\left(\mathbf{X}_i; t_j\right)\right| \tag{G.54}$$

converges to zero conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}$. We will show that (G.54) converges in mean to zero conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}$, which implies the desired result.

Let $\{\epsilon_i\}_{i=1}^n$ are independent Rademacher variables that are independent of $\{\mathbf{X}_i^\star\}_{i=1}^n$ and $\{\mathbf{X}_i\}_{i=1}^n$. We have that

$$\begin{aligned}
&\mathrm{E}_P\left[\sup_{t_k,t_j\in\Delta(P)}\left|\frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}_i^\star;t_k\right)g\left(\mathbf{X}_i^\star;t_j\right)-\sum_{i=1}^n \overline{p}_i g\left(\mathbf{X}_i;t_k\right)g\left(\mathbf{X}_i;t_j\right)\right| |\mathcal{A}_n\right]\\
&\le \mathrm{E}_P\left[\sup_{t_k,t_j\in[\underline{t},\overline{t}]}\left|\frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}_i^\star;t_k\right)g\left(\mathbf{X}_i^\star;t_j\right)-\sum_{i=1}^n \overline{p}_i g\left(\mathbf{X}_i;t_k\right)g\left(\mathbf{X}_i;t_j\right)\right| |\mathcal{A}_n\right]\\
&\qquad\le \mathrm{E}_P\left[\mathrm{E}_\epsilon\left[\sup_{t_k,t_j\in[\underline{t},\overline{t}]}\left|\frac{1}{n}\sum_{i=1}^n \epsilon_i g\left(\mathbf{X}_i^\star;t_k\right)g\left(\mathbf{X}_i^\star;t_j\right)\right|\right]|\mathcal{A}_n\right],
\end{aligned}\tag{G.55}$$

where the last inequality follows from an application of the from the symmetrization lemma (Lemma 2.3.1 in van der Vaart and Wellner, 1996) applied to conditional expectations. Now let $\mathbb{P}_n^\star$ be the empirical measure based on the bootstrap sample. Fix $\mathbf{X}_1^\star,\mathbf{X}_2^\star,\ldots,\mathbf{X}_n^\star$, and let $\mathcal{H}$ be a $\epsilon-$net in $L_1\left(\mathbb{P}_n^\star\right)$ over $\mathcal{GG}$. Then

$$\mathrm{E}_\epsilon\left[\sup_{t_k,t_j\in[\underline{t},\overline{t}]}\left|\frac{1}{n}\sum_{i=1}^n \epsilon_i g\left(\mathbf{X}_i^\star;t_k\right)g\left(\mathbf{X}_i^\star;t_j\right)\right|\right]\le \mathrm{E}_\epsilon\left[\sup_{h\in\mathcal{H}}\left|\frac{1}{n}\sum_{i=1}^n\epsilon_i h\left(\mathbf{X}_i^\star\right)\right|\right]+\epsilon \tag{G.56}$$

Furthermore, the RHS of (G.56) is less than or equal to

$$\begin{aligned}
&\sqrt{1+\log\left(N\left(\epsilon,\mathcal{GG},L_1\left(\mathbb{P}_n^\star\right)\right)\right)}\sup_{h\in\mathcal{H}}\left\|\frac{1}{n}\sum_{i=1}^n\epsilon_i h\left(\mathbf{X}_i^\star\right)\right\|_{\psi_2|\mathbf{X}^\star}+\epsilon\\
&\qquad\le\sqrt{1+\log\left(N\left(\epsilon,\mathcal{GG},L_1\left(\mathbb{P}_n^\star\right)\right)\right)}\sqrt{\frac{6}{n}}+\epsilon,
\end{aligned}\tag{G.57}$$

where the Orlicz norms $\|\cdot\|_{\psi_2|\mathbf{X}^\star}$ are taken over $\{\epsilon_i\}_{i=1}^n$ with $\{\mathbf{X}_i^\star\}_{i=1}^n$ fixed, and $N\left(\epsilon,\mathcal{GG},L_1\left(\mathbb{P}_n^\star\right)\right)$ is the minimal number of balls of radius ϵ in the $L_1\left(\mathbb{P}_n^\star\right)$ metric needed to cover the set $\mathcal{GG}$.

The VC property of $\mathcal{GG}$ implies $\sup_Q\log\left(N\left(\epsilon,\mathcal{GG},L_1\left(Q\right)\right)\right)<+\infty$, where the supremum is taken over all probability measures, Q. Hence, the RHS of (G.57) is bounded above by

$$\sqrt{1+\sup_Q \log\left(N\left(\epsilon,\mathcal{GG},L_1\left(Q\right)\right)\right)}\sqrt{\frac{6}{n}}+\epsilon, \tag{G.58}$$

which does not depend on $\{\mathbf{X}_i^\star\}_{i=1}^n$, $\{\mathbf{X}_i\}_{i=1}^n$, and P. Therefore, $\forall\epsilon>0$,

$$\mathrm{E}_P\left[\sup_{t_k,t_j\in\Delta(P)}\left|\frac{1}{n}\sum_{i=1}^n g\left(\mathbf{X}_i^\star;t_k\right)g\left(\mathbf{X}_i^\star;t_j\right)-\sum_{i=1}^n\overline{p}_i g\left(\mathbf{X}_i;t_k\right)g\left(\mathbf{X}_i;t_j\right)\right||\mathcal{A}_n\right]\leq \sqrt{1+\sup_Q \log\left(N\left(\epsilon,\mathcal{GG},L_1\left(Q\right)\right)\right)}\sqrt{\frac{6}{n}}+\epsilon. \tag{G.59}$$

This concludes the proof of this part upon realizing that

$$\sqrt{1+\sup_Q \log\left(N\left(\epsilon,\mathcal{GG},L_1\left(Q\right)\right)\right)}\sqrt{\frac{6}{n}}\to 0$$

as $n\to+\infty$.

Part 2. Since $\left(\hat{\Sigma}_w^\star(\eta)\right)^{-1}-\Sigma_w^{-1}=\left(\hat{\Sigma}_w^\star(\eta)\right)^{-1}\left(\hat{\Sigma}_w^\star(\eta)-\Sigma_w\right)\Sigma_w^{-1}$, we have

$$\begin{aligned}\left\|\left(\hat{\Sigma}_w^\star(\eta)\right)^{-1}-\Sigma_w^{-1}\right\| &\leq \left\|\left(\hat{\Sigma}_w^\star(\eta)\right)^{-1}\right\|\left\|\Sigma_w^{-1}\right\|\left\|\hat{\Sigma}_w^\star(\eta)-\Sigma_w\right\| &\text{(G.60)}\\ &\leq e_0\left\|\left(\hat{\Sigma}_w^\star(\eta)\right)^{-1}\right\|\left\|\hat{\Sigma}_w^\star(\eta)-\Sigma_w\right\|, &\text{(G.61)}\end{aligned}$$

which is less than or equal to $e_0^2 o_P(1)$ conditional on $\mathcal{A}_n$ uniformly in $\mathcal{M}(e_0)$ for large enough n by Part 1 of this lemma. □

H The Bootstrap Procedure of LSW

In this section, we outline the steps in the bootstrap procedure of Linton et al. (2010) (LSW henceforth) which is used in the MC simulation experiments in Section 5. LSW use an integral-type test statistic. In the setting of the paper, it is

$$\hat{T}_n = \int_{\underline{t}}^{\overline{t}} \left(\max \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g\left(\mathbf{X}_i; t\right), 0 \right\} \right)^2 dt. \tag{H.1}$$

The estimate of the contact set we use in the MC simulations is

$$\hat{\mathcal{C}} = \left\{ t \in [\underline{t}, \overline{t}] : \frac{1}{n} \sum_{i=1}^{n} g\left(\mathbf{X}_i; t\right) > -\eta_n(t) \right\}, \tag{H.2}$$

where the random numbers $\eta_n(t)$ satisfy property (4.2). In the MC simulations we use

$$\eta_n(t) = \hat{\sigma}_t \sqrt{\frac{2 \log n}{n}},$$

where $\hat{\sigma}_t^2$ is the sample analogue estimator of σ_t^2. Because of the continuity of the moment functions, $\hat{\mathcal{C}}$ will always have positive Lebesgue measure when it is nonempty.

The bootstrap DGP LSW use is the ECDF on the data. Let $\{\mathbf{X}_i^\star\}_{i=1}^n$ be a random sample from the ECDF of the data, then their bootstrap test statistic is

$$T_n^\star = \begin{cases} \int_{\underline{t}}^{\overline{t}} \left(\max \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g\left(\mathbf{X}_i^\star; t\right) - \frac{1}{n} \sum_{i=1}^{n} g\left(\mathbf{X}_i; t\right) \right], 0 \right\} \right)^2 dt, & \text{if } \hat{\mathcal{C}} = \emptyset, \\ \int_{\hat{\mathcal{C}}} \left(\max \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g\left(\mathbf{X}_i^\star; t\right) - \frac{1}{n} \sum_{i=1}^{n} g\left(\mathbf{X}_i; t\right) \right], 0 \right\} \right)^2 dt, & \text{if } \hat{\mathcal{C}} \neq \emptyset. \end{cases}$$

Letting B_n be the number of bootstrap replications, the approximate bootstrap p-value is defined as

$$\Upsilon_{B_n}^{\text{LSW}} = \frac{1}{B_n} \sum_{j=1}^{B_n} 1\left[T_{n,j}^\star \geq \hat{T}_n \right], \tag{H.3}$$

and one rejects H_0 if $\Upsilon_{B_n}^{\text{LSW}} \leq \beta$, where $\beta \in (0, 1/2)$ is a given nominal level.

I The Bootstrap Procedure of AS

In this section, we outline the steps of the bootstrap procedure proposed by Andrews and Shi (2010)(AS henceforth) which is used in the MC simulation experiments in Section 5. AS propose a Kolmogorov-Smirnov and Cramér von MIses test statistics for inference on possibly infinite number of conditional moment inequality conditions. Recall that the setting of the paper considers a continuum of unconditional moment inequality conditions, which the AS procedure covers. In this setting, the AS test statistics are identical, and it is given by

$$\hat{T}_n = \sup_{t\in[\underline{t},\overline{t}]} \left(\max\left\{ \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i;t\right)\right)/\hat{\sigma}(t), 0 \right\}\right)^2, \quad \text{where} \tag{I.1}$$

$$\hat{\sigma}^2(t) = \frac{1}{n}\sum_{i=1}^{n} g^2\left(\mathbf{X}_i;t\right) - \left[\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i;t\right)\right]^2. \tag{I.2}$$

Next we describe the steps for computing the bootstrap GMS critical value of AS in the setting of the paper. The critical value is obtained through the following steps.

1. Compute $\overline{\varphi}_n(t)$ for $t \in [\underline{t}, \overline{t}]$, where $\overline{\varphi}_n(t)$ is defined as follows. Let

   $$\xi_n(t) = \kappa_n^{-1}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i;t\right)\right)/\hat{\sigma}(t), \tag{I.3}$$

   where $\kappa_n = (0.3\log(n))^{1/2}$. Define

   $$\overline{\varphi}_n(t) = \hat{\sigma}(t)B_n 1\left[\xi_n(t) < -1\right] \quad \text{and} \quad B_n = (0.4\log(n)/\log\log(n))^{1/2}. \tag{I.4}$$

2. Generate B bootstrap samples $\left\{\mathbf{X}^{\star}_{i,s}\right\}_{i=1}^{n}$ for $s = 1, \ldots .B$ using the ECDF on the data.

3. For each bootstrap sample, compute $\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}^{\star}_{i,s};t\right)$, and $\hat{\sigma}^2_s(t)$ just as $\hat{\sigma}^2(t)$ is computed but with the bootstrap sample in place of the original sample.

4. For each bootstrap sample, compute the bootstrap test statistic $\hat{T}^{\star}_{n,s}$ as $\hat{T}_n$ is computed in (I.1)

but with $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i;t\right)\right)$ replaced by

$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_{i,s}^{\star};t\right) - \frac{1}{n}\sum_{i=1}^{n} g\left(\mathbf{X}_i;t\right) - \overline{\varphi}_n(t)\right)$ and with $\hat{\sigma}^2(t)$ replaced by $\hat{\sigma}_s^2(t)$.

5. Take the bootstrap GMS critical value $c_{n,1-\beta}$ to be the $1-\beta+\eta$ sample quantile of the bootstrap test statistics $\left\{\hat{T}_{n,s}^{\star}, s=1,\ldots,B\right\}$ plus η, where $\eta = 10^{-6}$.

For a given nominal level $\beta \in (0,1/2)$, the AS test rejects H_0 if $\hat{T}_n > c_{n,1-\beta}$.